CHAPTER 1

Medicine and Numbers
Where Is the Connection?

[…] for complete initiation as an efficient citizen of one of the new great complex world-wide States that are now developing, it is as necessary to be able to compute, to think in averages and maxima and minima, as it is now to be able to read and write.

H.G. Wells

Learning Outcomes

We shall discuss the following material in this chapter:

- Why physicians need to study medical statistics
- The utility of descriptive statistics
- The utility of inferential statistics

Medical Statistics : A Modern-Day Crystal Ball

At first glance, medicine and numbers may appear far too much apart, but a closer inspection will tell a different story. In days gone by, people turned to ball-gazers in search of answers for health, happiness and future. We now live in a universe of big data. Where crystal balls, distant stars or dried bones once conspired to rule upon citizens' lives, information now reigns supreme. A random set of numbers is no good until we apply the powerful tool of data analysis. It is the best means we currently have of making a meaningful pattern out of a group of random numbers. The insight we gain helps us to make predictions for the future. Statistics is our modern-day crystal ball: it is the collection and analysis of data that is our window to the past, interpreter of the present and guide to the future. With the widespread availability of statistical software, the calculation of statistics is no longer complicated. The key to getting the right answer from our crystal ball is to appreciate when to choose which statistics and to recognise the nuances of the reply.

Did You Know?

Figure 1.1 The scholar depicted in this Syrian postage stamp is **Abu Yusuf Al Kindi** (801 – 873), an Arab mathematician and polymath. He is credited with making one of the earliest known statistical analysis in his treatise ***On Deciphering Cryptographic Messages*** [2].

https://doi.org/10.1017/9781108973663.004 Published online by Cambridge University Press

Think About It!

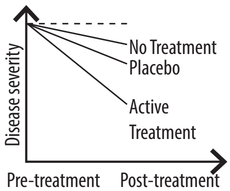


Figure 1.2 Comparison of disease severity between no treatment and active treatment [3]. © BMC Medical Research Methodology, CC BY 2.0 .

What does Figure 1.2 suggest?

Apart from numbers, can descriptive statistics be conveyed in any other fashion?

What is the limitation of descriptive statistics?

Why don't we discuss uncertainty while relaying descriptive statistics?

Why Do We Need to Study Medical Statistics?

Nobody expects a physician to study medical statistics to become a fully fledged statistician, although no one will stop you. Yet, a fundamental understanding of statistics is essential for all physicians in our everyday practice. In our everyday lives, we research every product we buy, compare ratings, look at reviews and feedback before making decisions. The modern man or woman makes a well-informed decision before committing to a product. Medicine is no different. We need to learn medical statistics so that we can make informed decisions about patient care.

Statistics gives us the tools to find a pattern from a random set of numbers borne out of observations, experiments and trials, and to communicate the results meaningfully. Understanding this common language allows both the researcher and the audience to make sense of the numbers. Informed decision-making is only possible if we can understand what kind of test was performed, why it was performed, what was the size and effect of an intervention, as well as its significance. Understanding this will allow us to differentiate between hype and truth.

Descriptive Statistics: Seeing the Wood and the Trees

A useful function of medical statistics (Figure 1.3) is to **describe** the data observed in an experiment or trial in a condensed format and relay the information in a way that is understood by readers. This part of medical statistics is known as **descriptive statistics**. Descriptive statistics help us to make sense out of the confusion of random numbers.

For example, if we measure weight, we shall employ descriptive statistics to understand as well as inform the reader about the average weight, the distribution from the lightest to the heaviest, and the most frequent weight observed etc.

Here is an example from a recent study where the authors wished to investigate the prevalence of musculoskeletal injury (MSK-I) among athletes [4]:

https://doi.org/10.1017/9781108973663.004 Published online by Cambridge University Press

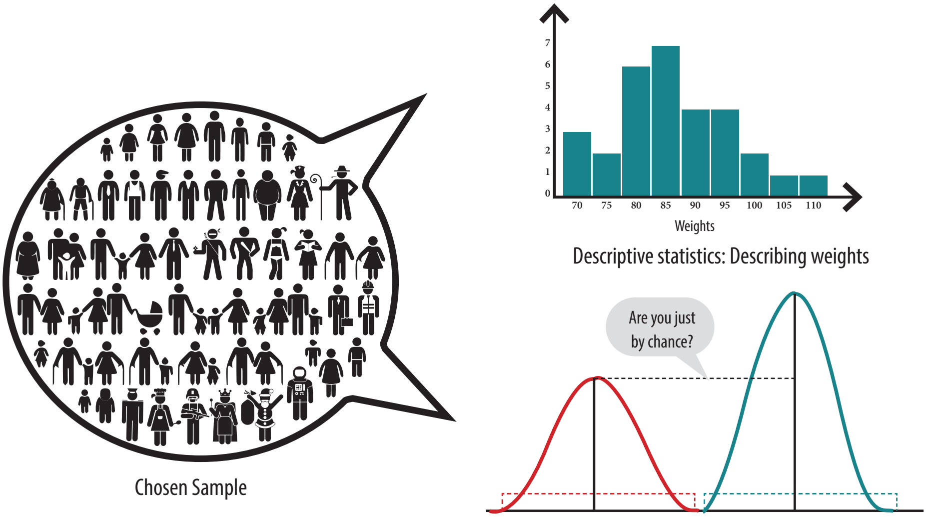
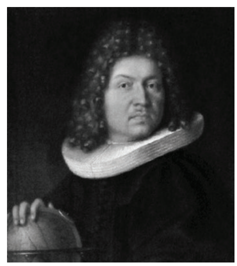


Figure 1.3 Descriptive versus Inferential statistics.

In this cross-sectional observational study, 627 athletes from rugby (n = 225), soccer (n = 172), combat sports (n = 86), handball (n = 82) and water polo (n = 62) were recruited at different sports training centres and competitions. The mean age was 25 ± 6 years, and 60% of the athletes were male […]. The MSK-I prevalence among all athletes was 76%, with 55% of MSK-I occurring in a joint, 48% occurring in a muscle and 30% being tendinopathy, and 19% of athletes had three investigated injuries. There was a predominance of joint injury in combat sports athletes (77%), muscle injury in handball athletes (67%) and tendinopathy in water polo athletes (52%).

Did You Know?



© Public domain [5]

Figure 1.4 Random events play a large role in our lives. **The law of large numbers** is central to understanding probability. It was first described by a Swiss mathematician **Jacob Bernoulli** (1655–1705). It states that if the same experiment is repeated manifold, the average of results gets closer to the real mean.

Inferential Statistics

When we observe a characteristic, it is seldom possible to investigate everyone with it. Instead, we choose a representative sample. Therefore, we need tools to be able to infer from our limited observational values to the unknown but much larger total population. By employing the tools available in **inferential statistics**, we can predict how everyone else with the same characteristic is likely to behave.

We can estimate the unknown population parameter, the plausible range of this parameter, if it differs between two

https://doi.org/10.1017/9781108973663.004 Published online by Cambridge University Press

Bullet Points

Descriptive statistics:

to summarise and present the observations.

Inferential statistics:

-to make inferences about the larger population from the available observations.

-to estimate the unknown population parameter.

-to calculate the plausible range of the effect size.

-to estimate the probability that differences in observations might be due to chance.

populations, the probability that chance or random variation might have affected the results, and how a particular characteristic is influenced by others, etc. Since we make inferences from a limited sample to a larger population, there is **always** a degree of **uncertainty** in **inferential statistics**. If we take the example of measuring weights, we shall employ inferential statistics to help us infer regarding the likely average weight of the whole population of interest, who is most at risk of obesity, and what we can do to prevent it etc.

For example, let's refer back to the above study [4]. Ideally, to investigate the prevalence of MSK-I one should ask every athlete if they have ever had MSK-I but this is not feasible. Therefore, the authors recruited a sample and recorded their observations. From the observations, the authors were able to make inferences. Here is more from the same abstract [4]:

> Age (≥30 years) was positively associated with joint (OR = 5.2 and 95% CI = 2.6–10.7) and muscle (OR = 4.9 and 95% CI = 2.4–10.1) injuries and tendinopathy (OR = 4.1 and 95% CI = 1.9–9.3) […] The analysis of associated factors (epidemiological, clinical and sports profiles) and the presence of MSK-I in athletes suggests an approximately 4-5-fold increased risk for athletes ≥30 years of age.

The authors undertook statistical tests and were able to conclude that age ≥30 was the most important risk factor for MSK-I. They also calculated the likely magnitude of the risk (OR: Odds Ratio). These tests allowed them to make educated guesses regarding the unknown population parameter that we would likely find if we studied the whole population. They also calculated the accuracy of their estimate (95% CI: Confidence Interval), and by calculating the p-value, the authors were able to estimate the probability that the observed differences between different groups of athletes could be affected by chance alone.

Think About It!

What is the difference between descriptive and inferential statistics?

The first step in our journey to understanding medical statistics is to appreciate that there are differences in individual characteristics that dictate what kind of statistical tests we perform. Let's begin in the next chapter.

https://doi.org/10.1017/9781108973663.004 Published online by Cambridge University Press

Take Home Messages

- Learning medical statistics enables us to:-

  - Make sense of data.
  - Communicate results effectively.
  - Understand the language of researchers.
  - Make up our minds regarding the effectiveness of an intervention.
  - Be an evidence-informed physician.

- Medical statistics is divided into two branches: descriptive statistics and inferential statistics.

- Descriptive statistics are necessary to summarise and present collected data.

- Inferential statistics allow us to make inferences from the collected data about the population at large even though we may not have observed the whole population of interest.

https://doi.org/10.1017/9781108973663.004 Published online by Cambridge University Press

Questions & Answers

Q: What does Figure 1.2 suggest?

A: This graph plots pre- and post-treatment disease state on the *x*-axis and disease severity on the *y*-axis. One can conclude that patients in the trial got better regardless of treatment. However, the magnitude of improvement was highest with active treatment, followed by that of placebo effect compared to no treatment alone. Therefore, when we observe improvement in disease activity, it is essential to appreciate that patients can get better without any treatment in the natural course of events. However, the margin of difference between the therapeutic effect, the placebo effect, and spontaneous improvement due to the natural history of the disease may not be the same. We plan to revisit the placebo effect in Chapter 9.

Q: Apart from numbers, can descriptive statistics be relayed in any other fashion?

A: Yes, descriptive statistics are often relayed visually by bar charts, pie-charts, histograms etc.

Q: What is the limitation of descriptive statistics?

A: The limitation of descriptive statistics is that we cannot make any inferences from the results or understand the underlying issues. For example, looking back at the previous study we understand that joint injury was most common and that athletes engaged in combat sports, handball and water polo were most susceptible to injury in their sample. We do not know if the results might be true for the general population [4].

Q: Why don't we discuss uncertainty while relaying descriptive statistics?

A: Descriptive statistics are only a summary of the observed sample; there is no uncertainty in this information.

Q: What is the difference between descriptive and inferential statistics?

A: Descriptive statistics are a summary of the observed sample. Inferential statistics attempt to make inferences from the results of the available sample to the broader population.