

This site is intended for healthcare professionals

ADVERTISEMENT

the PHARMACEUTICAL JOURNAL

[\(https://pharmaceutical-journal.com/\)](https://pharmaceutical-journal.com/)

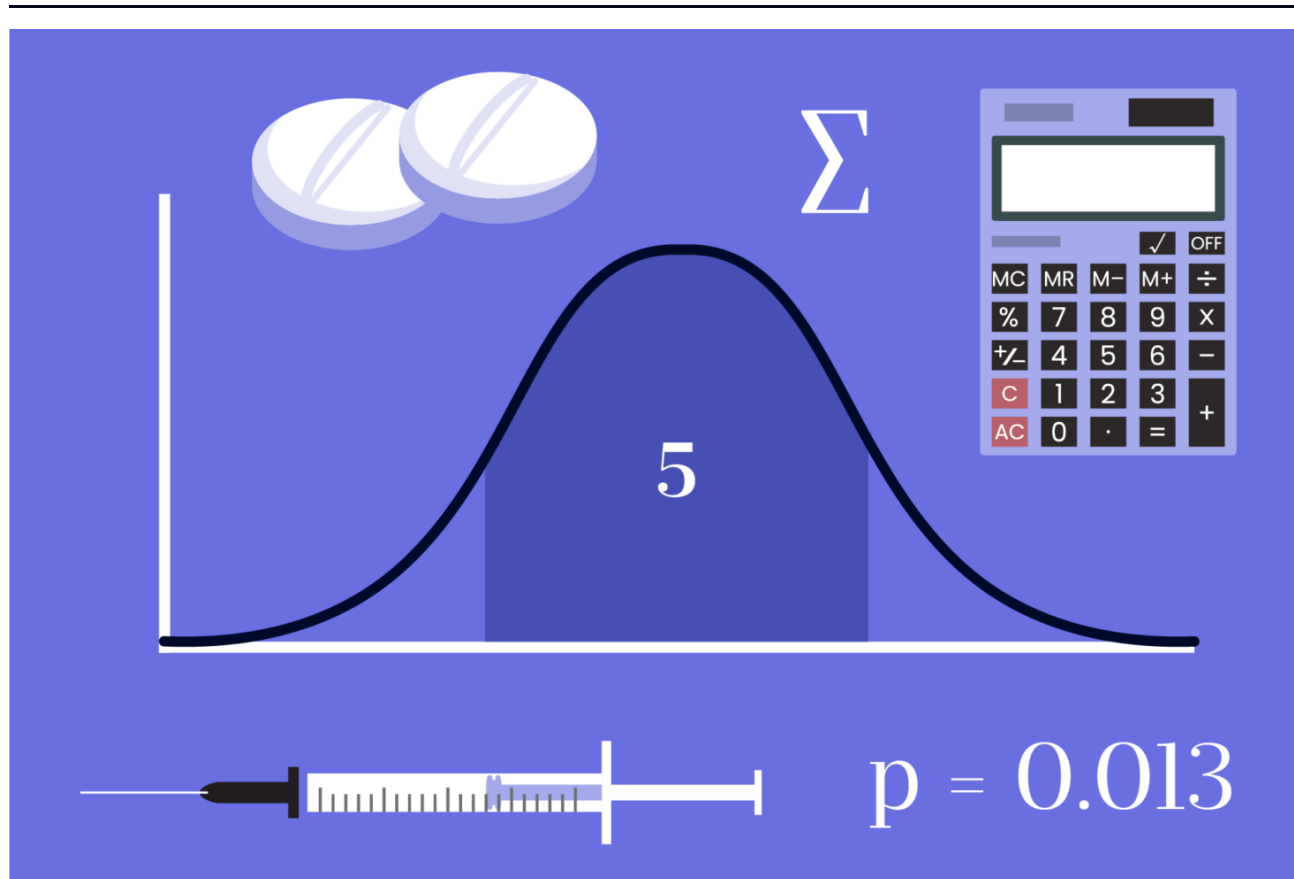
Home (<https://pharmaceutical-journal.com>) / Learning (<https://pharmaceutical-journal.com/article/ld/all>)

Evaluating study results: central tendency and variability and confidence intervals

Part of a comprehensive series showing how to evaluate a clinical study or research paper. This article explains some of the fundamental statistical concepts that can help interpret study data and consider how an entire population might respond

Research and evaluation

11 September 2026.By **Marie Abate** (<https://pharmaceutical-journal.com/author/marie-abate>) .



CHARLOTTE GURR

By the end of this article, you should be able to:

- Describe the measures of central tendency ([#h-measures-of-central-tendency](#)) of data, including mean, median and mode;
- Discuss and interpret important measures of the variability ([#h-measures-of-variability-i-e-spread-or-dispersion](#)) of study data, including standard deviation, standard error of the mean, range and interquartile range and define the term variance;
- Describe confidence intervals ([#h-confidence-interval](#)) and their use.

This article is the part of a comprehensive series exploring how to evaluate clinical studies when addressing information needs using a five-step process:

1. Identifying study type or design;
2. Appraising the journal, authors and study purpose;
3. Critiquing the methods used;
4. Understanding basic statistical tests.
5. Analysing study data and results, and the discussion section;

This article explores the fourth step: understanding basic statistical tests. It is recommended that you read it in conjunction with:

- '[Evaluating study results: statistical inference, hypothesis testing, significance](https://pharmaceutical-journal.com/article/1d/evaluating-study-results-statistical-inference-hypothesis-testing-and-significance) (<https://pharmaceutical-journal.com/article/1d/evaluating-study-results-statistical-inference-hypothesis-testing-and-significance>).'
- '[Understanding risk and clinical utility when evaluating clinical studies](https://pharmaceutical-journal.com/article/1d/understanding-risk-and-clinical-utility-dropouts-data-handling-and-the-discussion-section-when-evaluating-studies) (<https://pharmaceutical-journal.com/article/1d/understanding-risk-and-clinical-utility-dropouts-data-handling-and-the-discussion-section-when-evaluating-studies>).'

To access other articles in the collection visit the [hub page \(https://pharmaceutical-journal.com/information-skills-and-critical-evaluation\)](https://pharmaceutical-journal.com/information-skills-and-critical-evaluation) on *The Pharmaceutical Journal*.

Interpreting a study's data and the significance of the results are critical aspects of deciding whether – and how – the findings should be applied to clinical practice. This article focuses on the study results (i.e. data), including determining how most study patients responded to therapy (i.e. measures of central tendency), and how widely individual responses are spread around that central point (i.e. measures of variability). It is also useful to take a specific finding from a study and predict what its value would be in the entire target population. The confidence interval (CI) supplies a range of values likely to contain the actual population value and will be discussed in this article.

Measures of central tendency

Clinicians want to know how the 'typical' study patient responded to therapy – that is, the data's 'central tendency' – rather than simply review a list of individual patient values. Three measures of central tendency that studies report for their data are the mean, median and mode.

Mean

The mean is calculated as the arithmetic average of a data set. It is determined by adding each of the values in the data set and dividing the sum by the total number of values (n). For example, if serum potassium values (in mEq/L) for 10 patients in a study were: 4.1, 3.1, 5.2, 3.7, 5.1, 3.2, 4.8, 4.3, 3.9 and 5.1. The mean would be calculated as:

$$\frac{4.1 + 3.1 + 5.2 + 3.7 + 5.1 + 3.2 + 4.8 + 4.3 + 3.9 + 5.1}{10} = 4.25$$

Means are commonly reported in studies and provide a useful estimate of the central tendency (i.e. clustering) of continuous-level data. Means have also been used for ordinal-level data, particularly for questionnaire or survey results that use rankings such as 5 = excellent/strongly agree; 4 = very good/agree; 3 = average/neutral; 2 = poor/disagree; 1 = very poor/strongly disagree. When a mean is reported for ordinal data, keep in mind that the distance between numbers (e.g. between 1 and 2, or 4 and 5), is not necessarily equal. As a result, it is difficult to interpret what a decimal really signifies with mean values such as 3.7 versus 3.4 for ordinal-level data.

A problem with the mean is that it can misrepresent the central clustering (i.e. usual/typical value) when reported for data that have outliers, which is defined as a small number of very large or very small data points compared to the rest of the data set. For example, suppose a study measured 10 oestrogen concentrations in women and reported the following values (in pg/mL): 28, 29, 30, 30, 29, 28, 28, 30, 30, and 259. Owing to the outlier, the mean of these data is 52.1. This misrepresents – and is much higher than – the clustering of the rest of the data around the value of 29.

Although investigators will often provide means for ordinal data, use caution when interpreting these results. For example, a mean = 3.26 reported for a 4-point ordinal level Likert scale (e.g. 4 = Excellent; 3 = Good; 2 = Fair; and 1 = Poor) does not have the same mathematical properties as a mean = 3.26 mg/L reported for a continuous-level drug concentration. One cannot assume that 0.26 points above a score of 3 for 'Good' has the same meaning as 0.26 points above a score of 2 for 'Fair'.

Median

The median is the midpoint of a rank-ordered listing of all data points, which is also referred to as the 50th percentile. With an odd number of values, the median is the middle value. With an even number of values, the median is the average of the two middle values. For the oestrogen example above with 10 concentrations, the median is the average of the two middle values, 29 and 30, which would equal 29.5: 28, 28, 28, 29, 29, 30, 30, 30, 30, 259. Note the advantage of the median is that it is not affected by a small number of outlying data points – in this case, the value of 259. Thus, it better represents the central tendency of data that has one or more outliers, which are either very small or very large.

Look for the median as a better indicator of the central clustering of study data for non-normally distributed (i.e. skewed) data. The mean is best used for normally or near normally distributed data. How can you tell if data are non-normally distributed? The mean and median have the same value for normally distributed data – think of a bell-shaped curve. If the median and mean differ substantially, the data are skewed.

Mode

The mode is the most frequently occurring value in a data set. Although in theory it could be reported for continuous-level data, it makes the most sense when reported for data consisting of whole numbers. Why? If a study reports continuous data measured to one decimal place, and all the values are close but slightly different, there would be no mode. Only the mode can be used for nominal data. For example, suppose a study reports systolic blood pressures in patients using the following categories: <120 mmHg (23 patients); 120–149 mmHg (27 patients); ≥150 mmHg (18 patients). The mode is 120–149 mmHg, which is the category with the most (27) patients. A mode can be reported for ordinal data (e.g. if patients rank their satisfaction with therapy on a scale of 0–4 and most patients indicate a '3', 3 would be the mode).

The Table below summarises the measures of central tendency used for each scale/level of measurement.

Table: Use of the measures of central tendency

Measure	Scale/level of measurement
Mean	Ordinal data ^a Continuous data ^b
Median	Ordinal data Continuous data
Mode	Nominal data Ordinal data Continuous data ^c

^aUse caution when interpreting a mean value for ordinal-level data.

^bCan misrepresent central tendency for skewed data; median preferred for skewed data.

^cNot usually helpful here.

Worked example 1: central clustering of data

A study evaluating the efficacy of a herbal Chinese tea extract for treating hyperlipidaemia in 12 patients with diabetes reported that the serum cholesterol concentrations following 8 weeks of therapy were:

- Mean = 220 mg/dL;
- Median = 170 mg/dL.

Which value appears to provide a better estimate of the central clustering (i.e. tendency) of these data?

The median would provide the better estimate of central tendency in this example because the data appear skewed. If the data were normally distributed, the mean and median would be equal or close in value. Since they are substantially different, the data are likely skewed. With skewed data, outliers do not affect the median but will affect the mean.

Measures of variability (i.e. spread or dispersion)

In addition to knowing how a typical patient responded in a study (i.e. central tendency of data), clinicians would like to see how spread out the data (i.e. patient responses) were. Did every patient respond the same as the mean or median value, or were many higher or lower – better or worse – responses present? Measures of variability tell the reader the spread or dispersion of the individual study data points.

Measures of variability commonly reported in clinical trials include:

- Range;
- Interquartile range (IQR);
- Standard deviation (SD);
- Standard error of the mean (SE or SEM).

Range

The range is the spread (i.e. difference) between the highest and lowest value in a data set; although, many authors report it as the lowest and highest values. A related, more useful measure that provides a better indication of where the values fall within the range is the IQR. The IQR includes those individual values that fall between the 25th and 75th percentiles, which comprise 50% of the values. Authors usually report IQR as the values for the 25th and 75th percentiles. For example, suppose a study reported therapy adherence as the percentage of doses taken and listed the IQR as 55–96%. This IQR indicates that 50% of the adherence values were between 55% and 96% (i.e. the 25th and 75th percentiles). Thus, 25% of the values were $\leq 55\%$ and 25% were $\geq 96\%$. Note that outliers would be excluded from the IQR.

Variance

The variance is not usually reported in clinical trials, but you may see this term. It is essentially looking at the differences between each individual data point and the mean value. It is calculated as the average of the difference between each individual data point and the mean value, in which each difference is also squared.

Standard deviation

In contrast to the infrequently used variance, the square root of the variance is called the standard deviation (or variance = SD^2). The SD is one of the most reported measures of variability in clinical studies and gives the spread of the individual study values around the mean value for that measure. For a normal distribution of data, ~68% of the individual values will be found within ± 1 SD; ~95% of the values will be within ± 2 SD; and ~99% of the values will be within ± 3 SD of the mean value.

Key point

- Interquartile range (IQR) is usually reported with the median; these measures are preferred for skewed (i.e. nonnormally distributed) data. The smaller (i.e. closer together) the values in the IQR, the less spread out the individual values are;
- Standard deviation (SD) is usually reported with the mean. These measures are preferred for normally or near normally distributed data. The smaller the SD, the closer the individual study values lie to the study mean;
- If a study reports a mean ± 2.0 (mean $\pm$ SD), this does not indicate that the value of 2.0 represents two standard deviations. It simply indicates that the value of the SD for that measure is 2.0.

Studies will report a mean and SD as: mean (SD) or mean $\pm$ SD. Although it doesn't say '1 SD', the number listed would represent 1 SD. For example, suppose a study reported a mean (SD) age in years for their patients as 54.5 (8.7). In this case, approximately 68% of the study patients' ages – assuming a normal or near-normal age distribution – were between 45.8 years (54.5 – 8.7) and 63.2 years (54.5 + 8.7). Note that ~34% of the ages would be found on each side of the mean for 1 SD. About 95% of the study patients' ages would be between 2 SD or between 37.1 years (54.5 – 17.4) and 71.9 years (54.5 + 17.4). Remember that the percentages indicated above apply to normally or near normally distributed data. If the data are skewed, the SD cannot be interpreted in this way.

Worked example 2

Two studies examined whether counselling diabetic patients about their medications increased blood glucose control. The patients' mean (SD) fasting blood glucose concentrations following counselling were:

- Study 1: 160 (31) mg%;
- Study 2: 158 (45) mg%.

Which study, 1 or 2, reported greater variability in individual patient responses following therapy?

Study 2 shows greater variability of individual patient responses since the reported SD is larger. Assuming normally or near normally distributed glucose concentrations, ~68% of the individual patients' concentrations in study 1 were within 129–191 mg%, compared to 113–203 mg% for study 2.

Worked example 3

Investigators determined the efficacy of a new potassium supplement on serum potassium concentrations in 20 patients taking a diuretic who had hypokalaemia. Following 2 weeks of therapy with the new supplement, the serum potassium concentration was 4.1 ± 0.3 mg% (mean $\pm$ SD).

Interpret the meaning of the italicised statement.

The mean (i.e. arithmetic average) of the 20 individual patients' potassium levels was 4.1 mg%, which provides an indication of the 'central tendency' or 'typical' value for the data. The SD is 0.3 mg%, which indicates that ~68% of the individual patients' potassium concentrations were within 3.8–4.4 mg% (calculated from the mean of $4.1 - 0.3$ and $4.1 + 0.3$). This assumes that the data are normally or near normally distributed.

Standard error of the mean is often reported in clinical trials with study mean values; although, in most instances, it should not be reported. The standard error of the mean (i.e. SEM or SE) estimates how much the sample mean would vary if the study were repeated many times using different patient samples from the same underlying population. Because each study performed would include different patients, each study would likely produce a somewhat different overall mean value. If these sample means – from the different individual studies run – were plotted, they would form a distribution around the true – but unknown – population mean. If this distribution is approximately normal, about 68% of sample means would fall within 1 SEM above or below the unknown population mean.

In other words, SEM provides an estimate of the variability or precision of an overall study sample mean value for an outcome measure around the unknown population mean value. Note three important things here:

1. SEM describes the expected variation among study sample means. It is not referring to variability or spread of the individual patients' data within the study (the SD tells us that);
2. A small SEM indicates that the study sample mean is providing a fairly precise estimate of the population mean; and
3. SEM is generally not of much use to clinicians (it is more useful for calculating the values in a CI, see below).

The SEM is calculated from the SD as follows: $SEM = SD/\sqrt{n}$. The calculated SEM is always a smaller number than the SD. The problem with SEM is that clinical studies will sometimes report SEM instead of SD with a mean value, simply because SEM is a smaller number. Clinicians quickly reading a study without paying much attention to the abbreviation used (i.e. SD vs. SEM) might automatically assume they are seeing a SD with the mean instead of the SEM.

Key points

Use the SD to determine the spread of individual study patient responses around a mean value. If a clinical study only reports a SEM value, always calculate the SD. Why? Because SD provides more clinically useful information for applying findings to patients outside the study.

Why would investigators report SEM instead of SD? Probably to make the spread of individual study responses appear small (i.e. closer around the mean) than it actually is. If a study reports a mean $\pm$ SEM, consider whether there are any potential conflicts of interest present. For example, a pharmaceutical manufacturer that performed or sponsored the study would ideally want patients to respond to their drug in as consistent and uniform a manner as possible (i.e. as illustrated by a small SD). A large SD means there is substantial individual patient variability in drug response (i.e. individual responses could be much larger or smaller than the mean value). Thus, some patients could have a great response to therapy while others have little or no effect, or, perhaps, are even worse. To clinicians, this indicates that their patients in practice could respond unpredictably to the drug. As a result, they might hesitate to prescribe the drug — not what the manufacturer would desire.

Is there an important use for SEM? Yes, SEM is used to calculate the CI. As a clinician, use the CI instead to estimate the underlying population value (see '[CI \(#h-ci\)](#)').

Worked example 4: reporting the change from baseline

A study examined the efficacy of a new amlodipine + diuretic combination for hypertension treatment. A total of 200 patients were in the 16-week study. At the end of week 16, the investigators reported a mean (SEM) change from baseline in systolic/diastolic blood pressure as $-18.7 (0.8)/-10.9 (0.5)$ mmHg.

Was it appropriate for the SEM to be reported? Would standard deviation (SD) be better? If so, what would the SD be for this finding?

No, SEM should not have been used here. It does not provide the variability (i.e. spread and dispersion) of individual patients' study data. The SD is more useful for clinicians, who generally want to know how widely spread the individual patients' blood pressure responses were to the therapy. In many studies that inappropriately report SEM, there is a potential conflict of interest (e.g. manufacturer funding the study). The investigators may want the study variability to look small to clinicians who might quickly scan a study's results.

It is important to calculate the SD any time SEM is reported with the mean in a clinical study.

In this example, the SD can be calculated from the following equation:

$$\text{SEM} = \text{SD} / \sqrt{n}.$$

$$\text{Thus, SD} = \text{SEM} \times \sqrt{n}.$$

The SD for the systolic blood pressure would be $0.8 \times \sqrt{200} = 0.8 \times 14.1 = 11.5$. Notice what a difference it makes to see $-18.7 (11.5)$ versus the $-18.7 (0.8)$ reported. The SD for the diastolic blood pressure = $0.5 \times \sqrt{200} = 7.05$.

Confidence interval

We have covered commonly used measures of central tendency and variability to describe a study's data. It would also be useful to take a specific finding from a study and predict what its value would be in the entire target population (i.e. outside the study). Is there a measure that can do that? Yes, the CI supplies a range of values likely to contain the actual but unknown population value.

Let us take two examples to illustrate CIs. Suppose an election for a president or prime minister was nearing, and the following was reported: 'A telephone poll of 3,014 voting citizens representing persons from across the country showed candidate A had a 5% lead over candidate B, with 54% favouring candidate A and 49% favouring candidate B (i.e. 5% difference)'. Can we say with 100% certainty that in the actual election with all voters, candidate A would beat candidate B by exactly a 5% difference?

Now, suppose we take the election example and substitute a typical clinical trial. Let us simply replace the election wording with a clinical trial summary as follows.

Suppose a study compared the efficacy of two drugs and the following was reported: 3,014 patients were randomised to receive either drug A or drug B. The results showed that drug A had a 5% greater efficacy than drug B, with efficacy reported in 54% of patients taking drug A and 49% of patients taking drug B (i.e. 5% difference). Can we say with 100% certainty that in the patient population outside the study, drug A would have exactly 5% greater efficacy than drug B?

The answer for both examples is no. Why not? Because in the first example – even if the telephone poll truly represented the diversity of all voters – the poll still only looked at a sample of voters. The same is true for the drug efficacy study: only a sampling of patients from the population of interest, which is defined by the study's inclusion and exclusion criteria, participated in the study. In both examples, wouldn't it be nice to know, let's say with 95% certainty, what the election result or drug efficacy difference would likely be in the population outside the study? Is there a measure that could provide this information? Yes, and the CI is such a measure.

The CI provides the likelihood of what the population value would be for a specified study finding. The CI is usually stated as a 95% CI – occasionally a 90% or 99% CI is used – followed by a range of values likely to include the actual population value. A 95% CI is saying that one is 95% confident that the range of values provided will contain the population value for the specified outcome measure. With a 95% CI, there is still a 5% likelihood that the population value is completely outside the range given, but this is considered acceptable.

For example, suppose a study states the following: 'The clinical cure rate was 91.8% (89 of 97 patients) for levofloxacin compared with a cure rate of 82.4% (84 of 102 patients) for ciprofloxacin (9.4% difference between drugs; 95% CI = 2.1–16.8%)'. This CI is interpreted as follows: there is 95% confidence that in the population, the cure rate with levofloxacin will be from 2.1% to up to 16.8% percentage points higher than ciprofloxacin. The CI calculation is designed so that if individual studies are repeated many, many times, each using a different sample from the population, approximately 95% of the resulting CIs from those studies would contain the true population difference.

Key points about CI

Can't we simply take the actual findings from the study sample and apply those to clinical practice outside the study? Why is it important to look at the CIs reported?

Remember that the findings reported for the outcome measures in a study are from the sample, which consists of only those subjects in the study. The study sample is a relatively small group enrolled from the entire population of potential subjects. The population consists of all those individuals who would meet the inclusion and exclusion criteria specified in the study. In clinical situations, it is essentially impossible to study the whole population. Thus, a finding reported in a study might – but might not – reflect the actual value for that measure in the population. This is why the CI is very useful and should be referred to when analysing a study's findings. The CI provides a range of values that could reasonably represent the actual population value – at a specified level of confidence – based on the study findings.

What types of data can a CI be reported for?

A CI can be calculated for nominal- or continuous-level data. In a clinical study, it can be determined for efficacy rates or other outcome measures within a group, as well as for differences between groups.

How should a CI with a negative value be interpreted?

Since the CI provides a range of values likely to contain the actual (i.e. unknown) population value, negative numbers in the CI simply mean that the population value might represent a decrease for whatever that outcome measure was. Let's take the election/drug example from earlier and include the following CI: 'The results showed that Drug A had greater efficacy than Drug B, with efficacy reported in 54% of patients taking Drug A and 49% of patients taking Drug B (5% difference, 95% CI of -3% to 13%)'. This CI is given for the difference in efficacy between Drugs A and B, with a 5% difference reported in the study (Drug A efficacy [54%] – Drug B efficacy [49%] = 5% difference). The negative CI value of -3% indicates that Drug A efficacy might be 3% less than Drug B efficacy, a value of 0 would indicate that both drugs could have the same efficacy – one subtracted from the other would give a value of 0 – and the positive values indicate that Drug A could have up to 13% greater efficacy. Thus, this CI would be interpreted as: 95% confident that the population value for the difference in efficacy between Drug A and Drug B could range from 3% less efficacy with Drug A to 13% greater efficacy with Drug A compared to Drug B.

How should one interpret a very wide CI versus a very narrow CI?

A very wide CI means that the study estimate is not very precise, and the actual population value might assume any of the values in that range. For example, suppose a drug was found to be efficacious in 58% of patients, with a 95% CI = 18–98%. Population response rates from very low (18%) to very high (98%) are reasonable based on the study finding. Therefore, it is difficult to draw firm conclusions about the drug's clinical efficacy in practice.

Is it helpful if all the values in the CI range are large enough to be clinically important?

Yes. Let's look at two examples to illustrate.

Study 1 examines the difference in efficacy between two drugs (i.e. Drug A and Drug B) and finds a difference in the percentage efficacy rate (i.e. Drug A efficacy minus Drug B efficacy) of 35%, 95% CI for the difference = 30%–40%, $P < 0.0002$. This is statistically significant with 95% confidence that the population difference in efficacy between the drugs is likely to be from 30–40% greater with Drug A.

Study 2 examines the difference in efficacy between two drugs (i.e. Drug C and Drug D) and finds a difference in the percentage efficacy rate (i.e. Drug C efficacy minus Drug D efficacy) of 35%, 95% CI for the difference = 2%–68%, $P < 0.014$. This is statistically significant with 95% confidence that the population difference in efficacy between the drugs is likely to be from 2–68% greater with Drug C.

Which study's findings would be easier to apply clinically in practice? Study 1. Both studies found the same difference in percentage efficacy between the drugs (35%) in their study sample. However, all the CI values in the Study 1 CI range appear very clinically relevant, while in Study B, there are rather small population efficacy differences – perhaps only 2%, 3% or 4% – reflected in the CI range between the drugs. In summary, if all values in the CI range are large enough to be clinically important in practice, you can reasonably conclude that the treatment effect is likely to be clinically beneficial.

What factors affect the width of the CI range?

Several factors affect the width of the CI range:

- Level of confidence selected (90% CI narrower, 95% CI wider, 99% CI widest) – since there is less confidence with a 90% CI that the actual population value is within that range of values, the range will be narrower. Conversely, if one wishes to be very confident that the actual population value is reflected in the CI range of values, the 99% CI will provide the widest range of values;

- Sample size – as the sample size increases in a study, the better the population will be represented by the study sample and the narrower the CI range of values will be. Conversely, the smaller the sample size, the less well the population will be represented in the study sample, and the wider the CI range of values will be;
- SD of study sample – the greater the SD (i.e. individual variability of patient values) in a study, the more difficult it will be to predict the population value and the wider the resulting CI.

Worked example 5: confidence intervals (CIs)

A study compared the efficacy of green tea extract (64 patients) with placebo (56 patients) for the treatment of hypertension.

Patients receiving green tea extract had a mean decrease in diastolic blood pressure from baseline to the end of therapy of 3.1 mmHg (95% CI = 1.0–5.2 mmHg).

Interpret the meaning of this 95% CI

One is 95% confident that the mean decrease in diastolic blood pressure with green tea extract use in the population (i.e. those individuals outside the study who meet the study's inclusion/exclusion criteria) might be as small as 1.0 mmHg or as large as 5.2 mmHg. Since most of the values in this CI range are small, clinicians can conclude that the efficacy of green tea extract in the population will probably not be of clinical importance.

Suppose this study had reported a 90% CI instead. Which would describe the width of the 90% CI compared to the 95% CI?

- The same width;
- Wider than the 95% CI or;
- Narrower than the 95% CI.

A 90% CI would be narrower than the 95% CI — all else being equal. Since there is less confidence that the population value would be included in a 90% CI vs. a 95% CI range, the range of values will be smaller with a 90% CI.

Suppose this study enrolled 200 patients in the green tea extract group instead of 64 patients. How would this increased number of patients affect the width of the 95% CI reported?

The CI would be narrower with 200 patients because the greater the number of patients that a study enrolled, the better the study sample will represent the underlying population. Thus, the CI range estimated to contain the population value will be more precise (i.e. narrower)¹.

A summary of key points and how to apply the information from this article to practice follow.

Key points

- The mean and median are the two most commonly used measures of central tendency to indicate a 'typical' patient's response to therapy. The mean is often used for ordinal level data although, it is difficult to interpret decimals for non-whole numbers – use caution in this instance;
- When a mean is reported, look for the SD to provide an indication of the spread of the individual patients' responses around the mean for that measure. A small SD indicates that most of the individual responses were tightly clustered around the mean;
- When a median is reported, look for an IQR to provide an indication of the spread of the individual patients' responses around the median for that measure. A narrow IQR indicates that 50% of the individual responses were within a small range of values between the 25th and 75th percentiles;

- SEM should not be used as a measure of variability or spread of the data in clinical studies – SD should be provided instead. SEM is always smaller than SD, so investigators will sometimes report SEM to minimise the apparent variability of individual patient responses in their study;
- The CI should be used to provide readers with a range of values likely to contain the actual population value for that measure, at the level of confidence specified;
- With a 95% CI, there is still a 5% likelihood that the range of values reported will not contain the actual population value – this is an acceptable level of confidence;
- The CI width is affected by the level of confidence selected (90% versus 95% versus 99%), the sample size and the underlying SD for that measure.

How to apply to practice

- The median should be used instead of the mean when providing the central tendency (e.g. clustering and typical value) for skewed data (e.g. when both the median and mean are given and appear substantially different);
- Look at the value of a SD or IQR provided. A wide spread indicates the study patients had substantial variability in their responses to that outcome measure. This makes it more difficult to predict how patients outside the study would respond to therapy;
- Always convert any SEM values reported in a study to SD, using the equation:
 $SEM = SD/\sqrt{n}$ or $SD = SEM \times \sqrt{n}$.
 Use the SD to examine the variability (i.e. spread) of individual patient responses around the mean;
- When SEM is given in a study, consider whether there are any potential conflicts of interest for investigators that might explain why they chose to use that measure instead of the SD;
- Always use the CI to indicate what the population value is likely to be. If all the values within the CI range are likely to be clinically important in practice, it is more likely that the finding will be of clinical significance in the population;
- If a CI includes the value that indicates no treatment effect (0 for a treatment difference or change in outcome measure; 1 for a relative risk or odds ratio), that measure is not statistically significant.

Self-assessment questions

QUESTION 1

Which of the following measures is also described as the ‘50th percentile’?

A: Mean

B: Median

C: Mode

QUESTION 2

A study reports that the median blood concentration in persons who overdosed on drug X was 230 mg/L, with a mean concentration of 398 mg/L. What conclusion should be drawn about the data given the median and mean concentrations reported? What measure of variability should be reported in this circumstance?

QUESTION 3

Study A reports that the median estrogen concentration in 100 women using an oral contraceptive (OC) is 256 ng/mL and the mean concentration is 380 ng/mL. Study B similarly reports that the mean estrogen concentration in 150 women using the same OC is 380 ng/mL but does not report the median. Which of the following statements can be correctly concluded from these data?

A: If a few extreme outlying concentrations existed in study A, only the median was unaffected by them.

B: It was not appropriate for study A to report a median value for continuous-level data.

C: The individual patients in study A had estrogen concentrations that were similar to those in the patients in study B.

D: The individual patients in study A had estrogen concentrations that were very different than those in the patients in study B.

QUESTION 4

A study looked at the effects of the addition of an angiotensin-converting enzyme inhibitor (ACEI) to a low-sodium diet on proteinuria in diabetic patients. After 8 weeks of therapy, it was found that the addition of the ACEI reduced the extent of proteinuria by 49% (95% CI = 32–66%). Interpret the meaning of the statement in bold – include the numbers provided in the statement in your answer.

ANSWER QUESTIONS 5 AND 6 ABOUT THE FOLLOWING TWO PUBLISHED STUDIES:

Study A: 100 patients received a new cholesterol-lowering drug, and a cholesterol concentration during therapy of $185 \text{ mg}\% \pm 5 \text{ mg}\%$ (mean $\pm$ SEM) was reported.

Study B: 100 patients received this same new drug, and a cholesterol concentration during therapy of $185 \text{ mg}\% \pm 18 \text{ mg}\%$ (mean $\pm$ SD) was reported.

QUESTION 5

Which study had the greatest individual patient variability in serum cholesterol concentrations?

A: Study A

B: Study B

QUESTION 6

Explain precisely and completely what the $\pm 18 \text{ mg}\%$ refers to in Study B.

QUESTION 7

Toradol was compared with naproxen for treating acute low-back pain. Complete pain relief was experienced in 58% of patients receiving Toradol and 46% of patients taking naproxen (difference = 12%, 95% CI = –7% to 31%). The degree of back tenderness and the extent of flexibility were also measured. Which of the following is correct concerning the bolded term?

A: 95% confidence that the difference in pain relief between Toradol and naproxen in the study patients fell between –7% and 31%.

B: 95% confidence that the population difference in pain relief between Toradol and naproxen could range from 7% fewer patients to up to 31% more patients experiencing pain relief with Toradol.

C: 95% of the Toradol patients and naproxen patients experienced pain relief from 7–31% of the time.

QUESTION 8

Would the CI reported in question 7 be statistically significant? Yes or no?

Answer guidance

QUESTION 1

B: median.

QUESTION 2

The data appear skewed with outliers present since the mean is much higher than the median concentration. With skewed data, the interquartile range (IQR) is an appropriate measure of variability since it excludes the lower 25% and upper 25% of concentrations, which are likely to contain the outlying data points.

QUESTION 3

A: The data appear skewed since the mean and median concentrations are different, and only the median would be unaffected by a few outliers. Medians can be reported for continuous-level data so answer b is incorrect. No measure of variability (e.g. SD) was reported with the mean value in either study, so there is no way of knowing the variability of individual patient concentrations. Thus, answers c and d are also incorrect.

QUESTION 4

The bold statement is a 95% CI. It indicates that one is 95% confident that, in the population, the reduction in proteinuria in patients with diabetes with the use of an ACEI and a low-salt diet would range from 32% to 66%. The population would be those individuals outside the study who meet the study's inclusion/exclusion criteria.

QUESTION 5

A: study A. SD provides the variability in individual patient responses around the mean value. Thus, the SEM in study A needs to be converted to SD using the equation $SEM = SD / \sqrt{n}$. The SD reported in study A would be: $5 = SD / \sqrt{100}$. $sd = 5 \times 10 = 50$. This SD is much higher than the sd reported in study B, so greater (wider) individual patient variability around the mean value was seen in study A. Keep in mind that SEM should not be used in clinical studies in place of the SD. When used, always convert it to the SD.

QUESTION 6

This is the SD. Assuming the data are normally or near normally distributed, it indicates that ~68% of the individual patients' cholesterol concentrations in the study were within $185 - 18 \text{ mg\%} = \mathbf{167 \text{ mg\%}}$ and $185 + 18 \text{ mg\%} = \mathbf{203 \text{ mg\%}}$. The remaining 32% of the patients' cholesterol concentrations in the study were outside of this range. The mean ± 2 SD provides the range that contains ~95% of the individual study values, and the mean ± 3 SD provides the range that contains ~99% of the individual patient values.

QUESTION 7

B: 95% confidence that the population difference in pain relief between toradol and naproxen might range from 7% fewer patients to up to 31% more patients experiencing pain relief with toradol. The CI provides the range of values likely to contain the population value for that specified measure, at a certain level (%) confidence. In this example, the CI was reported for the difference in complete pain relief between toradol and naproxen ($58\% \text{ complete pain relief with toradol} - 46\% \text{ pain relief with naproxen} = 12\% \text{ difference}$). The CI is saying that in the population of patients outside the study (those persons who would still meet the specified inclusion/exclusion criteria), the difference between the two drugs might not be exactly 12%. But we are 95% confident that the difference could be anywhere from 7% less (i.e. it is a negative number) with toradol to up to 31% greater with toradol compared to naproxen.

QUESTION 8

No: The CI contains the value of 0, which means that there could be no difference between complete pain relief with toradol and naproxen in the population. Since the CI contains this value of no difference, it is not statistically significant.

Acknowledgements

This article was adapted from *Drug Information and Literature Evaluation, Second Edition*, previously published by Pharmaceutical Press.

A full list of resources and materials used to prepare the book can be accessed from the [bibliography \(https://pharmaceutical-journal.com/drug-information-and-literature-evaluation-second-edition-bibliography\)](https://pharmaceutical-journal.com/drug-information-and-literature-evaluation-second-edition-bibliography) page.

1. Swinscow T, Campbell M. *Statistics at Square One*. 10th ed. BMJ Books; 2002.

Last updated 21 September 2026 15:50

Citation

The Pharmaceutical Journal, PJ September 2026, Vol 317, No 8013;317(8013)::DOI:10.1211/PJ.2026.1.428149

Please leave a comment

You must be [logged in \(https://pharmaceutical-journal.com/my-account?redirect=https://pharmaceutical-journal.com/article/ld/evaluating-study-results-central-tendency-and-variability-and-confidence-intervals\)](https://pharmaceutical-journal.com/my-account?redirect=https://pharmaceutical-journal.com/article/ld/evaluating-study-results-central-tendency-and-variability-and-confidence-intervals) to post a comment.

You might also be interested in...



Understanding risk and clinical utility, dropouts, data handling and the discussion section when evaluating studies

(<https://pharmaceutical-journal.com/article/ld/understanding-risk-and-clinical-utility-dropouts-data-handling-and-the-discussion-section-when-evaluating-studies>)



Evaluating study results: statistical inference, hypothesis testing and significance

(<https://pharmaceutical-journal.com/article/ld/evaluating-study-results-statistical-inference-hypothesis-testing-and-significance>)



Pharmacy must protect its scientific identity

(<https://pharmaceutical-journal.com/article/opinion/pharmacy-must-protect-its-scientific-identity>)

ADVERTISEMENT