What Your Future Doctor Should Know about Statistics: Must-include Topics for Introductory Undergraduate Biostatistics

Brigitte Baldi[1] and Jessica Utts[1]
[1]Department of Statistics, University of California at Irvine, Irvine, CA 92697

Abstract
It's possible that your future doctor will take only one statistics course in her or his life, and that it will be an introductory course for undergraduate students planning a career in the health sciences. Therefore, it is important that we cover certain essential topics in that course, which may not be covered in the more general introductory statistics course. In selecting and presenting such topics, we should bear in mind that doctors also need to communicate probabilistic concepts of risks and benefits to patients who are increasingly expected to be active participants in their own health care choices despite having no training in medicine or statistics. It's also important that interesting and relevant examples accompany the presentation, because the examples (rather than the details) are what students tend to retain years later.

Here we present a list of topics we cover in the introductory biostatistics course that may not be covered in the general introductory course. We also provide some of our favorite examples for discussing these topics.

Key Words: Statistics education, biostatistics, undergraduate

1. Medical Schools and Statistics: a Tenuous Relationship

1.1 Statistics in Medical Schools' Curricula: from Absent to Required

The United States Medical Licensing Examination [1] assesses the clinical skills of medical school students as well as their knowledge and understanding of basic science concepts relevant to the practice of medicine. Statistics does not figure prominently in this exam. In fact, the exam's general description only mentions the requirements that examinees be able to "interpret graphic and tabular material." The more detailed exam description shows that "Quantitative Methods" is one of nine topics listed under "General Principles," which is one of twelve main topics in the first step (out of three steps) of the board exam [2]!

So perhaps it is not surprising that probability and statistics do not feature prominently in American medical schools. The Association of American Medical Colleges provides a list of American Medical Schools along with links to their websites [3]. A quick (non-exhaustive) search of these medical schools revealed very different approaches to the treatment of probability and statistics in their curricula. Some medical schools offer dedicated courses (for example, Albany Medical College, The Geisel School of Medicine at Dartmouth, University of California, San Diego School of Medicine) while others do not even mention the topic at all (for example, Columbia University College of Physicians and Surgeons, Emory University School of Medicine, University of California, Irvine School of Medicine).

1.2 Statistics in Admission Requirements: a Shifting Paradigm

A quick search of admission requirements for American medical schools revealed a similar mosaic of approaches to statistics education. Some schools do not mention statistics at all (for example, Albany Medical College, Boston University School of Medicine, Emory University School of Medicine, Mayo Medical School) while others recommend but do not request an undergraduate statistics course (Loma Linda University School of Medicine, University of Massachusetts Medical School, University of Michigan Medical School). The Medical College Admission Test (MCAT) has limited emphasis on statistics (its Physical and Biological Sciences Cognitive Skills section does include questions related to scientific hypotheses, data collection, and data interpretation). Some medical schools, however, require that students seeking admission have mathematical training which may be in the form of a statistics course (for example, The Geisel School of Medicine at Dartmouth, Georgetown University School of Medicine, Harvard Medical School, University of California, San Diego School of Medicine).

Interestingly, Harvard Medical School changed its course requirement options for students applying to enter in 2011 or beyond [4]. Specifically, the school acknowledges that computational skills and quantitative reasoning are required for contemporary scientific literacy but are not served adequately by a purely mathematical approach (calculus). The school now expects some statistics foundation in the applicant's undergraduate education, but gives flexibility in the actual method of instruction, encouraging courses that blend biology or health-related content with statistics coverage. The University of Massachusetts Medical School likewise recommends an undergraduate course in statistics because of the increasing emphasis on evidence-based medicine [5].

1.3 Undergraduate Introductory Statistics: Options

The increased emphasis on evidence-based medicine creates a greater need for educating future physicians in the general domain of quantitative reasoning, probability, and statistics. Perhaps we will soon see more medical schools requiring at least one undergraduate course in introductory statistics. But would such a course truly serve the specific needs of practicing physicians? Statistics is a field with vast applications and introductory statistics courses are likely to reflect this (with examples ranging from economics, engineering, life sciences, social sciences, to general interest surveys, along with the seemingly compulsory gambling examples).

An introductory biostatistics course could better serve pre-med students by ensuring that concepts key to their future careers are discussed. To keep such a course equivalent to a general introductory course, health-related concepts must be covered without redesigning radically the statistics content of the course. One simple approach relies on a strategic choice of in-class examples so that the health content can be discussed while teaching the statistics content, rather than in addition to it. This has the added advantage that students tend to remember interesting examples far longer than procedural details. In this article, we showcase some examples that can be used in class to serve this dual purpose.

2. Probability and Continuous Variables for Disease Definition

2.1 Natural Variability

An important medical issue is how we define a particular disease or medical condition. Some may be obvious enough to define by their presence or absence (a wound has broken the skin or not, a bone is fractured or not, a biopsy shows cancerous cells or not).[1] Yet others may be in a grey area because the variable used to distinguish the healthy from the pathological is continuous. How high is too high for blood pressure or total cholesterol? How long is too long for gestation? How low is too low for bone mineral density? And now that the American Medical Association has recognized obesity as a disease [6], we ask how high is too high for a person's body mass index.

Every introductory statistics course will cover the concept of continuous random variables. Utilizing examples of variables used in disease definition helps future doctors understand that natural variation exists, both within individuals and within populations. It is also an opportunity to discuss how scale matters and how threshold values are inherently arbitrary to some extent.

2.2 Implications of a Chosen Threshold

Disease definitions change at times, and sometimes new categories or conditions get added. For example, current medical guidelines define high cholesterol, a risk factor for heart disease, as total blood cholesterol values of 240 mg/dl or above. But a category defined as "elevated" or "borderline high" cholesterol represents total cholesterol values between 200 and 240 mg/dl [7]. We may ask how common such values are. All introductory statistics courses cover normal distributions. Let's model total cholesterol in a population of middle-aged men using a normal distribution with mean 222 mg and standard deviation 37 mg [8]. This distribution is shown in Figure 1, with the areas representing high and borderline high cholesterol values highlighted. Normal calculations show that approximately 31% of this population has high cholesterol and another 41% has borderline high cholesterol. Together, the two groups make up 72% of this population, or nearly three-quarters. The extra padding, while at first glance moderate and reasonable, has a tremendous impact on the population targeted for treatment.

A recent NPR article titled "Why Do People Still Die of Heart Disease?" [9] discussed the progress made treating heart disease, especially through the use of a class of cholesterol-lowering drugs called statins. When asked if this drug class, the most prescribed in the United States, should be given to so many people, the invited expert answered that "these drugs have been so effective that some have advocated giving them to virtually everybody over the age of 45 or 50." This is not actually that far from our computation of 72% of middle-aged men. However, the idea is a radical rethinking of medicine. It redefines age as a pathology in-and-of-itself, and it is a departure from medicine's core ethical value of doing no harm.

The consequences of shifting the threshold for treatment based on a continuous random variable can be illustrated with other examples. For instance, osteoporosis is defined as very low bone densities, lower than 2.5 standard deviations below the mean bone density of young adults [10]. Among older women, modeling bone density using a normal distribution with mean -2 and standard deviation 1 [11] shows that osteoporosis afflicts

[1] Although, of course, the severity of the condition is rarely a black-and-white issue.

approximately 31% of this population, as seen in Figure 2. Adding a new category called "low bone mass" or "osteopenia" with values between -2.5 and -1 adds another 53% to be considered for medical treatment. We may take for granted that elderly women are fragile individuals, and even not question the soundness of treating 84% of them. But how would the same guidelines translate to the reference population of young adults, with mean 0 and standard deviation 1? A quick back-of-the-envelope calculation to illustrate in class the use of the 68-95-99.7% rule tells us that we would consider roughly 16% of all young adults to have bone densities so low that they would need medication. Undergraduate students hoping to work in healthcare should be encouraged to ponder such a conclusion.

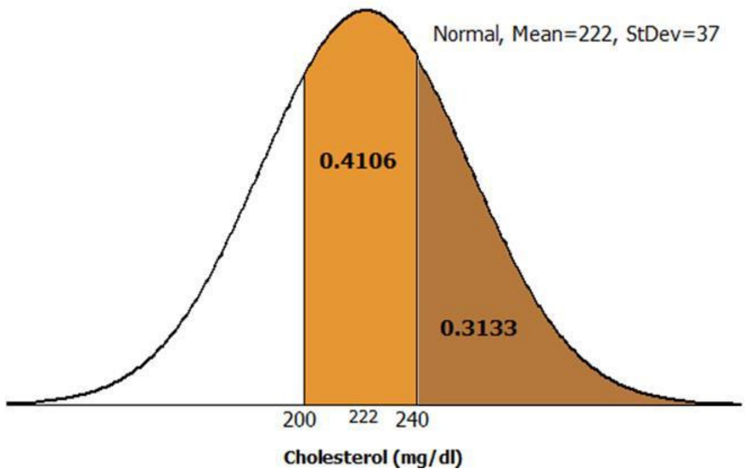

Figure 1: Distribution of total blood cholesterol levels among middle-aged men, modeled using a normal curve with mean 222 and standard deviation 37 mg/dl. The shaded areas represent elevated (borderline high) cholesterol levels (200 to 240 mg/dl, 41%) and high cholesterol levels (240 mg/dl and above, 31%).

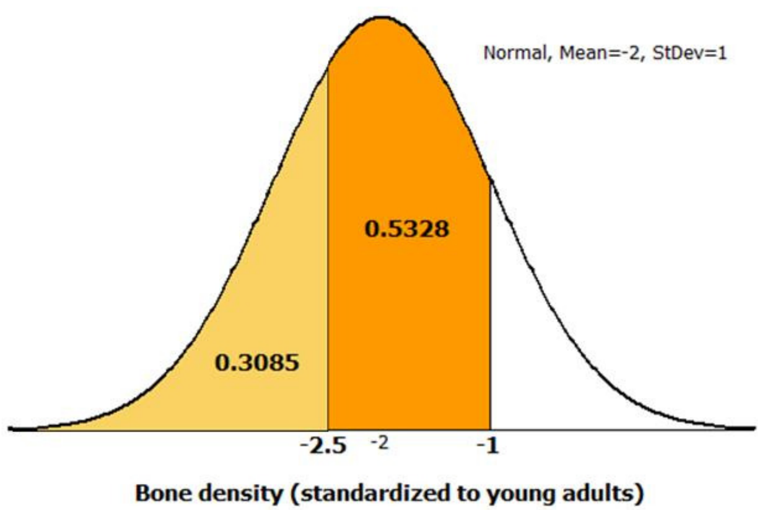

Figure 2: Distribution of standardized bone densities among elderly women, modeled using a normal curve with mean -2 and standard deviation 1. The shaded areas represent osteoporosis (values below -2.5, 31%) and osteopenia (values between -2.5 and -1, 53%).

3. Conditional Probabilities and Diagnostic Tests

Not all introductory statistics courses cover conditional probabilities, in part for time constraints, and in part because students struggle greatly with conditional probabilities, and a deep understanding of conditional probabilities is not strictly necessary to understand basic statistical inference. Leaving out conditional probabilities can be a reasonable option for a general student audience, but it would be a serious omission when serving undergraduate students planning a career in health sciences. In particular, these future professionals should gain an intuitive understanding of diagnostic and screening tests.

3.1 Factors Affecting the Positive Predictive Value

Studies have shown that doctors often misinterpret the nature of the information provided by diagnostic and screening tests, even though they know the definitions of a test's sensitivity, specificity, and positive predictive value (PPV) [12]. Yet, this understanding is at the core of recent changes in health policies. For instance, up until the summer of 2012, the U.S. Food and Drug Administration had refused approval to over-the-counter HIV screening because of concerns with public reaction to false-positives [13, 14]. Inversely, the 2009 recommendation by the U.S. Preventive Services Task Force against routine mammography screening for women in their 40s caused an uproar [15], despite the fact that the PPV in this age group is very low.

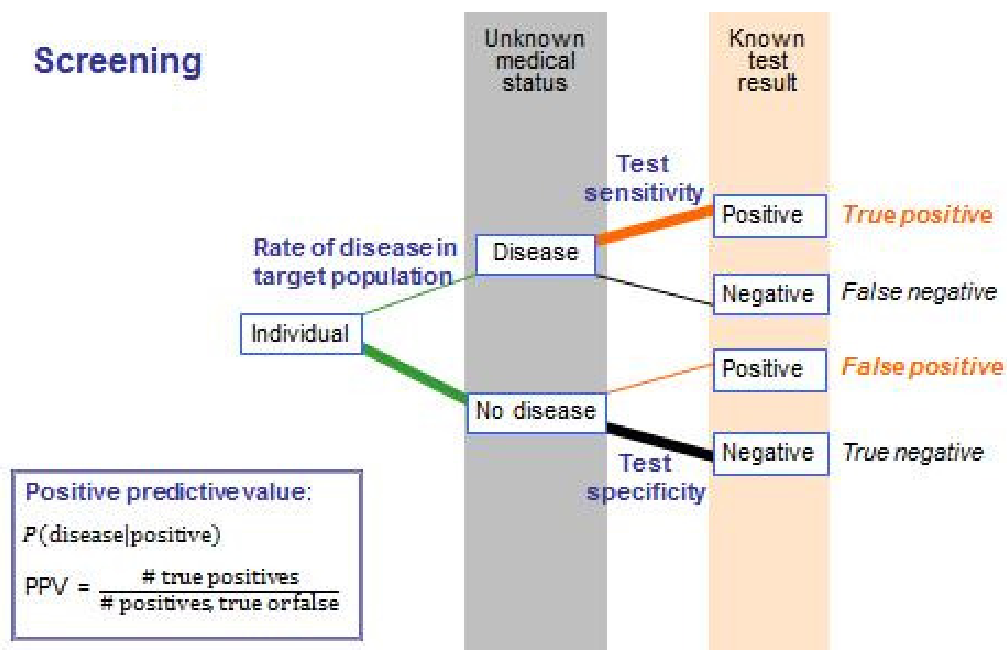

Figure 3: Tree diagram representing a screening test for a medical condition. The three thicker branches show the more likely outcome at each point, when a diagnostic test is used for screening purposes (individuals without the disease, green; sensitivity, orange; specificity, black). The positive predictive value, PPV, is influenced by all three.

Diagnostic tests are a great classroom example to illustrate the use of probability trees and two-way tables to compute conditional probabilities. They also tend to really capture students' attention. Figure 3 shows a generic probability tree in the context of a medical test used for screening purposes. That is, everyone in the target population is supposed to be tested and the disease rate in this target population is relatively low (thank goodness!).

Students should understand that the PPV, which represents the probability that a person receiving a positive test result truly has the disease or condition, depends on a number of factors. Obvious to all is the fact that the test should be reliable and produce correct answers as often as possible (sensitivity and specificity). More often ignored is the importance of the disease rate in the target population. When the population has a very low disease rate, most of the individuals tested will not have the disease and will receive either a true negative or a false positive. This greatly reduces the PPV of a test and, therefore, its usefulness. This means that a screening test might not be very useful if administered to a low-risk population.

3.2 Screening Choices and Public Health

3.2.1 Selecting the target population

The first choice to make is that of the target population intended for the screening test. A simple class example can help students understand some of the reasons behind the recent changes in breast cancer screening guidelines using mammography—while at the same time learning the mechanics of PPV computation. If we consider a sensitivity of about 85% and a specificity of about 95% for mammography in the United States [16], what determines the PPV of screening for breast cancer is the cancer rate in a given age group, as shown in this equation:

$$\mathrm{PPV} = \frac{\mathrm{rate} \times \mathrm{sensitivity}}{(\mathrm{rate} \times \mathrm{sensitivity}) + ((1 - \mathrm{rate}) \times (1 - \mathrm{specificity}))}$$

Using estimates from the National Cancer Institute, we obtain a different PPV for women in their 40s, 50s, and 60s, as shown in Table 1. The PPV for women in their 40s is very low at only 20%, meaning that only 1 in 5 who receive a positive test result from a routine mammography actually has breast cancer. This, and other reasons such as a lack of clear evidence that early detection in this age group actually saves lives on average, is why routine mammography in this age group is not systematically recommended any longer [17].

Table 1: Approximate positive predictive value of mammography for American women of various age groups.

Age group	**Breast cancer rate**	**Positive predictive value, PPV**
40 to 49	1 out of 67 (1.49%)	20%
50 to 59	1 out of 35 (2.86%)	33%
60 to 69	1 out of 28 (3.57%)	39%

3.2.2 Improving the test's performance

Another way to improve the PPV of a screening test is to improve its actual intrinsic performance, that is, its sensitivity and specificity. The equation above shows that the test sensitivity affects both the numerator and the denominator of the PPV, whereas specificity affects only the denominator. Looking back at Figure 3, we can see that a high specificity is particularly important for screening tests because the disease rate for screening tends to be very small. Having a higher specificity means having a lower percent of false positives. This is good for the PPV and it is especially good for patients, who are often considered not sophisticated enough to deal with false negatives. This perception explains why it has taken until 2012 for the U.S. Food and Drug Administration to approve the first over-the-counter HIV testing kit, the OraQuick In-Home HIV Test [14].

The OraQuick test provides an interesting classroom example to illustrate how probability values such as sensitivity and specificity are derived. The findings are based on a large survey of 4410 adults from a high-risk population who did not yet know their HIV status and had it confirmed later with comprehensive further medical examinations [18]. The results are shown in Table 2. OraQuick has an estimated sensitivity of 92.92% and an estimated specificity of 99.98%. This is an extraordinarily high specificity.

Table 2: Outcomes of the OraQuick In-Home HIV Test on a sample of 4410 adults from a high-risk population who did not yet know their HIV status. Each person's actual HIV status was later confirmed via several other medical examinations.

105 true positives (HIV and positive test)	**1 false positive** (no HIV and positive test)
8 false negatives (HIV and negative test)	**4296 true negatives** (no HIV and negative test)

The other merit of this example is that it should help students understand that statistical illiteracy has a high price, both financial and human. The Centers for Disease Control and Prevention estimates that about 25% of all HIV-positive Americans are not aware of their serologic status. This is thought to be one of the primary reasons why new HIV infections remain high in the United States, with about 40,000 new cases every year.

3.3 Communicating Probabilities to Patients

Clearly, doctors must understand diagnostic and screening tests as they are the primary consumers of such medical tests. But they also need to explain a test's objectives and a test result to their patients. If undergraduate students and doctors find conditional probabilities challenging, how do we expect untrained patients to fare? Fortunately, there are alternatives to Bayes' formal theorem that can make conditional probabilities accessible to all. This usually involves discussing outcomes in terms of frequencies rather than probabilities [12].

One Kaiser Permanente patient pamphlet makes a great classroom example to help students who struggle with probabilities. The pamphlet describes the possible outcomes of a PSA test, and is adapted from a brochure by the American Academy of Family Physicians [19]. The PSA test is used to screen older men for prostate cancer. The concept of the graphic is displayed in Figure 4.

What truly puzzles students about this example is just how easy and obvious probability computations are in this format. Out of 100 outcomes, we would expect 10 positives: 3 true positives (green) and 7 false positives (orange). The remaining 90 negatives would break down into 1 false negative (yellow) and 89 true negatives (blue).

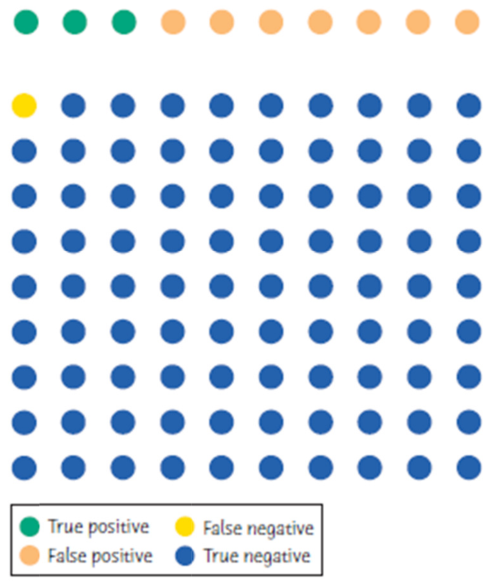

Figure 4: Graphic display of the expected outcomes of 100 PSA tests, adapted from a Kaiser Permanente's pamphlet found amidst various other patient-education material during an office visit. The original graphic was developed by the American Academy of Family Physicians.

From that we easily compute:

PSA sensitivity = P(positive | cancer) = 3/4 = 75%
PSA specificity= P(negative | no cancer) = 89/96 $\approx$ 92.7%

Rate of prostate cancer in target population = P(cancer) = 4/100 = 4%
PPV of PSA test in target population = P(cancer | positive) = 3/10 = 30%

That is, some men screened for prostate cancer with the PSA test will get a positive test result, but patients can see that only 30% of men with a positive test result actually have prostate cancer (and 70% don't). The American Urological Association (AUA) no longer recommends routine PSA testing for prostate cancer for men without special risk factors and younger than 55 or older than 69. For men ages 55 to 69, the AUA strongly recommends a personalized discussion between patients and doctors to decide for or against screening, in part because of the low PPV of the PSA test and in part because of the potential harms associated with treatment [20].

4. Statistical Versus Practical Significance and Patient Communication

4.1 Numbers in Context

Why is the PSA test not systematically recommended for routine prostate cancer screening, but mammography for routine breast cancer screening is still recommended for some age groups? After all, the PPVs for both tests are similar. And, if anything, mammography has more direct potential for harm, through the use of radiation, than the PSA test, which requires only a blood sample. These two examples offer a great opportunity to remind students that numbers truly have meaning only when they are placed within a context.

Mammography and PSA may have similar PPVs, but the point of a screening test is what will be done with the information it provides. It turns out that prostate and breast cancers have very different natural prognoses and treatment efficacies. The PSA test is of questionable use in part because treatment for prostate cancer has limited survival benefits on average (most older men die of other causes and would not have noticed their prostate cancer without the screening), but the potentials for harm are many, substantial, and relatively likely [19, 20]. In the end, treatment may provide a worse outcome than no treatment. Table 3 shows some of the factors that the AUA suggests physicians discuss with their patients before deciding on a PSA test. Mammography, on the other hand, appears to have more benefits, especially in some age groups [16, 17, 21].

Table 3: Possible outcomes of treatment for prostate cancer. Information reproduced from a 2003 American Academy of Family Physicians brochure.

	Radiation	**Surgery**
Improved survival	*Unknown*	*Unknown*
Death from treatment	2 in 1000	1 in 200 younger men 1 to 3 in 100 older men
Impotence (difficulty with erection)	40 in 100	30* to 90 in 100 *nerve sparing surgery
Any Incontinence (loss of urine control)	60 in 100	32 in 100
Complete Incontinence (lose complete control of urine)	1 in 100	7 in 100
Urinary Stricture (makes it difficult to urinate)	5 in 100	12-20 in 100
Any rectal Injury (discomfort/trouble with bowel movements)	11 in 100	30 in 100

4.2 Looking Beyond the *P*-value

Numbers are almost meaningless without context, and this is especially salient in the introductory statistics curriculum when discussing *P*-values. Instructors take great pains to explain that statistical significance is not the same as practical significance or importance. Again, it is possible to drive this statistical point while at the same time introducing important healthcare concepts.

The NPR article cited in section 2.2 also discusses the efficacy of cholesterol-lowering drugs [9]. An invited expert says, “We have high quality clinical research trials involving several hundred thousand patients. And in most high-risk populations, they lower the risk of a heart attack, stroke or death by anywhere from 25 to 35 percent.” That sounds pretty impressive, but is it? And what does it mean to a lay person?

Here is one example comparing the use of the cholesterol-lowering drug gemfibrozil with a placebo. Middle-aged men with elevated cholesterol levels were randomly assigned to take gemfibrozil or a placebo daily for 5 years. The researchers recorded the number of individuals experiencing a fatal or nonfatal heart attack over the whole study period [22]. The results are shown in Table 4, and are statistically significant ($X^2 = 6.1$, $P = 0.014$, or a two-sample z test). Qualitatively, at a level suitable for an undergraduate introductory statistics course, the findings can be summarized as follows:

> (a) Gemfibrozil led to a 34% reduction of the risk of heart attack (relative risk reduction, RRR). That is, the group taking gemfibrozil had 34% fewer heart attacks, fatal or not, than the placebo group in the 5-year treatment period.
> (b) The reduction in heart attacks from taking gemfibrozil corresponded to 1.4 percentage points (absolute risk reduction, ARR), or 1.4 heart attacks (fatal or not) prevented for every 100 patients treated over five years.
> (c) 71 patients needed to take gemfibrozil daily for five years to prevent 1 heart attack, fatal or not (number needed to treat, NNT).

These three statements focus on the negative outcome, heart attack. The information can also be reframed to say that taking gemfibrozil daily for five years increased the probability of not having a heart attack from 95.9% to 97.3%. All four summaries reflect the same statistically significant data, but they give very different impressions of gemfibrozil's efficacy.

Table 4: Does the cholesterol-lowering drug gemfibrozil help reduce the risk of heart attack? A randomized experiment compared the number of patients with a heart attack over a 5-year period for middle-aged men assigned to either the drug or a placebo.

	Heart attack	**No Heart Attack**	***n***	$\hat{p}$
Gemfibrozil	56	1995	2051	.0273 (2.73%)
Placebo	84	1946	2030	.0414 (4.14%)

Historically, the relative risk reduction has been most prominent. However, there is a growing recognition of the need to provide a more comprehensive and balanced description of health risks. The website www.thennt.com is a great source of examples that can help you contrast treatments by comparing their benefits and harms, in both percentage and number form. Our students, and future healthcare professionals, need to understand these issues at the core of evidence-based medicine, not only to enable them to make the best medical decision for their patients, but also to inform their patients in ways that are easy to understand.

5. Conclusion

There are many other examples that can be used during class to serve a dual purpose of basic statistics education and illustration of specific healthcare topics. For example, when introducing data acquisition and experimental design, one can bring up the increased emphasis on evidence-based medicine—with counterexamples. In the distant and the recent past, a number of studies have eventually shown that treatments regularly used on patients have no more benefit than a placebo, more potential for harm, and substantially higher cost. Here are just a few: mammary artery ligation [23] and percutaneous myocardial laser revascularization [24] for the treatment of angina, lavage and arthroscopic débridement for osteoarthritis of the knee [25], and a recent series of multi-million dollar proton therapy centers to treat a whole range of cancers at twice the cost without evidence of added benefits [26].

Students will be exposed to all these concepts in medical school, and likely with a more in-depth treatment. But if an undergraduate introductory statistics course is the only comprehensive view of probability and statistics they will get, we need to make sure that the examples we use are interesting and relevant enough that students remember them long after finals week.

Specifically, we should aim to teach undergraduate students heading for medical and public health professions:

How probability and statistics impact medical issues
How to explain probability and statistics concepts to patients
How to evaluate scientific/statistical evidence critically

We have found that this can be done reasonably well by selecting concrete examples that lend themselves to discussing these important concepts. The actual statistical content of the introductory statistics course does not need to be altered substantially—if at all. However, to cover a wide range of important health-related topics, it might be necessary to offer a dedicated introductory biostatistics course for pre-med undergraduate students.

References

[1] USMLE, www.usmle.org.
[2] www.usmle.org/pdfs/step-1/2013content_step1.pdf.
[3] AAMC, www.aamc.org/students/applying/requirements/msar/toolkit/351636/requiredpremedicalcourseworkandcompetencies.html.
[4] hms.harvard.edu/departments/admissions/applying/requirements-admission.
[5] www.umassmed.edu/som/admissions/reqs.aspx.
[6] AMA Adopts New Policies on Second Day of Voting at Annual Meeting: Obesity as a Disease, June 18, 2013.
[7] A Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III).
[8] National Center for Health Statistics, *Health, United States, 1995*, 1996.
[9] July 27, 2013, www.npr.org/templates/story/story.php?storyId=206034807.
[10] WHO Scientific Group on the Assessment of Osteoporosis at the Primary Health Care Level, 2004.
[11] E. S. Siris et al., "Identification and fracture outcomes of undiagnosed low bone mineral density in postmenopausal women—results from the National Osteoporosis Risk Assessment," *Journal of the American Medical Association*, 286 (2001), pp. 2815—2822.
[12] G. Gigerenzer and A. Edwards, "Simple tools for understanding risks: from innumeracy to insight," *British Medical Journal*, 327 (2003), pp. 741—744.
[13] A. A. Wright and Ingrid T. Katz, "Home testing for HIV," *New England Journal of Medicine*, 354 (2006), pp. 437—440.
[14] "FDA approves first over-the-counter home use HIV test kit," FDA News Release, July 3, 2012.
[15] "Breast guidelines test American tolerance for risk," *Reuters*, November 20, 2009.
[16] U.S. Preventive Services Task Force, "Screening for breast cancer: recommendations and rationale," February 2002, Agency for Healthcare Research and Quality, www.ahrq.gov.

[17] J. S. Mandelblatt et al., “Effects of Mammography Screening Under Different Screening Schedules: Model Estimates of Potential Benefits and Harms,” *Annals of Internal Medicine*, 151 (2009), pp. 738—747.

[18] OraQuick In-Home HIV Test customer package insert information, downloaded from www.orasure.com.

[19] “Should you get a PSA test?,” 2006 Kaiser Permanente pamphlet based on “Should you get a PSA test? A patient-doctor decision,” available from the American Academy of Family Physicians at www.aafp.org. Copyright © 2003 American Academy of Family Physicians. Permission is granted to reproduce this material for nonprofit educational uses.

[20] H. Ballentine Carter et al., “Early Detection of Prostate Cancer: AUA Guidelines,” *American Urological Association Education and Research*, 2013.

[21] www.cancer.gov/cancertopics/pdq/screening/breast/healthprofessional, National Cancer Institute, Physican Data Querry.

[22] M. H. Frick et al., “Helsinki Heart Study: primary-prevention trial with gemfibrozil in middle-aged men with dyslipidemia. Safety of treatment, changes in risk factors, and incidence of coronary heart disease,” *New England Journal of Medicine*, 317 (1987), pp. 1237—1245.

[23] For a review of several randomized clinical trials of historical importance on this procedure, see F. G. Miller, “The Enduring Legacy of Sham-Controlled Trials of Internal Mammary Artery Ligation,” *Progress in Cardiovascular Diseases*, 55 (2012), pp. 246—250.

[24] M. B. Leon et al., “A Blinded, Randomized, Placebo-Controlled Trial of Percutaneous Laser Myocardial Revascularization to Improve Angina Symptoms in Patients With Severe Coronary Disease,” *Journal of the American College of Cardiology*, 10 (2005), pp. 1812—1819.

[25] J. B. Moseley et al., “A Randomized Trial of Arthroscopic Surgery for Osteoarthritis of the Knee,” *New England Journal of Medicine*, 347 (2002), pp. 81—88.

[26] “Proton Beam Therapy Heats Up Hospital Arms Race, Kaiser Health News and NPR news report, May 31, 2013.