

Understanding the Odds: Statistics in Public Health

ABSTRACT

In the world of public health and medicine, researchers are often trying to discover new ways of understanding and preventing diseases and other negative health outcomes. When public health researchers want to examine the relationship between some sort of exposure, like smoking, and a disease, such as lung cancer, they will often start by calculating what is called an odds ratio. An odds ratio is a comparison of odds between people who were exposed and people who were not exposed. However, odds ratios can be tricky to understand, even for experienced researchers. In this article, we will break down the odds ratio by reviewing the concepts and calculations of probability and odds. We will also discuss how to interpret an odds ratio, and how these ratios can be useful in real-world applications.

Key Words: Statistics, Probability, Odds, Odds Ratio, Public Health

WHAT ARE THE ODDS?

Imagine it's the middle of winter and you and your friends are camping by a lake. You turn to your friend and ask, "What are the odds you jump in the lake?" Immediately your friend protests and says, "Zero!" You understand this to mean that there is no way your friend will be jumping in the freezing lake—but have you ever considered what "odds" really means? Many times, the term "odds" is actually used incorrectly. However, odds are a vitally important concept in **biostatistics** [1]. More specifically, odds can be used to calculate odds ratios, which explain the relationship between measurable characteristics that scientists are interested in, such as the relationship between smoking and lung cancer. As such, odds ratios can be used in important areas of research that have real-world implications, making this a critical concept for young scientists who are looking to enter the world of public health.

PROBABILITY AND ODDS

To understand the odds ratio, we first need to understand probability and odds. **Probability** refers to the likelihood that an event will occur. Probability is a number between 0 and 1, where 1 is a 100% chance that the event will occur. Let's think about a deck of cards. In a deck of 52 cards, there are 13 cards of the heart suit. If you shuffled the deck of cards and picked the first card off the top of the deck, the probability of getting a heart suit would be:

$$\frac{13 \text{ (\# of hearts)}}{52 \text{ (\# of cards in the deck)}} = \frac{1}{4} = 0.25$$

This would give you a probability of 0.25, or a 25% chance of drawing a heart. Understanding probability is the basis for understanding odds and the odds ratio.

Odds are simply the ratio of two probabilities: the probability that the event *will* occur divided by the probability that the event *will not* occur [2]. Recall that a probability of 1 is a 100% chance of occurring. Therefore, if you subtract the probability of a specific event occurring from 1, you are left with the probability of the event *not* occurring (**Figure 1**).

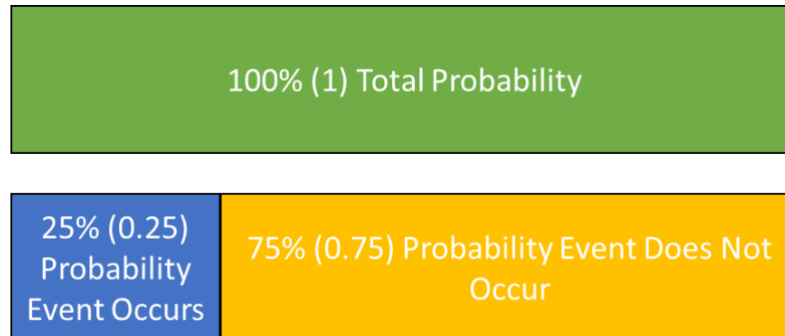


Figure 1 – The total probability is the combination of the probability that the event *will* occur and the probability that the event *will not* occur. Total probability always equals 1.

Using the previous example of a deck of cards, the odds of drawing a heart would be:

$$\frac{\text{Probability of drawing a heart}}{\text{Probability of NOT drawing a heart}} = \frac{0.25}{1 - 0.25} = \frac{0.25}{0.75}$$

This would give you an odds of 0.33, or 1/3. This can be stated as a 1 to 3 odds of drawing a heart. We can understand this to mean that we expect to observe one success (heart) for every three failures (non-heart), or that a failure is three times as likely as a success [3].

ODDS RATIO

When working with **odds ratios**, researchers are interested in the relationship an **exposure** and an **outcome**. As in our earlier example, a researcher may be interested in the relationship between smoking (exposure) and lung cancer (outcome). However, exposures aren't always negative events that cause a disease—sometimes they can be preventative. An example of a preventative exposure would be a vaccine, or the use of protective equipment such as gloves.

The odds ratio measures the relationship of an exposure and a disease through division. The group on the bottom of the fraction (the divisor) is the **reference group**, which is usually the unexposed group. Typically, the odds ratio compares the odds of the outcome they are interested in in the exposed group to the odds of the unexposed group. This can be hard to follow, so we often use a type of table called a **contingency table** to organize this information [2]. Let's use a hypothetical example of eating sugary candy as an exposure and getting a cavity as the outcome. The columns in the contingency table (**Figure 2**) represent the outcome status (yes or no), and the rows represent the exposure status (yes or no).

		Cavity		
		Yes	No	
Candy	Yes	 35	 10	45 
	No	 15	 40	55
Total		50 	50 	

Figure 2 – A 2 x 2 contingency table (two squares by two squares) for sugary candy (exposure) and cavities (outcome). This table shows the number of people in each combination of the two categories: eats sugary candy (yes/no) and has a cavity (yes/no). This data is not real; it is meant for educational purposes only.

Based on Figure 2, let's calculate the odds ratio for those who eat sugary candy. The odds ratio for compares the odds that an outcome will occur for those who *have* had the exposure versus the odds that the outcome will occur for those who have *not* had exposure. In this circumstance, the odds ratio would be the odds that a person has a cavity if they eat sugary candy versus the odds a person has a cavity if they do not eat sugary candy.

First, we can calculate the numerator of this ratio (**Figure 3A**). The numerator is the odds of having a cavity if a person eats sugary candy. As previously discussed, odds are the probability that the event will occur, divided by the probability that the event will not occur. The probability that someone will have a cavity, given that they eat sugary candy, would be 35 (number of individuals who have a cavity and eat sugary candy) divided by 45 (total number of people who eat sugary candy). The probability that someone will not have a cavity, given that they eat sugary candy, would be 10 (number of individuals who do not have a cavity and eat sugary candy) divided by 45 (total number of people who eat sugary candy). This would give us 3.5.

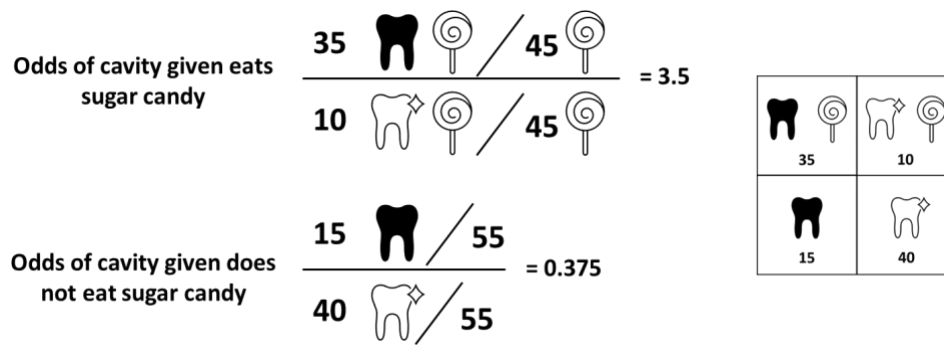


Figure 3 – Calculations for (A) odds of a cavity for those who eat sugary candy and (B) odds of cavity for those who do not eat sugary candy.

Next, we can calculate the denominator of this ratio (**Figure 3B**). The denominator is the odds of having a cavity if a person does not eat sugary candy. The probability that someone will have a cavity, if they do not eat sugary candy, would be 15 (number of individuals who have a cavity and do not eat sugary candy) divided by 55 (total number of people who do not eat sugary candy). The probability that someone will not have a cavity if they do not eat sugary candy would be 40 (number of individuals who do not have a cavity and do not eat sugary candy) divided by 55 (total number of people who do not eat sugary candy). This would give us 0.375.

Finally, the odds ratio is the ratio of the two odds we just calculated. The calculation would be as follows:

$$\frac{\text{Odds of cavity given eats sugary candy}}{\text{Odds of cavity given does not eat sugary candy}} = \frac{3.5}{0.375} = 9.33$$

The odds ratio of having a cavity for those who eat sugary candy compared to those who do not eat sugary candy is 3.5 divided by 0.375, which gives us 9.33. However, 9.33 may not intuitively mean anything to you, so let's discuss how you interpret odds ratios.

INTERPRETING ODDS RATIOS

There are three main interpretations for an odds ratio: no association, protective factor, and risk factor. If an odds ratio is equal to 1, this means that there is no association between the exposure and the odds of disease [4]. This is because, for an odds ratio to be equal to 1, the odds of disease given exposure and the odds of disease given no exposure are the same (any number divided by itself is 1). If the odds are the same, this would mean that the exposure does not affect the probability of disease. In the cavity example, you could think of the odds of eating name-brand sugary candy and getting a cavity, compared to the odds of eating generic sugary candy and getting a cavity. These two groups will likely have the same odds of getting a cavity, and thus their odds ratio will be equal to 1.

If the odds ratio is less than 1, this means that the exposure is a protective factor. What this means is that the exposure is associated with lower odds of disease [4]. As an example, an odds ratio of 0.5 means that there is a 50% decrease in the odds of disease if you have the exposure. An example of an exposure with a protective factor would be brushing your teeth twice a day. If you brush your teeth morning and night, you have smaller odds of getting a cavity compared to those who do not brush teeth morning and night.

Finally, if the odds ratio is greater than 1, this means that the exposure is a risk factor. This means that the exposure is associated with higher odds of disease [4]. As an example, if the odds ratio is 1.5, the odds of disease after being exposed are 1.5 times greater than the odds of disease if you were not exposed. Another way to think of it is that there is a 50% increase in the odds of disease if you are exposed. For our sugary candy example, an odds ratio of 9.33 means that sugary candy is a risk factor for getting a cavity. We could also state that if you eat sugary candy, your odds of getting a cavity are 9.33 times greater than the odds of getting cavity for those who do not eat sugary candy.

CONCLUSION

Researchers use odds ratios in many aspects of research. From the testing of new medical techniques, to vaccination trials testing the effectiveness of vaccines, to examining the impact of environmental factors like pollution or UV rays, odds ratios can help researchers understand how exposures effect the odds of disease. Naturally, odds ratios are often much more complex in real-life applications. However, these calculations are fundamental concepts for those working in public health. In light of the recent COVID-19 pandemic, it will be critical for upcoming young scientists to understand odds and odds ratios, so they can make a difference in the world of public health.

Glossary

Biostatistics: The discipline concerned with the treatment and analysis of numerical data derived from biological, biomedical, and health related studies.

Probability: The proportion of times an even is expected to occur in the long run, which are always a number between 0 and 1.

Odds: The odds of an event is the proportion of successes divided by the proportion of failures.

Odds Ratio: A measure of association which measures the odds of an event happening in one group compared to the odds of the same event happening in another group.

Exposure: Any factor, condition, or characteristic that may be associated with an outcome of interest.

Outcome: The outcome of a study is a broad term for any defined disease, state of health, health-related event or death.

Reference group: A group to which an individual or another group is compared.

Contingency table: A table that displays the number of subjects observed at combinations of possible outcomes for the two variables.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 2046122. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

We would like to thank the members of the Ryu Computational Lab Group for their review and contribution to this article, including Riley Brenner, Kurtis Bertauche, Alexander Choi, and Korben DiArchangel.

REFERENCES

1. Gerstman B. *Basic Biostatistics: Statistics for Public Health* . 2nd ed. Jones & Bartlett Learning (2015).
2. Agresti A. *Statistical Methods For The Social Sciences*. 5th ed. Pearson (2018).
3. Agresti A. *Wiley Series in Probability and Statistics*. 3rd ed. John Wiley & Sons (2019). Available at: <http://www.wiley.com/go/wsps>
4. Szumilas M. Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Academie canadienne de psychiatrie de l'enfant et de l'adolescent* (2010) **19**:227–229. Available at: <https://pubmed.ncbi.nlm.nih.gov/20842279>