10.5005/jp-journals-10043-0011

CONTINUING MEDICAL EDUCATION (CME)

Medical Statistics Made Easy for the Medical Practitioner

Tiny Nair

ABSTRACT

Understanding basic medical statistics is important in today's medical practice, not merely as an academic exercise but to translate medical information into day-to-day patient care. This review article tries to address basics of medical statistics to the end user.

Keywords: Clinical trial, Probability, Standard deviation, Statistics.

How to cite this article: Nair T. Medical Statistics Made Easy for the Medical Practitioner. Hypertens J 2015;1(2):63-67.

Source of support: Nil

Conflict of interest: None

INTRODUCTION

Most practicing physicians view medical statistics as a complex mathematical topic alien to biological science and medicine. Statistical jargons create a sense of fear and a compulsive response to avoid delving into this arena. But most medical publications cite, quote and depend on is statistical terms, which we need to understand as an end user. Understanding a medical publication, be it a drug trial, a case report, an epidemiological study or a meta-analysis, needs knowledge of medical statistics. The intention of this article is to simplify statistical terms, so that the reader can differentiate good robust publications from statistically weak ones.

STATISTICAL TERMS

Probability

Probability (p) is the most commonly used term in statistics. Probability is an indicator of how much of the result (outcome) can happen just by chance. If we toss a coin, it has a 50% chance of showing a head or a tail. If we continue to toss hundreds of time, the chance of it showing head or tail tends to be about half (50%). This can be mathematically put as 50/100 = 5/10 = 0.5.

Head

Department of Cardiology, PRS Hospital, Thiruvananthapuram Kerala, India

Corresponding Author: Tiny Nair, Head, Department of Cardiology, PRS Hospital, Thiruvananthapuram, Kerala, India e-mail: tinynair@gmail.com

This means that a probability value of (p) = 0.5 indicates that there is 50% probability that the result may be happening just because of a chance.

In similar sense, different p-values would mean the following:

p = 0.05 = 5/100 = 1/20 chance of the result being accidental

p = 0.01 = 1/100

p = 0.001 = 1/1000

To be statistically significant, we need to look for a p-value of at least less than 0.05.

Statistically speaking:

$p < 0.05$—statistically significant

$p < 0.01$—highly significant

$p < 0.001$—very highly significant

Always look at the sample size, effect size and confidence interval before giving a judgment on p-value alone.

Sample Size and Effect Size

p-value is inversely proportional to effect size and sample size. This means if we are testing the effects of a new blood pressure medicine is reducing blood pressure by 20 mm Hg, but was tested in only 100 patients, the p-value may be significant but the sample size is too small to be significant. In similar lines, a drug reducing blood pressure by merely 1 mm Hg may return a significant p value, when tested in 1 million population but the effect size is too small to be really worthwhile statistically.

To avoid statistical error, today, the statistician calculates the minimum sample size that would be required to show a difference of results depending on the prevalence of the event rate in the natural course of the disease in the population being studied.

Sample Size Calculation

To decide on the sample size of a trial, the following factors need to be considered:

Sample size calculation
• Effect size
• Significance level
• Power of a trial

Effect Size

If a 'old' drug A is effective in 30% population and 'new' drug B in 40%, the effect size is calculated as:

B–A
40–30 = 10% (absolute)
and 30/40 = 75% (relative)

Significance Level

If we are looking at 5% significance level, we have a p-value of 0.05. If a study is not significant but gives an impression of being significant (generally because of borderline p-value), it is called a type 1 or alpha error.

The alpha error can be calculated from the nomogram (Table 1) (for two-tailed test, it is 1.96).

Power

If in a trial, the number of sample is small, a significant trial may look statistically nonsignificant. This is called beta error. A repeat trial with a larger sample size is likely to correct such error. Power of a trial is 1-β. The usual power of a trial is kept at 80% (0.8).

Power = 1-β

Sample size varies directly with significance level (precision) and power and inversely with effect size:

$$\text{Sample size } (\alpha) = \frac{\text{significance } (\alpha) + \text{power } (1\text{-}\beta)}{\text{Effect size}}$$

Odds and Odds Ratio

The term odds mean a disease or effect happening *vs* not happening.

Supposing that 10 out of 100 patients of acute myocardial infarction (MI) would die, the odds are 10 will die and 90 will live.

So the odds are 10/90 = 0.11 (happens/not happen).

Now a medical paper says that there is a new drug ABC shows benefit in reducing death rate of MI. On being treated with the new drug, only 2 out of 100 acute MI cases died. This means 2 dies and 98 live.

So odds for this new treatment are 2/98 = 0.02.

Odds ratio = control odds/treatment odds = 0.11/0.02 = 5.5

This means treatment with this new drug reduces chance of death by 5.5 times.

Risk and Risk Ratio

Risk is a similar term that means disease or effect out of the entire population. In the same example of the heart attack above, the risk of death is 10 out of 100 (not 10/90 as in odds).

Table 1: Alpha error at different significance levels

Alpha error	5%	1%	0.1%
two-sided	1.96	2.57	3.29
One-sided	1.65	2.33	—

Risk of death in MI 10/100 = 0.10 (happens/total)

Similarly, risk after being administered ABC has a risk of 2/100 = 0.02.

Risk ratio = 0.10/0.02 = 5 (very close to odds ratio)

Note that, in most cases, odds ratio and risk ratio is close.

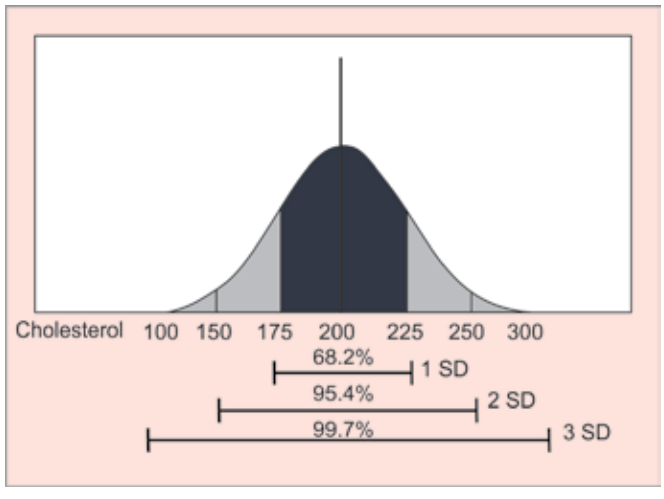
Now consider, unlike the example of MI, a disease has a mortality of 90% (90 out of 100 die). In such a scenario, the odds would be 90/10 (died 10 survived) returning a value of odds of 9, while the risk ratio would be 90/100 = 0.9. So, dichotomy between odds and risk indicate high event rate in control group and this may corrupt a study.

Standard Deviation

Any physiological parameter will have variation from people to people. If we take sufficient number of people and plot their values, they tend to follow a normal or nominal distribution (a bell-shaped curve). The central line in the curve is the mean (akin to arithmetic average). As we go farther away from the mean, the degree of deviation (from mean) increases. One standard deviation (SD) covers 68.2% of data spread from mean. Two SD covers more, up to 95.4% data and 3 SD covers almost a whooping 99.7% data (Graph 1).

Example

Let us look at the example of serum cholesterol in a population of a town of 8,000 people (Graph 1). Most people have a mean cholesterol of 200 mg (central line), while some as low as 100 and as high as 300. In the graph, 1 SD is 25 mg of cholesterol. This means, if we look at ± 1 SD, the value would range from 175 to 225. One SD constitutes of 68.2% of the population data in any normal distribution (5,456 people) the entire dark group. If we take 2 SD, it covers the cholesterol values from 150 to



Graph 1: Hypothetical standard deviation (SD) curve of distribution of serum cholesterol (mg%) levels in a population

250 covering 95.4% of the population (7,632 people). Three SD covers 99.7% of the population from the cholesterol level of 100 to 300 (7976 of 8000). This means 3 SD practically cover entire data set.

How to use SD Intelligently

You read a paper, which says that, after a stroke, the patients have stayed in hospital for 8 ± 3 ($p < 0.01$). Look pretty impressive? The number ±3 denotes 1 SD and that is 68.2% patient stayed in the hospital from 5 to 11 days (8 ± 3).

Now, multiply 3 × 3 SD to get the 3 SD values (covering 99.7%). This value is 9. This means the patients would have stayed in the hospital from –1 (minus 1) day to 17 (8 ± 9) days. This is mathematical impossibility. So, this study SD is unacceptable.

Confidence Interval

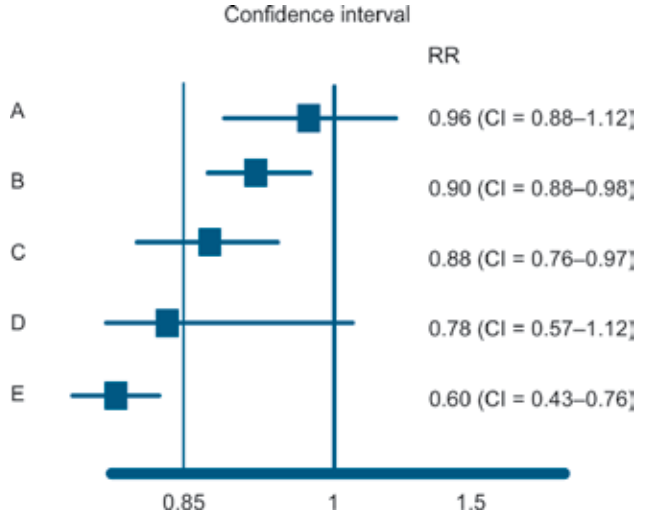
Confidence interval (CI) is the interval which indicates the lowest and the highest chances of the occurrence in the particular trial. In simple terms, a confidence interval of 18 to 34 means that the next time you do the same study the result may be as low as 18 or as high as 34.

A confidence interval is good (acceptable) if:

- It is tight, meaning that the variation or speed is small.
- It is on the same side of unity. A confidence interval of 0.67 to 1.32 means the risk of problem may be as low as 67% (reduced by 33%) or increased by 32% meaning poor correlation. So, confidence interval on one side of unity makes it statistically significant.
- Just being on one side of unity does make it statistically significant. But is it clinically relevant? A value below 0.85 (85%) makes it clinically significant.

Example (Graph 2)

- Confidence interval = 0.86 (CI = 0.88–1.12)—the value wraps around 1 and hence it is statistically not significant.
- Confidence interval = 0.83 (CI = 0.88–0.98)—the upper value is below 1. So, it is statistically significant. However, clinical significance of a value is considered only when it is below 0.85. So, in this case, it is statistically significant, but clinically not significant.
- Confidence interval = 0.78 (CI = 0.76–0.97)—here, the upper value is below 1, while the lower one is below 0.85. Hence, it is statistically significant and may be clinically significant.
- Confidence interval = 0.78 (CI = 0.57–1.12)—here the value wraps around 1, hence statistically not significant. However, since the lower value is below 0.85, it may still be clinically significant. In such cases, a repeat trial is indicated.



Graph 2: Different variations in confidence interval (CI): A—not significant; B—statistically significant but clinically not significant; C—statistically significant and may be clinically significant; D—statistically not significant and may be clinically significant; E— statistically and clinically significant

- Confidence interval = 0.59 (CI = 0.43–0.76)—in this case, both the upper and lower values are below 0.85, so it is statistically as well as clinically significant.

Study Example

How to interpret a Clinical Trial

A clinical trial tells us whether a new therapy is better than older one. If so it changes our clinical practice pattern. So, we need to carefully look into the results to decide how much impact it would have on our clinical practice tomorrow. Most new studies try to show that the effect of the study medicine is better than the present standard of care, some try to prove noninferiority.

Example

Journal report: New Drug XYZ reduces heart failure mortality by 20% study with a new drug XYZ' or heart failure.

One thousand patients of heart failure given standard care 5 died at end of 1 year.

Another 1,000 patients of heart failure was given XYZ, 4 died at end of 1 year.

Absolute risk reduction (ARR)—the real thing = 0.5–0.4 = 0.1%.

But the authors would put it in this way:

Relative risk reduction (RRR)—RRR 0.1 prevented out of 0.5 = 0.1/0.5 = 20%.

So, XYZ reduced death by 20%.

The logical next step would be to check for p-value, number needed to treat (NNT) and CI.

Look for p-value. Accept if $p < 0.05$.

Many noninferiority trials try to show whether the drug or treatment is 'not worse' compared to the present gold standard, but do not ask whether it is better or not. This means it is checking one end (single tail) of the standard bell curve and not both the ends (double tails). A single tailed trial needs a higher p-value to be significant (at least $p < 0.01$).

Ask for confidence interval, if wraps around unity, reject it (e.g. 0.56–1.23 goes to junk).

Enquire prespecified endpoints.

Calculate NNT—NNT is number needed to treat to prevent one event (here one death).

Number needed to treat is calculated by dividing 100/APR = 100/0.1 = 1000.

So you need to treat 1000 patients to prevent one death, NNT is significant (acceptable) if more than 50. This trial it is 1000, so not significant.

TYPES OF TRIALS

The Flow Chart 1 shows the various types of trials in clinical practice. While observational studies are important in deciding prevalence of a disease or its complications, interventional studies which are randomized and placebo controlled are the most robust.

Study Design

There may be several types of clinical trial depending on what exactly we need to try. Today, the cardiovascular (CV) event rates (death, MI, stroke) of most diseases are so low that to detect a difference made by a 'new' drug may be quite low in absolute numbers. This necessitates the need for larger trials with more power to detect even a mild benefit of the 'new' drug.

Randomized Trials

Active (new drug) group is allocated randomly to avoid bias. In a nonrandomized trial, a new drug may be given only to patients with milder disease, resulting in an erroneous interpretation that the drug is better. Randomization avoids that bias.

Placebo Controlled

Today, all patients need to be given standard of care for diseases. The active new drug group is given the new drug over and above the standard therapy, making the standard of therapy as placebo.

Multicentric

To make trials larger in number, multiple centers collaborate together to get larger numbers. Multicentric trials also ensure a homogenous mixture of patients of different ethnicity, different socioeconomic backgrounds, making it easy to decide whether this therapy is consistent among all segments of population.

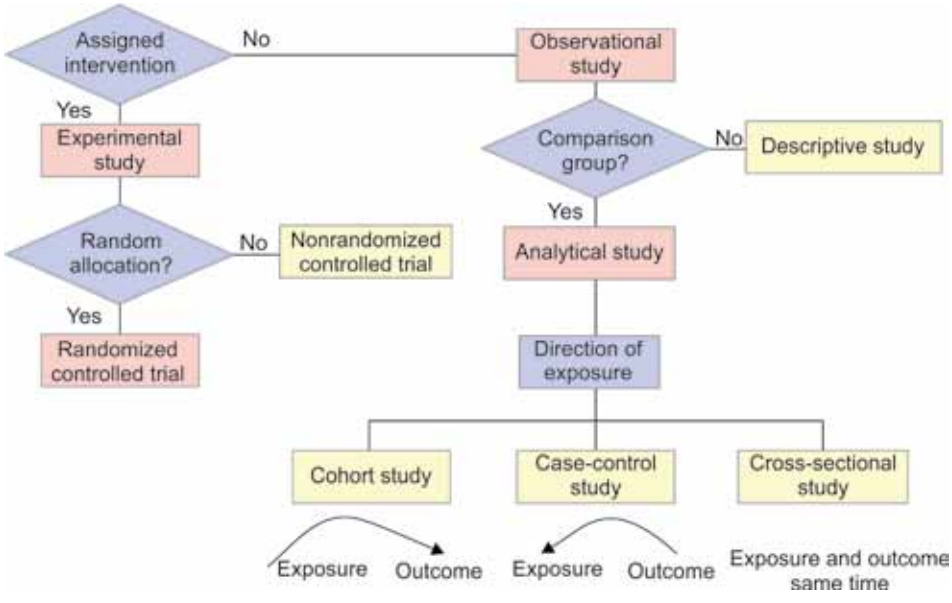
Blinded

Blinding makes the patients and the treating physicians blind to what therapy they are on. The analysis of the trials is done by analysts and statisticians who are also blinded.

End Points

These are prespecified prior to the beginning of the trial. End points can be primary and secondary. Most CV trials have primary end points as a combination of death, nonfatal MI and nonfatal stroke. Nonprespecified

Flow Chart 1: Type and design of clinical trials



end points, analyzed at a later date after the trial (as an after thought) are called posthoc analysis which has less statistical power of prediction.

PROBE Design

The downside of large RCTs are the cost. One way of reducing cost is to do an open trial but with endpoints blinded. These trials are called prospective randomized open-labelled end-point blinded (PROBE) trial. Here, the endpoints are blinded, giving it power to analyze significance, but many consider it inferior to RCTs.

CONCLUSION

Understanding the basics of medical statistics help the clinician in making important judgments and decisions. It helps the clinician to make evidence-based changes in the practice, or junk it depending on the merits. A step-by-step approach by looking at ARR, p-value, NNT and confidence interval can separate statistically robust data.

REFERENCES

1. Essential Evidenced-based Medicine, Dan Mayer, Cambridge Medicine. 2nd ed.
2. Harris M, Taylor G, Dunitz M. Medical Statistics Made Easy.
3. Magnello E, Loon BV. Statistics, A graphic Guide, Icon Books, UK.
4. Nair T. Medical Statistrics and Clinical Trials. 1st ed. Wiley Blackwell.
5. Sanjay K, Diamond G. Trial and error: how to avoid commonly encountered limitations of published clinical trials. J Am Coll Cardiol 2010;55:415-427.