

Numeracy

Volume 1, Issue 2

2008

Article 5

False Positives and Referral Bias: Content for a Quantitative Literacy Course

Stuart Boersma*

Teri Willard†

*Dept. of Mathematics, Central Washington University, boersmas@cwu.edu

†Dept. of Mathematics, Central Washington University, willardt@cwu.edu

False Positives and Referral Bias: Content for a Quantitative Literacy Course

Stuart Boersma and Teri Willard

Abstract

An extended study of accuracy in medical screening is presented as a useful application to increase students' quantitative reasoning skills. Two detailed examples are presented. The first explores the frequency of obtaining false positive results from a medical screening tool while the second examines the issue of referral bias and its effect on the apparent sensitivity and specificity of the screening tool. Results from student assessments indicate that the activity increases one's ability to define terms such as "false positive" and "false negative" and increases one's ability to read and compute with information obtained from a two-way table. Teacher assessment results indicate that the activity is challenging and could be used in existing high school or college classrooms. Additionally, links to a student activity, instructor notes, and Excel calculation tool are provided.

KEYWORDS: sensitivity, specificity, referral bias, two-way tables, medical screening

Introduction

A recent *San Francisco Chronicle* article (Guthrie 2004) tells a depressing story of a man who believed for eight years that he had HIV. Jim Malone suffered immensely (physiologically, socially, and mentally) from this medical misdiagnosis. While this occurrence of a *false positive* result in an HIV diagnosis is perhaps rare, the scare inflicted by other medical screening tools is a very real and frequent phenomenon. *The Boston Globe* recently reported that “nearly half the healthy men and women 55 and older who underwent regular screening for four leading cancers received at least one test result incorrectly suggesting they might have cancer, called a false positive” (Allen 2007). These false positive results cause a large number of people to worry unnecessarily and often undergo needless surgical procedures.

Newspaper readers are also often confronted with individuals claiming that they have been falsely accused of some misdeed because of a false positive test result. Consider, for example, that Floyd Landis was recently stripped of his *Tour de France* trophy because he tested positive for performance-enhancing steroids. Landis’ doctor has claimed that Landis was the victim of a false positive result (ABC News 2006).

Stories such as these – together with drug testing in the workplace and in our school athletic programs – make the subject of false positives an important and timely topic for our students’ consideration. For added interest, we note that *Mathematics Magazine* has recently published a satirical poem (Memory 2007) claiming that the suicide of the fictional literary gentleman Richard Cory could have been the result of not understanding the true likelihoods of receiving false positives. Richard Cory, one might recall, is the title character in a song adapted by Simon and Garfunkel from a narrative poem (1897) of the same title by the American poet and playwright, Edwin Arlington Robinson. As Professor Memory notes in a footnote, the fatal faulty logic he refers to in his new poem is Mr. Cory’s confusion of the probability of having the disease if one is positively diagnosed with it vs. the probability of being positively diagnosed if one has the disease.

The phrase *quantitative literacy* has evolved over the years and has now influenced many national and state-level mathematics standards (See Madison and Steen 2008 for a detailed history of *quantitative literacy* and *numeracy*.). For the authors of this paper, *quantitative literacy* will refer to the ability to reason critically with quantitative information. Or, more basically put, a *quantitatively literate* individual is capable of performing simple mathematical tasks that arise in a complex realistic setting. Because Central Washington University is dedicated to increasing the level of quantitative literacy for its students, we have been developing quantitatively rich student activities. We have found that studying the

phenomenon of false positives provides an ideal educational setting for students to practice some simple mathematical skills (proportions, proportional reasoning, percents, and the notion of likelihood or probability) in a complex realistic setting (the world of medical screening tools). Furthermore, we have found, a more-complete understanding of the quantitative subtleties of medical screening tools enriches the public, personal, and professional lives of teachers as well as students.

Examples and discussions surrounding the probability of receiving a false positive result appear in some standard college textbooks, although they are not common. Typically, when these examples appear, they are given a fairly cursory treatment. We know of three notable exceptions. Bennett and Briggs (2005) spend about a page discussing a false positive in the context of mammograms and include several problems on the topic in the homework exercises. Garfunkel (1998) presents several extended activities that allow students to explore some of the subtleties surrounding medical testing. Sevilla and Somers (2007) offer the only extended example we are aware of that uses the terminology typically associated with discussions of the effectiveness and reliability of medical screening tools.

We have found no educational materials that explore the most obvious question: *How does one know if a “positive” result is a true positive or a false positive?* It turns out that the answer to this question uncovers yet more counterintuitive quantitative subtleties. In this paper, we give an overview of the role of quantitative reasoning in studying the phenomenon of false positive test results. We provide classroom-ready materials and notes to instructors for them in Related Files 1 and 2, respectively. In the paper, we report results of student and teacher assessments of these materials. Related File 3 is an Excel tool to aid interested readers in exploring the calculations.

Terminology

Consider the following scenario:

Scene: Doctor’s Office. Doctor and patient are present.

Doctor: As you know, last week we ran a standard screening test for ailment *X*. I’m sorry to have to inform you today that the results have come back positive.

Patient: Does this mean I have *X*?

Doctor: Well, the test is considered to be very reliable. In fact, studies have shown that this test correctly identifies 99% of all patients with X and correctly screens out 99% of those who do not have X . Of course while we can't be entirely sure until we perform the customary surgery, there is a very high probability that you have X .

Patient is clearly distraught at the strong possibility of having ailment X and having to undergo more expensive and invasive testing procedures.

The patient needs to recognize that the doctor did not answer her question. The doctor implied, and perhaps believed, that the chance for a false positive is 1% and hence not very likely. This conclusion, however, is not correct as some simple computations can show.

Before examining the topic further, we need the basic terminology:

- *Sensitivity*: The probability (or likelihood) that a test produces a “positive” result when a patient is known to have the ailment.
- *Specificity*: The probability (or likelihood) that a test produces a “negative” result when a patient is known to be ailment free.
- *False Positive (FP)*: The situation when a patient receives a “positive” result even though the patient is ailment free.
- *False Negative (FN)*: The situation when a patient receives a “negative” result even though the patient has the ailment.
- *True Positive (TP)*: The situation when a patient correctly receives a “positive” result.
- *True Negative (TN)*: The situation when a patient correctly receives a “negative” result.

In our scenario of the patient facing the diagnosis of ailment X , both the sensitivity and specificity of the test were 99%, and the patient was clearly interested in the probability that her test was a false positive. We can assume that she was hopeful that the positive was indeed a false positive

In many cases, she might be interested in the probability of a false negative as well. For instance, if the ailment is life threatening but can be effectively treated if detected early enough, then she would want to make sure that the negative test result does in fact mean that she is ailment free.

As one looks into the reliability of screening tests, one finds that the sensitivity and specificity are often reported. However, it is very unlikely to find

any further discussion of the probability that the test produced a false negative or a false positive. To see why consider the following example.

Example

The Medical University of South Carolina Doctoring Curriculum (MUSCDC 2007) provides information for a realistic example.

An ELISA (Enzyme-Linked ImmunoSorbent Assay) is a biochemical technique to detect the presence of particular antibodies. The example concerns an ELISA that is developed to diagnose HIV infections. Assume that it has both a sensitivity and specificity of 99.9%. Assume the test is applied to a million people, 1% of whom are infected with HIV. How many of those million people will receive a false positive? What is the probability that a positive result is a false positive?

The quantitative information can easily be organized into a two-way table (Fig. 1). Because 1% of the population under consideration is HIV+, 10,000 people are HIV+ while the remaining 990,000 are HIV-. This information provides the totals for the two columns, and one can start to fill in the table. The sensitivity value of 99.9% implies that the ELISA will return a positive result on 99.9% of those who are truly HIV+. Thus, the true positives (TP) are $0.999 \times 10,000 = 9,990$. Similarly, the given specificity indicates that 99.9% of the HIV- subjects will be correctly identified; the number of true negatives (TN) is 989,010. The entries for false positives (FP) and false negatives (FN) are easily obtained by subtraction from the column totals. The row totals are easily found by addition, which completes the table. We have found that 990 of the million people who are tested actually receive a false positive.

GENERAL POPULATION	HIV-positive	HIV-negative	TOTAL
ELISA-positive	9,990 (TP)	990 (FP)	10,980
ELISA-negative	10 (FN)	989,010 (TN)	989,020
TOTAL	10,000	990,000	1,000,000

Figure 1: Distribution of Population According to ELISA Screening Tool

A little more calculation is required to answer the question that is of interest to the patient: *what is the likelihood that a positive reading is false*. The table reveals that 10,980 people tested positive (first row total), and 990 of them were not HIV+. Thus 990/10980, or approximately 9%, of those who tested positive

received false positives. The fact that nearly 10% received a false positive when the test was “99.9% reliable” should be a bit startling.

Similarly, 0.001% received a false negative result. Again, this is quite different from what one might expect (assuming that one would expect $100\% - 99.9\% = .01\%$). This figure is $1/10^{\text{th}}$ of the “anticipated” value).

Calculations like these show that the probabilities for receiving either a false positive or false negative are not obvious. As one begins to investigate the calculations more carefully, one sees that the reason for the high probability of a false positive reading is directly related to the fact that the preponderance of the population in the above example is HIV–; therefore, relatively speaking, there are very few positive results. Thus the few *false* positives among them can make for a large percentage.

In a classroom setting it is very important for students to wrestle with the process of filling out a table like the one in Figure 1. Once students become familiar with these types of calculations, they can investigate other situations by using the same test on a population with a much different prevalence of HIV in the population. By keeping track of the probabilities for the two different false readings, one can begin to understand how greatly they are affected by that underlying factor – the prevalence of the ailment that is being tested for. Figure 2 shows these relationships in a situation where a screening tool has a sensitivity of 90% and a specificity of 85%. Related File 3 is an Excel tool to help with the calculations and graphs presented in this example.

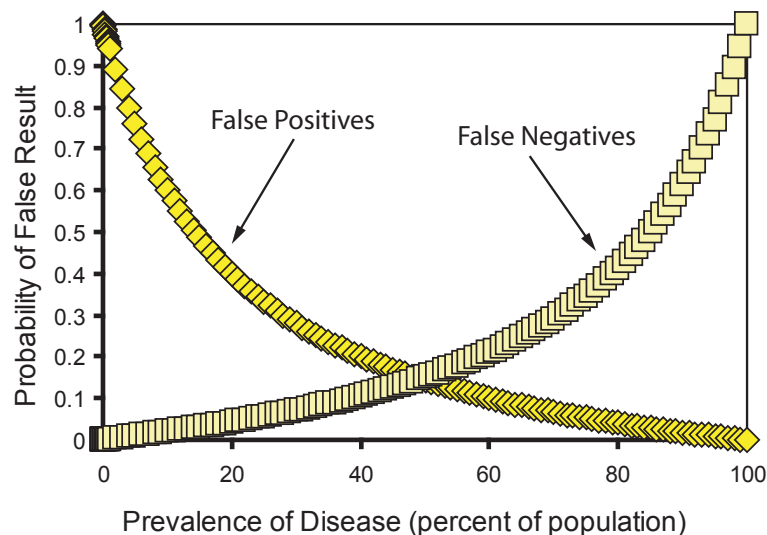


Figure 2: The relationship between the probability of receiving false results and the prevalence of the disease within the population. These values assume a sensitivity of 90% and a specificity of 85% and were calculated with the Excel tool using a prevalence increment of 0.001 from 0 to 0.01, an increment of 0.01 from 0.01 to 0.1, and an increment of 0.1 from 0.01 to 1.

Simply knowing how “reliable” a medical screening tool is (its specificity and sensitivity) does not give a patient the more-relevant information: *How likely is it that I do or do not have this ailment, given the result of the test?* As the graph shows, the probability of receiving a false positive or false negative result depends on the prevalence of the ailment in society.¹ Patients need to be armed with this concept when faced with making decisions to pursue costly and sometimes dangerous medical procedures.

An example like this one from MUSCDC (2007) illustrates how some simple mathematics – percents and a basic concept of probability – can shed light on the interpretation of screening tests. However, such examples also bring up another important question: *Which patients are the “true” positives as opposed to the “false” positives?* Without knowing this, one does not really have information about the sensitivity and specificity of the tool.

An Extended Look at Medical Screening

Consider the basic problem of how one might determine the sensitivity and specificity of a particular screening tool. One might first screen n patients and see how many positives and negatives the screening tool identified. But which of the positive readings are true positives? Which of the negative readings are true negatives?

After the initial screening, some patients may be referred by a doctor for a more-definitive diagnostic test. While the more-diagnostic test may have the attraction of being completely conclusive (i.e. it produces no false readings), it could have the downside of being very invasive, expensive, time consuming, dangerous, or some combination of these negatives. As it is neither feasible nor ethical to submit all patients to invasive testing for definitive confirmation of the presence or absence of a disease, doctors send only some patients for these further tests. Because it is the second test which is considered infallible, only those patients who undergo both tests contribute information that can be used to calculate the sensitivity and specificity of the original screening tool. This creates what is known as *referral* or *verification bias* in the medical literature.

A concrete example involves Coronary Artery Disease (CAD) which is the result of various plaques forming on the walls of the arteries that supply the heart

¹ Exploring the relationships between the probabilities of false readings, prevalence of disease, sensitivity, and specificity can also make for a very rewarding classroom activity or extended student project. One can use the Excel tool of Related File 3 to easily generate graphs similar to Figure 2 or, for a more advanced course, analyze the underlying conditional probabilities. Additionally, studying the effect of repeated iterations of a drug test (see Lyublinskaya 2005) is an interesting mathematical experience.

with oxygen and other nutrients. CAD can be difficult to detect in its early stages and often leads to sudden heart attacks. Exercise stress testing can be used as a diagnostic tool in helping to detect the presence of CAD. This non-invasive procedure can help doctors identify those patients who should be referred for more-definitive testing. The following example is based extensively on Danias and Parker (2002).

We assume that exercise stress testing has a specificity of 85% and a sensitivity of 75%. For the following calculations, assume additionally that we are studying a population of 10,000 patients of whom 25% have CAD. One can now produce the two-way table in a manner similar to that for the HIV example. This information is displayed in the left-hand table in Figure 3. For the sake of brevity, the row and column totals have been omitted.

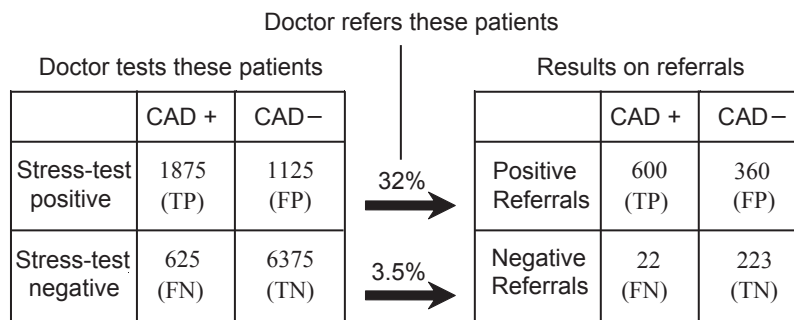


Figure 3: Tracking patient test results from a screening tool to a definitive test.

Now, assume that 32% of the 3000 patients who had positive stress-test results are referred to the hospital for the invasive coronary angiography (which, recall, we assume to be 100% reliable). Further, assume 3.5% of the 7000 patients who had negative stress-test results are also referred for coronary angiography. We will refer to the 32% and the 3.5% as the positive and negative referral rates respectively. As mentioned earlier, there are many reasons why all patients are not subjected to the definitive test. Also, the stress test itself may not, in reality, be a simple binary (positive/negative) test. Rather, results of stress testing may include a range of “scores” which help a doctor to decide if further testing is warranted (see Danias and Parker 2002 for a discussion). In order to investigate referral bias in a simple setting, we will assume that those referred for coronary angiography are a random sample of the patients in the study. That is, it is assumed that doctors typically refer 32% of patients who receive a positive

stress test (regardless of actual presence or absence of CAD). That means that 32% of the 1875 TPs and 32% of the 1125% of the FPs are referred for coronary angiography, where they are correctly diagnosed. Similarly, 3.5% of the 625 FNs and 3.5% of the TNs are referred.

A second two-way table can now be filled out to show the number of patients who were eventually definitively diagnosed as having or not having CAD (right-hand table in Figure 3). It should be emphasized at this point that only the values in the right-hand table are actually known (The numbers TP and FP in the left-hand table are unknown, because the true sensitivity and specificity are not known.). Focusing on the right-hand table, there were 622 patients who were eventually diagnosed with CAD (first column total). Six hundred of these actually received a positive indication based on the stress testing. So the *apparent sensitivity* of the screening tool would be $600/622 \approx 96\%$. This *apparent sensitivity* value is quite a bit higher than the true sensitivity of 75%. Similarly, the data from the right-hand table yield an *apparent specificity* of 38%, whereas the true specificity was assumed to be 85%.

The Excel tool in the related file can be used to change the various parameters involved in this example as well. It turns out that the discrepancy between the true and apparent sensitivities and specificities does not depend on the prevalence of the disease in the population, but rather on the referral rates of the doctors. More specifically, when the two referral rates are equal there is no difference between the true and apparent values. As seen in this example, when the positive referral rate is larger than the negative referral rate the true sensitivity is less than the apparent sensitivity while the true specificity is larger than the apparent specificity. Conversely, when the positive referral rate is lower than the negative referral rate these relationships are reversed. Thus, the medical practice of referrals can greatly affect the perceived reliability of a medical screening tool.

In our classroom activity (Related File 1), we end with an exercise which has students devising methods to correct for this referral bias. That is, we give the students information in the form of the right-hand table (so they have information regarding the *apparent sensitivity* and *apparent specificity*) and they are asked to “backout” the true sensitivity and specificity for the original screening tool.

Clearly, medical-testing and screening-tool analysis is a complicated field, and the subject is fraught with moral and ethical complications. Regrettably, not all aspects can be touched on here. Referral bias can certainly be avoided if a random sample of patients is subjected to both the diagnostic screening tool and the definitive test. In many cases this may be feasible and desirable. In others, the inherent dangers of the definitive test may make this approach less attractive. In fact, by understanding this simplified example, one can begin to construct a mathematical framework to correct for referral bias, thus eliminating the need for random medical testing (see Punglia et al. 2003).

The Classroom Activity

The related files for this paper include the classroom-ready activity and a detailed set of notes for instructors. We believe that the activity is suitable for students at both the high school and college level. High school students who have been successful in a course such as Algebra II or college students in a quantitative reasoning course (e.g., a “Math for Liberal Arts” course) should be adequately challenged by this activity. The mathematics is as simple as that used in the two examples above, but some students may be challenged by the reading and critical reasoning skills demanded in this context-rich environment.

The first part of the activity centers on the relationship between the likelihood of receiving a false positive diagnosis and the prevalence of the ailment in the studied population. The context for these calculations is the effectiveness of a new ELISA to identify the presence of HIV – how does it compare with a “gold standard” test, the Western Blot? The second part of the activity uses the CAD example we have discussed to investigate the phenomenon of referral bias. The activity has been reviewed by 12 high school and college mathematics teachers and class-tested in three college classes with a total of 92 students. Plans are currently underway to test this activity in a high school classroom as well.

Student Learning

Three quantitative reasoning classes with two different instructors at Central Washington University used this activity in the classroom. Each implementation used approximately two class periods (100 minutes). Students worked on the activity in groups in two of the classes, and the third worked on it as an instructor-facilitated class.

An identical assessment instrument was administered before and after students completed the activity. This instrument assessed students’

- ability to define the terms “true positive,” “true negative,” “false positive,” and “false negative,”
- knowledge of the likelihood of receiving a false positive on a very reliable diagnostic test,
- ability to define the terms “specificity” and “sensitivity”, and
- ability to calculate various percentages from information organized in a two-way table.

Two classes (henceforth called “Group A”) had been exposed to a discussion of false positives (at a level similar to that found in Bennett and Briggs 2005) while one class (Group B) had had no prior exposure to this topic.

In the pre-test, 72% of students in Group A could correctly define the term “true positive” as compared with 5% of students in Group B. Interestingly, more students could correctly define “false positive” (83% of Group A, and 45% of Group B) than could define “true positive.” We conjecture that the ordering of the items on the pre-test could have influenced this phenomenon. Students were first asked to define “true positive” and then to define “false positive.”

Very few students (about 1%) showed any prior knowledge of the specialized vocabulary “sensitivity” and “specificity.” About 35% of the students were able to correctly recall the definitions on the post-test.

Two items on the assessment instrument asked students to calculate various percentages from information in a two-way table (much like in Table 1). In Group A, about 70% of students who could not correctly accomplish this task on the pre-test could perform the calculation on the post-test. For Group B only about 30% of students who could not do this on the pre-test showed improvement on the post-test. We attribute the difference here to the fact that Group A had worked with two-way tables earlier in the course.

One item on the instrument was designed to test students’ knowledge of the fact that the probability of a false positive cannot be deduced from the “effectiveness” of the test:

“The medical test to diagnose reptile virus is found to be 99% effective. You have just received a positive test result. How likely is it that you have the disease?”

A correct response would indicate that there was not enough information present to answer the question. Specifically, one would need to know the prevalence of reptile virus in the population. Unfortunately, very few students gave such responses. In Group B, no students answered correctly on either the pre- or post-test. In Group A, three students (5.5%) answered correctly on both the pre- and post-test while six students (11%) answered incorrectly on the pre-test and correctly on the post-test. However, many students (both groups) gave responses such as “don’t know!” While such a response was scored as incorrect, there is the possibility that some students used this response in place of a response such as “not enough information”. This item will be altered in future assessments to help overcome this possible ambiguity in student responses.

The student assessment thus provided the following information:

- Students may have greater familiarity with a term such as “false positive” as opposed to “true positive.”
- Students in Group A who spent more individual time on the computations out-performed those in Group B (on the computational part of the post-test) who had a more passive role in the activity.

- Having students just work through the activity is not enough. The instructor needs to spend some time emphasizing the main points and reminding students of the importance of the calculations they have been performing..

Teacher Reaction

Six high school teachers, three instructors from two-year colleges, and three instructors from four-year colleges read this activity. Nine of these 12 teachers felt the activity was easy for them to understand, but felt that their students would not easily understand it without additional guidance from the teacher. In particular, many teachers felt that students would have difficulty with the section on referral bias if no teacher support was given. Teachers should be forewarned that students will need extra guidance and motivation for this part of the activity. We feel that this section should not be omitted as it provides students with an actual example from the medical literature. Eight of the teachers felt that they could use the activity in an existing course and all but one of the teachers indicated that they learned something new by reading through the activity.

As has been noted, two instructors at Central Washington University piloted the activity. In this section, we will refer to them as Instructor A and Instructor B (who taught students from Group A and Group B respectively). Instructor A was quite familiar with the two-way tables and the topic of medical testing after teaching a quantitative literacy mathematics course several times in the past and was, therefore, at ease teaching this lesson; however, Instructor A had not examined the effect of the prevalence of the disease in the population on the sensitivity and specificity of medical tests and, in fact, was not familiar with these terms. Instructor A directed the class to work in groups until the section about referral bias. Instructor B was quite inexperienced with both the tables and the topic and was apprehensive about teaching the lesson. Instructor B led the students through all parts of the activity as an entire group much like a typical lecture.

Both instructors were surprised by the results of the pre- and post-tests. They both believed that they might have missed some key points in the lesson. Neither instructor had experience with comparing pre- and post-test results for a lesson and welcomed the opportunity to examine what the students had learned, or did not learn. The instructors felt that students may have thought that the material was not important to remember because their work was not graded; they both felt that they would have seen better results on the post-test if students were somehow made more accountable for the material. In addition, both instructors felt that additional practice problems would have benefited the students.

After piloting the materials, one instructor proposed allotting three, or possibly four, days to the lesson. Here is a brief outline of a proposed schedule for the lesson.

Day 1:

1. Students take the pre-test.
2. Students perform a short skit similar to the fictional patient/doctor dialog above.
3. The teacher leads students through Part 1.
4. In groups, students do a practice problem like that in Part 1.

Day 2:

1. The teacher reviews the percents and results from Part 1.
2. The teacher leads students through Parts 2 and 3 and compares the results to Part 1 results.
3. In groups, students do some practice problems like those in Parts 2 and 3.

Day 3:

1. The teacher reviews the results of Parts 1 through 3 and discusses the significance to society.
2. The teacher leads students through Part 4, Case 1.
3. In groups, students do a practice problem similar to Part 4, Case 1, and then also do Part 4, Case 2.

In addition, the instructors felt that a graded writing assignment in which the students discuss the results and implications may lead to deeper understanding. Both instructors thought that most students appreciated the material on false positives and its relevance to their lives. The instructors both would like to do the activity again, but probably at a different point in the quarter.

Conclusion

We have presented an activity, together with student and teacher assessment, that can be used in a quantitative-literacy mathematics course for high school and college students. Quantitative literacy courses are being designed to help our future citizens become informed consumers of statistical and numerical data in hopes that they can make judgments that will better their lives and the course of society as a whole. From this experience, we have concluded – and the students concur – that the subject of false positives in medical testing is an appropriate subject to include in such a course .

Acknowledgments: The authors would like to thank Richard Trudgeon and teachers from the Yakima Valley Project (Ryan Downing, Lisa Frazier, Doug Lewis, Jahan Lohrasbi, George Lopez, Wanda Merz, Jamie Nordstrom, and

Tamara Wiley,) for helping in the assessment of the class activity. There were also many reviewer comments which helped form the final draft of this manuscript. Portions of this work have been funded by the Transition Math Project and Central Washington University.

References

- ABC News. 2006. Landis' Doctor: Test is 'False Positive'. [abcnews.go.com] July 29, 2006.
- Allen, S. 2007. Cancer scares grow as screening rises. *The Boston Globe* [boston.com], September 30, 2007.
- Bennett, J. and W. Briggs. 2005. *Using and Understanding Mathematics: A Quantitative Reasoning Approach*. 3rd ed. Boston: Pearson Addison Wesley.
- Danias, P.G. and J. A. Parker. 2002. Novel Internet-based Tool for Correcting Apparent Sensitivity and Specificity of Diagnostic Tests to Adjust for Referral (Verification) Bias. *Radiographics*. 22: e4-e4. <http://radiographics.rsna.org/cgi/content/abstract/22/2/e4>
- Garfunkel, S. 1998. *Mathematics: Modeling Our World*. Cincinnati, OH.: South-Western Educational Publishing.
- Guthrie, J. 2004. False diagnosis of HIV discovered after 8 years. Veteran's life severely affected after VA doctor made mistake. *San Francisco Chronicle*, August 28, 2004.
- Lyublinskaya, I. 2005. How Fair Is the Drug Test? *Mathematics Teacher*. 98 (8): 536.
- Madison, B.L. and L.A. Steen. 2008. Evolution of Numeracy and the National Numeracy Network. *Numeracy*. 1(1), Art. 2. <http://services.bepress.com/numeracy/vol1/iss1/art2/>
- Medical University of South Carolina Doctoring Curriculum (MUSCDC). 2007. Practical examples using sensitivity, specificity, gold (reference) standard, positive predictive value, and negative predictive value. MUSCDC., Medical University of South Carolina Doctoring Curriculum. [<http://www.musc.edu/dc/icrebm/sensitivity.html>]
- Memory, J.D. 2007. Why Richard Cory Offered Himself. *Mathematics Magazine*. 80(4):273.
- Punglia, R.S, A.V. D'Amico, W.J. Catalona, K.A. Roehl, and K.M. Kuntz. 2003. Effect of verification bias on screening for prostate cancer by measurement of prostate-specific antigen. *The New England Journal of Medicine*, 349 (4): 335-342.
- Sevilla, A. and K. Somers. 2007. *Quantitative Reasoning: Tools For Today's Informed Citizen*. Emeryville, CA: Key College Publishing.