Teaching Clinical Applications of the Bayes' Theorem in a QBIC Statistics Course

Ramon Gomez
Florida International University, 11200 SW 8th Street, Miami, FL 33199

Abstract
A widely used application of conditional probability is found in the health sciences field while evaluating the diagnostic ability of clinical tests. Positive and negative predictive values are major indicators in this assessment that represent applications of the Bayes' theorem not usually included in regular statistics courses for undergraduates. This paper describes the present author's experience in teaching clinical predictive values for a first statistics course at Florida International University. Students were biology majors enrolled in a special program identified as QBIC, an acronym for Quantifying Biology in the Classroom. The program is interdisciplinary and emphasizes the use of statistics for analyses of biological/biomedical data. QBIC scholars were exposed to various technology resources while learning predictive values. This teaching-learning approach demonstrated to be highly effective and provided QBIC students with a very important tool for future biostatistics courses and research activities.

Keywords: Bayes' theorem, Predictive values, QBIC, Technology resources

1. Introduction

The BIO2010 report (National Academy of Sciences, 2003) produced by a National Research Council committee offered recommendations for a new curriculum in biology that emphasizes a strong foundation in quantitative sciences. This report recommended in particular an increased exposure of biology students to statistical methods. Following this approach, a special undergraduate program for selected biology majors was inaugurated at Florida International University (FIU) in 2007. This undergraduate program identified as QBIC, an acronym for "Quantifying Biology in the Classroom" is rigorous and interdisciplinary. QBIC is a block program with a lock-step curriculum designed to bolster student links between subject areas. It highlights the use of mathematics and statistics for analyses of biological/biomedical data and to this end includes two statistics courses during the sophomore year (Tashakkori et al, 2008).

The QBIC Statistics I course is a three credit-hours class covering descriptive statistics, probability and inferential statistics based on a single sample. This course was developed by the present author following the program guidelines established by Dr. Samuel Shapiro, now retired and Professor Emeritus at FIU. Several non-standard topics for introductory statistics courses, accentuating biological/biomedical applications, were added to our QBIC course (Gomez, 2010 & 2011). The subject of clinical predictive values as an application of the Bayes' theorem was among them.

This paper reports the present author's experience in teaching clinical predictive values for a first statistics course at Florida International University. The implemented teaching approach demonstrated to be highly effective in providing QBIC students with an important tool for future biostatistics courses and biomedical research activities.

2. Method

2.1 Course design and organization

The QBIC Statistics I course has been taught during the fall term of the students' sophomore year, with the present author as the instructor, since 2008. The text book, Biostatistics: A Foundation for Analysis in the Health Sciences by Wayne Daniel, requires mid-level mathematical prerequisites. It includes many proposed exercises for which real data from the health sciences is utilized to illustrate statistical applications. The course encompasses contents from chapter 1 to 7, covering a typical range of topics: descriptive statistics, probability, and inferential statistics based on one sample. The Bayes' Theorem falls under chapter 3 (Basic Probability) where clinical predictive values are discussed as an application. While teaching predictive values various statistical tools are introduced to the students such as contingency tables, probability trees and the multiplication rule, as well as a number of clinical concepts.

The traditional approach to teach Statistics consists of using a board during lectures, a textbook as a reference, and supplementary material posted on a website. However, the appropriate use of technology has been identified as a significant contributor to the students' understanding. Incorporation of technology resources can make college teaching of statistics more effective as it improves the quality of instruction, encourages students' active learning, and provides them with psychological incentives (Garfield, 1995; Higazi, 2002). Furthermore, a committee created by the American Statistical Association that produced the Guidelines for Assessment and Instruction in Statistics Education (GAISE by ASA, 2005), recommended the use of technology for introductory statistics courses at college level. In that regards, QBIC scholars are exposed to various technology resources for the Statistics classes, including PowerPoint presentations for lectures and a variety of computational tools, which facilitate students' understanding (Hulsizer & Woolf, 2009). A course pack developed by the present author and equipped with the PowerPoint slides is made available to the students at the university bookstore.

The QBIC Statistics I course includes twenty eight class meetings, two per week, seventy five minutes each. A computer connected projector, in conjunction with an eighty inch screen, is used for the PowerPoint presentations. A dry erase board is also available as a supplement for class discussions. Students are required to bring the course pack with the PowerPoint slides to every class and hence, limited notes are needed during lecture time. Consequently, students are more focused on the discussion of statistical concepts and exercises solution.

2.2 Class example

Troponin-I is a protein that plays an important role in muscle contraction. Several reports in the literature indicate that cardiac Troponin-I is released into blood within hours of myocardial infarction symptoms. Levels of Troponin-I (ng/mL) are significantly increased as a result of a heart attack. An immunoassay test based on this protein was developed in the 1990's by Baxter Diagnostics Inc. to aid in the diagnosis of Myocardial Infarction (MI).

A clinical study evaluating the Troponin-I test performance was conducted by Baxter as part of the process for test approval by the Food and Drug Administration (FDA). A total of 314 patients with chest pain arriving to the emergency rooms at four US hospitals were evaluated for MI using:

- Physician examination
- Electro-Cardio-Gram (ECG)
- Existing gold standard blood test for MI

The 314 patients were classified as having MI and non-MI based on this evaluation. The Troponin-I test was then given to the patients with two possible outcomes of Positive (+) or Negative (-) for MI, using a cut-off level previously determined. The Table 1 below summarizes the results of this clinical study as they are presented to the QBIC students.

Table 1: Clinical Data

Test result		Patient status: MI	Patient status: Non-MI	
		MI	Non-MI	
	+	119 TP	2 FP	121
	–	9 FN	184 TN	193
		128	186	314 Total

Source: Baxter Diagnostics Troponin-I Package Insert

The assessment of the test's performance, including the concepts of true positive (TP), true negative (TN), false positive (FP), false negative (FN) results as well as the clinical sensitivity and specificity, is introduced to QBIC scholars. Clinical sensitivity and specificity are performance characteristics of the test that do not involve the frequency of the disease in the given population. The calculation of sensitivity and specificity is desribed as follows:

Clinical Sensistivity = TP / MI

Clinical Specificity = TN / Non MI

Using data from Table 1 students are able to calculate the clinical sensitivity and specifity as 93.0% and 98.9% respectively. Then, it is shown that these two test performance characteristics can be seen as conditional probabilities:

Clinical Sensitivity = P (+ | MI)

Clinical Specificity = P(- | Non MI)

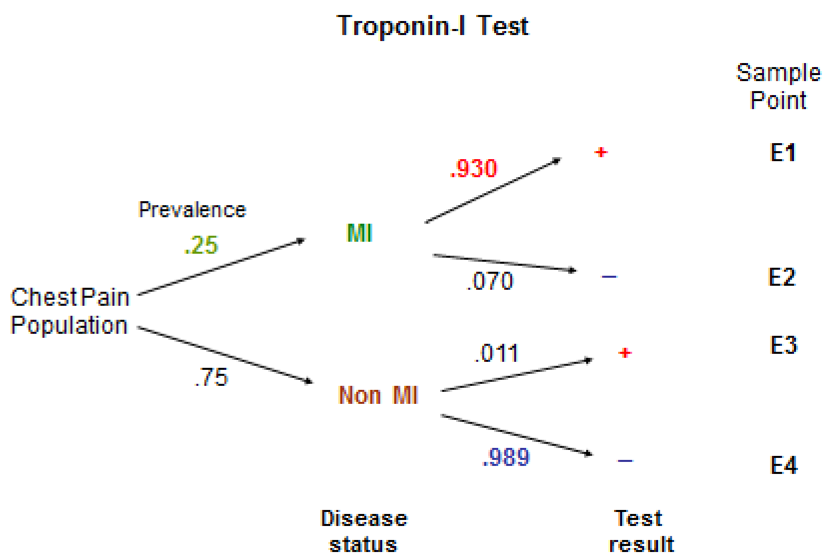
However, the questions for a clinician and a patient are different: What is the probability that a person with a positive result truly has the disease? What is the probability that a person with a negative result does not have actually the disease? These other two probabilities represent clinical concepts called Predictive Values and involve the frequency or prevalence of the disease. These crucial concepts are introduced to the QBIC scholars and then identified by conditional probabilities as follows:

Predictive Value Positive = P(MI | +)

Predictive value Negative = P (non MI | -)

Predictive Values calculation depends on the clnical sensitiviy and specificity of the test as well as the prevalence of the disease in the given population. At this point, a tree diagram including these pieces of information is presented to the students, assuming an MI prevalence of 25% among the population of chest pain patients arriving to emergency rooms. Four sample points are obtained and their probabilities calculated by using the multiplication rule (Figure 1).

Figure 1: Probability Tree



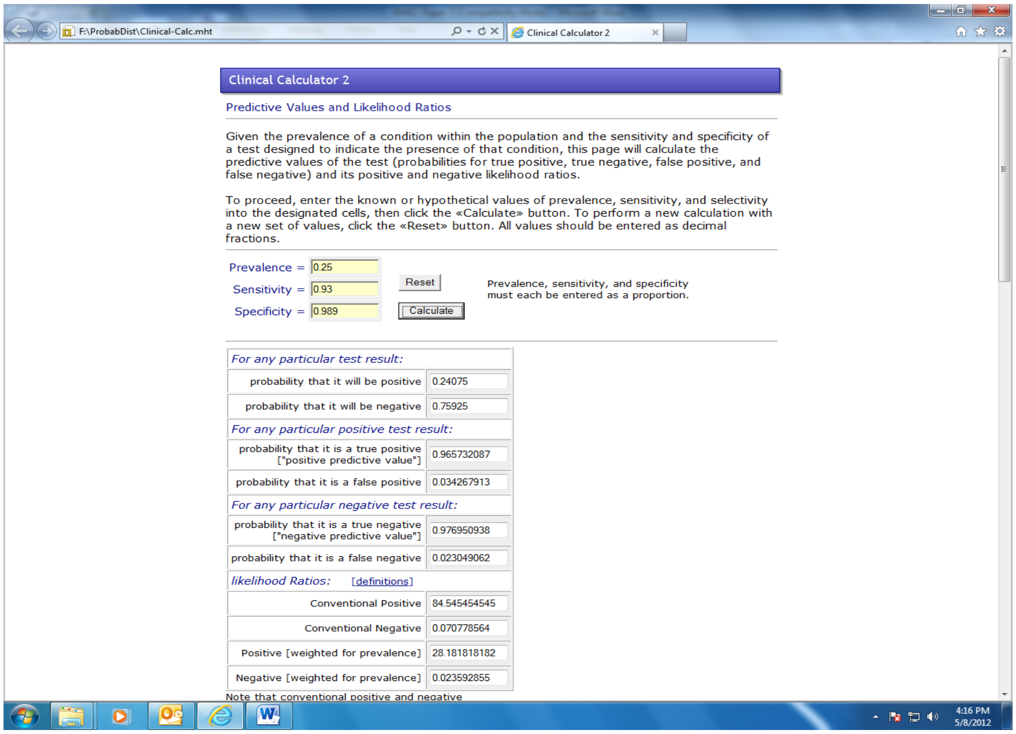
The predictive values, as posterior probabilities, can be determined by using the Bayes' Theorem and hence, a hand calculation is illustrated using the probabilities for the four sample points from Figure 1 as follows:

P(MI | +) = P(E1) / [P(E1) + P(E3)]

P (non MI | -) = P(E4) / [P(E2) + P(E4)]

The hand calculations of PVP = .967 and PVN = 0.976 are then validated in class using an online calculator (***http://vassarstats.net/clin2.html)*** that requires the input of disease prevalence, sensitivity and specificity (Figure 2). The use of this online calculator is highly recommended to the students as a tool for the computation of predictive values that saves time and prevent numerical errors..

Figure 2: Online Calculator for Predictive Values



In this case it is observed that both positive (PVP) and negative (PVN) predictive values are very high probabilities indicating that the test can be used as an aid in the diagnosis of the disease. Other examples are discussed with the students where the PVP is low while the PVN is very high suggesting that the test is more appropriate for screening of the disease/medical condition rather than for its diagnosis

3. Discussion and Conclusions

Clinical predictive values can be considered a non-standard topic for introductory statistics courses at college. Typically this subject is taught as part of undergraduate or graduate courses for students majoring in Statistics/Biostatistics. However, teaching predictive values provides a significant motivation to any undergraduate students taking introductory statistics courses and facilitates their understanding of the Bayes' Theorem. The basic application discussed in this paper, where the sample space is partitioned in only two complementary events and real data was used, illustrates a very constructive scenario for the students' learning process. As a measure of learning assessment, predictive values were included in one partial test and the final exam for QBIC scholars with very satisfactory results. Moreover, their motivation and satisfaction was high as evidenced by the 81% of excellent-very good opinions of students about the quality of instruction during the period 2008-2011.

This teaching-learning approach demonstrated to be highly effective and provided QBIC students with a very important tool for future biostatistics courses and clinical research activities. The successful implementation of this approach also showed the feasibility of teaching predictive values to undergraduates with very little statistical training.

References

American Statistical Association. (2005). Guidelines for assessment and instruction in statistics education (GAISE): College Report. www.amstat.org/education/gaise.

Baxter Diagnostics Inc. (1995). Stratus cardiac Troponin-I fluorometric enzyme immunoassay.

Daniel, W. W. (2009). Biostatistics: A foundation for analysis in the health sciences, 9th edition. Hoboken, NJ: John Wiley & Sons.

Garfield, J.B. (1995). How students learn statistics. International Statistics Review, 63, 25-34.

Gomez, R. (2010). Innovations in teaching undergraduate statistics courses for biology students. *Review of Higher Education and Self Learning, Vol. 3, Issue 7, 8-13*

Gomez, R. (2011). Teaching Innovations for QBIC Statistics Courses at Florida International University. USCOTS 2011 Proceedings.

Higazi, S. M. (2002). Teaching statistics using technology. ICOTS6 Proceedings, 2002.

Hulsizer, M. C. & Woolf, L. M. (2009). A guide to teaching statistics. Malden, MA: Wiley-Blackwell.

National Academy of Sciences. (2003). BIO2010: Transforming undergraduate education for future research biologists. www.nap.edu/catalog/10497.html

Tashakkori, A. Reio, T. G., & Rincon, D. (2008). Quantifying biology in the classroom: An innovative program that integrates the biological sciences, physical sciences, mathematics, and statistics. *Evaluation Report*, Florida International University.