“TREATING THE MANY TO BENEFIT THE FEW”: USING EXAMPLES FROM HEALTHCARE TO MAKE STATISTICS EXCITING AND RELEVANT

Taiwo Amoo, Department of Business Management, Murray Koppelman School of Business Brooklyn College of the City University of New York, 2900 Bedford Avenue, Brooklyn, NY 11210, 718-951-5000 x2075, tamoo@optonline.net

Hershey H. Friedman, Department of Business Management, Murray Koppelman School of Business, Brooklyn College of the City University of New York, 2900 Bedford Avenue, Brooklyn, NY 11210, 718-951-5000, x.friedman@att.net

Martin Raphan, Brooklyn, New York, martin.raphan@law.nyu.edu

ABSTRACT

This paper describes an innovative way to make the introductory statistics course interesting and relevant to students. The authors suggest that a patient deciding whether to take a prescription drug or undergo a medical procedure should understand basic evaluation criteria like the number needed to treat (NNT) and number needed to harm (NNH), which are described in this paper. We also describe how information presented to consumers regarding the efficacy of drugs often focuses on the wrong statistical measures. Brochures and advertisements for medications often focus on the relative risk reduction (RRR), when NNT is more helpful in evaluating the efficacy of a treatment.

Keywords: Misuse of statistics, medical errors, healthcare statistics, ARR, RRR, NNT, NNH.

INTRODUCTION

There are several courses – most are in the area of mathematics – where students feel that they will never use the material and thus there is no point in learning it. Some students might openly ask the instructor, “Will I ever use this information? Why do I have to know this?” (Rarely is this question asked about, say, basketball, swimming, sex-education, or film studies.)

Business analytics is making statistics a hot field again. Today, employers are seeking to hire employees who have quantitative skills [4]. This makes it important to make the introductory statistics course exciting and relevant. Moreover, many employers are looking for employees with critical thinking skills [4]. The purpose of this paper is to demonstrate how NNT and related measures can be used in the introductory class to demonstrate the value of statistics and at the same time sharpen students’ critical thinking skills.

The following statistics can be used to answer the question, “Why do I have to know statistics?” According to the Institute of Medicine, about 100,000 hospitalized people die each year from medical errors that could have been prevented; 250,000 people die annually from mistakes made by the American healthcare system. Every year, 75,000 hospital patients die from an infection picked up in a hospital. Almost everyone knows that heart disease and cancer are the major causes of death in the United States; few people realize that medical errors are the third major cause of death [6]. People who understand some basic statistics will know when to actually undertake a procedure that will require hospitalization.

The costs and benefits of medical testing

This discussion can be tied into statistical errors. A discussion of Type I and Type II errors can be tied into problems associated with false positives in medicine. Indeed, students may mistakenly believe that the more medical tests done the better. They may also think that someone who wants to live a long life should see doctors every month for a checkup. They do not understand that the dangers of conducting too many medical procedures may be greater than that of conducting too few.

Developments in medical technology have provided doctors with tests that can probe for various conditions, and the sensitivity of such tests has only increased with time. This allows doctors to discover maladies earlier, and with higher probability than ever before. Technology has also allowed for medical tests that are cheaper, and can be performed more frequently. This does not mean, however, that increased testing is a pure good. The more tests a doctor performs, the greater the likelihood of finding true problems, but, at the same time, performing more tests can create increased costs, which go beyond the monetary costs of performing the test.

One such cost comes from the danger of overdiagnosis [10]. As our ability to see more of what is going on inside the body via the use of high-resolution scans increases, so does the likelihood that there will be an overdiagnosis. Welch, Schwartz, and Woloshin report that in people with no gallbladder disease symptoms, approximately 10% will exhibit gallstones in ultrasound scans; 40% of people without any symptoms will show damaged knee cartilage (meniscal tear) with MRI scans; MRI scans will show bulging lumbar discs in more than 50% of people with no back pain. In fact, a recent study consisting of a sample of 1,000 people with no symptoms willing to undergo a total-body CT screen, 3,000 abnormalities were found; 86% of subjects had at least one. The danger of overdiagnosis means that there is a diminishing return to performing additional tests. As more tests are performed, the greater the likelihood of finding true problems, but the greater the likelihood of finding spurious problems.

Another problem is that such tests can cause much suffering. Beyond the fact that many tests, such as mammograms and prostate exams can be frightening and uncomfortable, many tests can also have harmful side effects that last a lifetime. For example, many patients who receive radical prostatectomies suffer from harmful side effects that include incontinence and impotence [2]. It is not clear, then, whether the increased probability of finding a true instance of disease is worth the harm caused to multiple patients. Bach (2012), for example, notes that “with routine mammography, you’d have to screen more than 1,000 women in their 40’s to prevent just one breast cancer death.” The guidelines for prostate exams and surgery were changed when it became known that more than 80% of radical prostatectomies performed in the United States are unnecessary. Apparently, only one out of 48 men have their lives extended by this type of surgery [2].

NNT / NNH

How, then, to measure the value of additional testing against its costs? One potential way of measuring the value of an additional test is to estimate the amount by which it increases the lifespan of the patient. However, if not properly calculated, this statistic can be misleading. Suppose people with a certain disease live to the age of 65 on average. If the disease is discovered when people are 60, then we can say that the average person with the disease lives for five years, or that the survival time is 5 years. Suppose with better technology we are able to discover the disease earlier, say when the typical patient is 50. Even if patients continue to have the same lifespan, living to the age of 65, it would appear that the survival time has increased to 15 years. Thus, because the survival time is calculated from the time the disease is *discovered*, early detection gives the false impression of increasing lifespan. Generally, early diagnosis may not result in a longer life but simply increase the time the disease is discovered until death. Thus, while it is “true that patients diagnosed early have better survival statistics than those diagnosed late,” this does not mean that early diagnosis actually helps [10].

Another statistic that can be used to measure the effectiveness of a procedure, test or drug is the difference of occurrence in a treated group and occurrence in a control group. However, such statistics can be manipulated to mislead as well. Suppose an experiment is done and a large sample of people are randomly assigned to two groups; one group takes a placebo and the other group takes an experimental drug for, say, five years. Suppose, at the end of the study, 3 out of every 100 people in the placebo group had strokes while 2 out of every 100 people in the experimental group had strokes. Would it be correct to say that the drug reduced the number of strokes by one-third (from 3 to 2)? This makes the drug sound quite effective, but is arguably misleading. Only one person out of 100 benefitted from taking the drug; 99 out of 100 got nothing out of taking the drug. This example is not fiction. This is essentially what Pfizer did in promoting its statin, Lipitor [5]. Pfizer ran a campaign targeted to consumers that declared: “Lipitor reduces the risk of heart attack by 36%... in patients with multiple risk factors for heart disease.”

More generally, such studies often suffer from a misunderstanding about the meaning of statistical significance. Statistical significance is about statistics – how sure are we that the difference observed is real, and not just the result of some random deviation? It is possible for the difference between two populations to be statistically significant, yet completely “insignificant” in a practical sense. I may be 100% certain that I have a penny in my pocket, but that does not make its value significant. Similarly, it may be true that a drug has decreased the incidence of stroke from 3 in 100 to 2 in 100, but that does not mean that this decrease is of sufficient importance to warrant the risks involved in taking it.

Another statistic that is used to measure the effectiveness of a procedure, test or drug is Number Needed to Treat (NNT). NNT was defined as “the number of patients that would need to undergo a particular treatment over a specific time period in order to see their health improve beyond what would have happened had they done nothing or had they undergone a different treatment " [5]. Suppose only one in 100 patients who take a drug are cured. Then, for every patient who receives a benefit from taking the drug, there would be 99 who received no benefit, or were adversely affected. In this case, since 100 patients must take the drug for one patient to be positively affected, the NNT is 100. A related statistic, which focuses on the potential harm caused by a treatment or test is the number needed to harm, or NNH. Obviously, when evaluating a course of treatment, it is important to consider both numbers.

NNT can help clarify the difference in populations. Consider the test above, where a drug reduced the incidence of stroke from 3 in 100 to 2 in 100. In this case, only one person out of 100 benefitted from taking the drug; the NNT is 100. When stated this way, the potential ineffectiveness of the drug is more apparent. There is evidence that the NNT for low-risk patients using statins for five years is 250 (Carey, 2008). If the NNT were made available to the public, it might result in reduced medical costs and better health. Incidentally, medical experts say that one should not take a drug with an NNT of over 50 (Carey, 2008). This, of course, assumes that the NNH is not a problem. If it is, then even if NNT is low, the concurrent likelihood of being harmed by the course of treatment might mitigate against taking it.

As another example, consider sinusitis. It is a quite common condition, with 20 million cases a year. It is usually the result of a virus, in which case antibiotics are ineffective. Despite this, many physicians will prescribe antibiotics anyway; 20% of all antibiotic prescriptions in the United States are for sinusitis. (http://www.thennt.com/nnt/antibiotics-for-radiologically-diagnosed-sinusitis/) Analyzing the statistics for use of antibiotics in treating sinusitis shows that the NNT is 15, while the NNH is 8. (http://www.thennt.com/nnt/antibiotics-for-radiologically-diagnosed-sinusitis/) This means that only one out of every 15 people with sinusitis was helped by taking antibiotics, while one out of 8 were harmed (vomiting, rash, and/or diarrhea from the medication). There are likely a number of patients who, presented with this knowledge, would choose not to take the medication.

The following example, further illustrates the different ways of presenting data from the same study, and how they can be used to mislead:

> You read that a study found that an osteoporosis drug cuts the risk of having a hip fracture in the next three years by 50%. Specifically, 10% of the untreated people had a hip fracture at three years, compared with 5% of the people who took the osteoporosis drug every day for three years. Thus 5% (10% minus 5%) less people would suffer a hip fracture if they take the drug for 3 years. In other words, 20 patients need to take the osteoporosis drug over 3 years for an additional patient to avoid a hip fracture. 'Cuts the risk of fracture by 50%' represents a relative risk reduction. 'Five per cent less would suffer a fracture' represents an absolute risk reduction. 'Twenty patients need to take the osteoporosis drug over 3 years for an additional patient to avoid a hip fracture' represents a number needed to treat" [8].

The ARR (absolute risk reduction) is 5%; RRR (relative risk reduction) is 50%; and NNT is 20. Advertisements and brochures would stress the 50% RRR, creating the false sense of effectiveness. This sort of practice may help explain why we spend too much on health care and have little to show for it. Patients may fall into the trap of choosing treatments with high relative risk reduction, but with large NNTs, resulting if few patients actually helped. "Treating the many to benefit the few" is an apt description of health care in the United States [9].

Discussion

With the abundance of medical information thrown at a patient, it is important to give them some basic numbers that they can use to form their decisions. We have seen that there are a number of ways to present this information, and that some of them can be misleading. In particular, we have seen that advertisements sometimes tout a large relative reduction in likelihood of incidence of a disease, while the absolute reduction may be quite small. We have discussed the NNT and NNH, and have seen that those statistics may be more helpful in evaluating treatment.

This does not mean that NNT and NNH are without flaws. In particular, while these statistics are indicative of the likelihood of being helped or harmed by a drug, they do not indicate the magnitude of various benefits or harms. For example, we indicated that an NNT of 15 and NNH of 8 might mitigate against a treatment for sinusitis. Suppose instead that a treatment for terminal cancer had the same numbers – one out of every 15 patients were totally cured, while one out of every 8 patients treated were not cured, but developed a rash. Would the evaluation of the treatment still be negative?

Conclusion

This paper demonstrates how a discussion of healthcare measures such as NNT and NNH may be used in an introductory statistics course to make it innovative and exciting. In addition, it can be used to enhance the critical thinking of students. Students will see with real world data why statistical significance is not sufficient and a practical importance measure such as NNT is also needed before a drug is prescribed. This paper also can be used as a tool to teach how statistical data can be used in an unethical way to prescribe harmful drugs, i.e., recommending drugs when the NNT measure is high. NNT and NNH can also be used as part of a discussion involving the different types of errors in research. It is especially important for students to see that statistics is not a dry course with meaningless formulas. On the contrary, understanding how statistical data can be used and misused in healthcare, may save someone’s life.

REFERENCES

[1] Bach, P. B. (2012, June 5). The trouble with 'doctor knows best.' *New York Times*, Health, D6.
[2] Blum, R. H. and Scholz, M. (2010). *Invasion of the prostate snatchers*. New York: Other Press.
[3] Carey, J. (2008, January 28). Do cholesterol drugs do any good? *Business Week*, 52-59.
[4] Gerstein, M. & Friedman, H. H. (2016), Rethinking Higher Education: Focusing on Skills and Competencies. *Psychosociological Issues in Human Resource Management*, 4(2), 104-121.
[5] Heisel, W. (2010, May 21). NNT can be TNT for blowing up pharma marketing claims. *Reporting on Health*. Retrieved from http://www.reportingonhealth.org/blogs/nnt-can-be-tnt-blowing-pharma marketing-claims
[6] Laliberte, R. (2016). 12 ways the healthcare system may be harming you. AARP Bulletin, 18-25.
[7] Linville, P. W., Fischer, G. W., & Fischhoff, B. (1992). AIDS risk perceptions and decision biases in J. B. Pryor and G. D. Reeder (Eds.), *The social psychology of HIV infection*. Hillsdale, NJ: Erlbaum.
[8] Napoli, M. (2011, May 15). Most drugs don't work… *Center for Medical Consumers*. Retrieved from http://medicalconsumers.org/2011/05/15/most-drugs-dont-work/
[9] O'Dowd P. (2001). Treating the many to benefit the few: The modest reach of preventative medications medicine and health. *Medicine and Health* /*Rhode Island*, 84, January, 4-8.
[10] Welch, H. G., Schwartz, L. M., and Woloshin, S. (2011). *Overdiagnosed: Making people sick in the pursuit of health*. Boston: Beacon Press.