A Simple Explanation of Statistical Data for Product Managers

by Neil Dickinson

About the writer: Neil Dickinson is Senior Science Writer at Dice. Neil has been involved in the pharmaceutical industry for nearly 40 years, working with all the major companies and across dozens of famous brands.

Have you ever wondered what the statistics actually mean in the data that support the core claims made on behalf of your brand? Or how percentage differences between the effects of an active drug and a control, such as a placebo, are calculated?

As medical copywriters, we are asked to explain these calculations often enough to suggest that there may be widespread misunderstanding out there. That's understandable and nothing to be embarrassed about: after all, we are marketers not mathematicians.

So we have prepared this little information leaflet that gives you the VERY BASICS.

Statistics can be a tough subject to get your head around: the double negatives of disproving a null hypothesis is just the start! So while this leaflet is not as detailed or complex as you'd find in university material, it may help you to grasp the important basics.

1. Calculating the percentage difference of effect

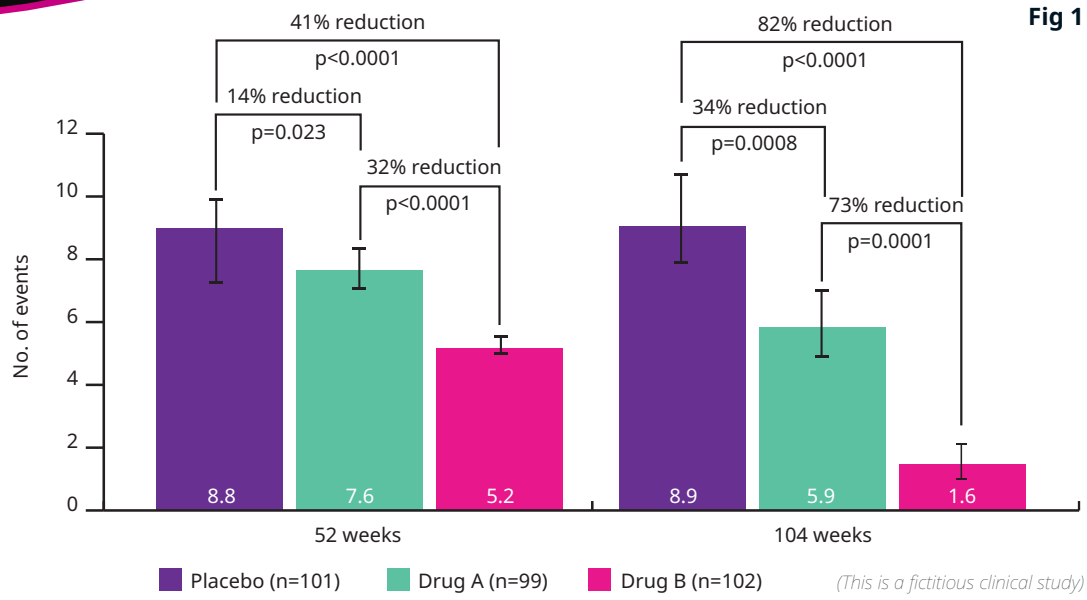


Fig 1

In this chart, there are several claims for % reduction in the number of events in patients receiving placebo compared with an active drug. Here's how the 82% reduction on the right hand chart is calculated:

- 8.9 – 1.6 = actual reduction of 7.3. You then divide this number by the number you are actually comparing the effect with, in other words the placebo number of lesions (8.9).

- 7.3 ÷ 8.9 = 0.8202... (x 100 for the percentage) = 82.02...% (rounded down to 82%)

Try it yourself for the 73% reduction on the right hand chart or all the figures on the left hand side.

1a. Rounding up or down

Remember the conventional rule: if a decimal number is 0.5 or over then it can be rounded up to the nearest whole number (e.g. 5.62 can be rounded to 6); if it's less than 0.5 then it can only be rounded down to the nearest whole number (e.g. 5.12 can be rounded to 5).

So in the examples above, the actual percentages on the right-hand graph are 82.02%, 72.88% and 33.71%. According to the conventional rule these are rounded to 82%, 73% and 34%.

1b. Watch out for calculation differences in the published data

It's not infrequent to find that percentages presented in charts in published papers seem to be incorrect when you check them using the methods just described. The most likely reason for this is that the figures used by the authors in the calculations are not the rounded numbers you see in the chart but the more precise numbers from their raw data.

An alternative reason for the calculation differences could be human error; simply, the author made a mistake in their calculations. Without having access to the detailed data, a decision has to be made as to whether or not to use the published numbers or those calculated by you – and that's a discussion to have with your medics.

2. P values

2a. What are p values and why are they useful?

In a nutshell, p values give you a snapshot of the relevance of a data set and allow you to make a quick decision about the "significance" of any differences that might be suggested by the data. This means they try to identify if something is true or if it happened by chance.

2b. Why is p<0.05 considered significant?

This has become a widely accepted value (largely down to 'tradition') and it is based on some sound reasoning. A p value of <0.05 means that a certain result suggested by the data has less than 5% (or 1 in 20) chance of simply being 'down to chance'. In other words, it is very likely that the results are a genuine demonstration of cause and effect.

To demonstrate, try a simple coin toss. If you were to toss a coin and get heads three times in a row, you might smile at the vaguely unusual result, but you wouldn't find it remotely extraordinary and you certainly wouldn't suspect the coin of being weighted unfairly.

In fact the odds of three heads in a row are 1 in 8 (p=0.125). What about four in a row? The odds are now 1 in 16 (p=0.0625), but you probably still wouldn't suspect a dodgy coin. You'd still consider the result simply being 'down to chance' because the p values are still greater than 0.05.

Suspicions might start to enter your mind if you had 5 in a row, the odds of which are 1 in 32 (p=0.03125). As you can see, we are now into the p<0.05 territory as described above. So at this point, you may well be right to start questioning the coin and the more consecutive heads you get, the odds get longer and the p value gets smaller. The smaller it gets, the more you have cause to doubt that the result of, say, 10 heads in a row, is a result of random chance and you have to start questioning the coin.

So a low p value suggests that your drug really is having a genuine effect that can't be put down to chance and the threshold for that suggestion is p<0.05, which 'feels' about right in the context of the coin tossing example.

2c. What effect does sample size have?

P values are also affected by sample size and the magnitude of effect. Generally, the larger the sample size, the more likely the study will find a significant relationship, if one exists. As the sample size increases, the impact of random error is reduced. In the data presented above, the patient numbers are quite high, lending credibility to the data.

3. Relative Risk (RR), Odds Ratios (OR) and Hazard Ratios (HR)

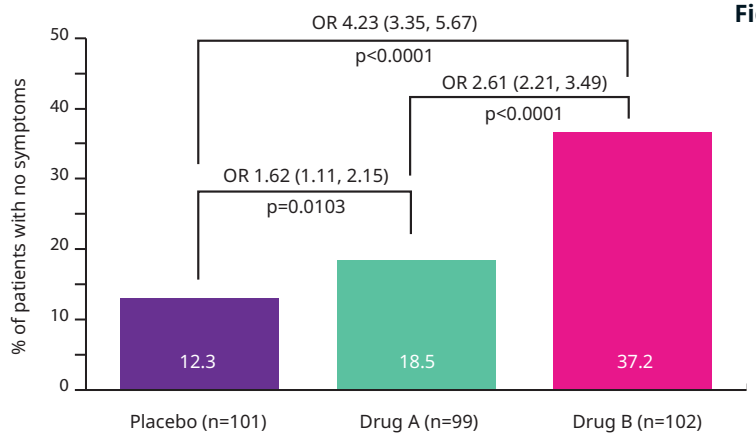
As stated earlier, this short leaflet is not intended to be a crammer for a statistics exam, but merely to help you understand how data that you encounter is calculated. These three terms will be familiar to everyone who has ever looked at some clinical papers, and the differences between them can be quite subtle. I recommend an excellent paper for anyone who wants to delve deeper (see reference 1).[1]

(Note that all the RR/OR/HR numbers that you see in a paper are 'point estimates', which means that they are each a single value representing a range of possible values. The 'true' figure is impossible to know without investigating 100% of the population – and that's why we use statistics!)

3a. So how are they calculated?

The OR is literally a ratio of the odds of an event occurring. In this example an OR of 1.0 means that patients on a drug are equally likely to have no symptoms as the patients on the placebo. An OR of less than 1 means that the patients on a drug are less likely to have no symptoms than those on placebo. An OR of greater than 1 means that the patients on a drug are more likely to have no symptoms than the patients on placebo. The odds are the probability of an event occurring divided by the probability of that event not occurring. (odds = x/1-x)

Fig 2



(This is a fictitious clinical study)

So the odds of patients having no symptoms if taking placebo are calculated as 0.123/1-0.123 = 0.140. For patients on the study Drug A, the odds are 0.185/1-0.185 = 0.227. [In case you're wondering where the numbers 0.123 and 0.185 come from, they are the percentage figures of 12.3 and 18.5 expressed as decimals].

The OR is then calculated as 0.227/0.140 = 1.62. This means that by this analysis, a patient treated with Drug A is 62% more likely to have no symptoms than a patient on placebo.

Alternatively, the authors could have chosen to calculate the relative risk (RR) to drug intervention versus placebo.

RR is the probability of an event happening in one group divided by the probability of the same event in a different group. So in this case, using the same figures as in the OR calculation, RR=0.185/0.123 = 1.50. This result means that the patients taking Drug A are 1.50 times more likely to have no symptoms than the patients in the placebo group.

So the number is similar to the OR but not the same. Generally speaking, an OR can overestimate and magnify risk especially in a very large population or very common disorder, but it is up to the authors of the paper to decide their choice of analysis.

HR is broadly equivalent to RR, but is more useful when the risk is not constant with respect to time. HR uses information collected at different times and that's why you typically see it used in the context of survival over time (the classic Kaplan-Meier curves).

Calculations of HR are similar on a superficial level to RR, but to a non-statistician, the complexities are subtle. But if you see an HR of 0.5, it means that the relative risk of an event in one group is half the risk of the same event in the other group. In fact, more broadly speaking, a hazard ratio of less than 1.0 implies a treatment effect.

4. What about confidence intervals (CI)?

In short, confidence intervals give more context to a statistical analysis than a raw p value or OR result.

"Confidence intervals provide information about statistical significance, as well as the direction and strength of the effect." [2]

The norm in clinical studies is to quote 95% CIs. In the top OR result in Figure 2, it means that there is a 95% probability that the 'true' result will lie between 3.35 and 5.67.

How do CIs help in data interpretation?

For example, if you only provided data from the two left hand bars in Figure 2 you might, at first glance, conclude that taking Drug A seems to provide enough of a clinical effect compared with placebo. The p value is reasonably small (0.0103) and the OR of 1.62 suggests, as stated earlier, that Drug A could offer a 62% improved chance of being free of symptoms than placebo.

However, if you look at the CI for the OR, it ranges from 1.11 to 2.15. This means that the TRUE treatment effect COULD BE as low as 1.11, or just an 11% improvement.

When you look at the bigger picture, if you compare that to the comparison of Drug B to the placebo then it's clear that not only is the OR much more impressive (4.23) and more highly significant (<0.0001), but that even the 'worst case' result from the CI is still an impressive 3.35 – which means that this regime is AT LEAST over three times as likely to make patients symptom free compared with placebo.

If this was a real study, it would clearly indicate that Drug B was more effective than Drug A when they are compared with placebo.

Summary

- P values: the smaller the better, but must be <0.05 to make a claim of significance
- RR/OR/HR: they are all ratios to help demonstrate the difference of effect between groups. UK code recommends that absolute rates are also given wherever possible.
- CI: the range of possible 'true' results either side of the 'point estimate'. The bigger the study population, the tighter the CI range.

References

1. Stare J and Maucort-Boulch D. Metodoloski Zvezki 2016; 13(1):59-67
2. du Prel J-P et al., Dtsch Arztebl Int 2009;106(19):335-9

I would like to thank the team at PharmaReview for their kindness in proof-reading and sense-checking this document.

Level 2, St Cloud Gate
St Cloud Way
Cookham Rd
Maidenhead, UK
SL6 8XD

t: +44 (0)1628 397830
w: dice-comms.co.uk