

## MEDICAL STATISTICS MADE SIMPLE

# Why Medical Researchers Need Statistics

*You already collect the numbers. Statistics is the discipline that turns those numbers into something a reader, a reviewer, and a patient can trust. This chapter is the reason the rest of the book exists.*

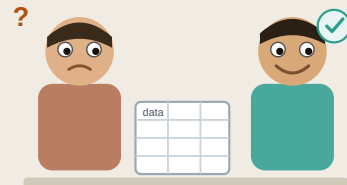
🕒 Opening chapter · Read once, then let it frame every chapter that follows

*"The spreadsheet is full. Two groups, a column of blood pressures, forty patients in each arm. And still the question sits there, unanswered: is the difference I can see on the screen real, or is it noise dressed up as a finding?"*

— The moment every researcher meets statistics, whether they planned to or not

Dr. Sarah Whitfield, a cardiology registrar at a county hospital in Ohio, USA, has just closed a six-month audit of two blood-pressure drugs. Her numbers look clean. Drug A brought systolic pressure down by an average of 12.4 mmHg, Drug B by 8.1 mmHg. She can see the gap. What she cannot say, standing in front of her supervisor, is whether that 4.3 mmHg gap is a genuine effect of Drug A or just the ordinary scatter you would get from any two groups of forty patients. She has the data. She does not know which test tells her what the data mean.

This is not a knowledge gap she should be ashamed of. It is the exact point where every clinical dataset stops being a private impression and has to become a public claim.



*"Same spreadsheet, two reactions: the clinician sees a puzzle, the statistician sees a plan."*

Statistics is the bridge. It is the set of tools that lets you say, with a stated level of confidence, whether the pattern in your patients is likely to hold in the next hundred patients you will never meet.

 **CLINICAL REALITY CHECK — A MISSING STATISTICIAN COSTS PAPERS**

Altman, Goodman and Schroter surveyed authors submitting original research to the *BMJ* and *Annals of Internal Medicine*. Methodological input was reported for **73% of papers**, and research carried out **without any statistical assistance was more likely to be rejected without even being sent out for review**. The lesson for Dr. Whitfield is plain: the analysis is not a formality you bolt on at the end. It is part of whether your work is read at all.

---

*Altman DG, Goodman SN, Schroter S. How statistical expertise is used in medical research. JAMA. 2002;287(21):2817–20.  
doi:10.1001/jama.287.21.2817 · PMID 12038922*

## What this chapter gives you

A clear answer to the question every new researcher secretly asks: why do I have to learn this at all? By the end you will see statistics not as a hurdle between you and publication, but as the process that carries your work from a folder of numbers to an evidence claim a reader can act on. You will meet the difference between data, information and evidence, see where statistics enters at every stage of a study, and watch a small blood-pressure comparison change from a hunch into a defensible result. Chapter 6 then shows you how to pick the exact test; this chapter shows you why the choice is worth making with care.

# 73%

of submitted papers had **methodological input** from a statistician or epidemiologist; work without it was more often rejected before review.

---

*Altman et al. JAMA. 2002. PMID 12038922*

# 47

of 100 papers sent for statistical review used an **inappropriate analysis** — nearly half reached for the wrong test.

---

*Prescott & Civil. Injury. 2013. PMID 23182752*

# 3

levels a study climbs with statistics: **data** → **information** → **evidence**. Each step is a method, not a leap of faith.

---

*Framework after Yan et al. 2020. PMID 29209453*

# 1. Statistics is what turns "I noticed" into "I can show"

Every clinician is already a pattern-spotter. You notice that a ward feels busier on Mondays, that one antibiotic seems to clear infections faster, that a certain patient group keeps returning. Those noticings are the raw material of research. What they are not, on their own, is evidence. The gap between "it seems that" and "we found that" is exactly the gap statistics is built to close.

Dr. Whitfield can see that Drug A lowered blood pressure more than Drug B. The honest problem is that if he repeated the whole audit next month with a fresh forty patients per arm, the averages would not land on 12.4 and 8.1 again. They would wobble. Statistics is the language for describing that wobble and asking whether the gap he saw is bigger than the wobble he should expect. Without it, he is left arguing from a single snapshot, and a reviewer will say so.

## EVIDENCE ANCHOR — THE ERRORS ARE PREDICTABLE, SO THEY ARE PREVENTABLE

Yan and colleagues reviewed the recurring statistical problems in medical and biomedical research and grouped them by stage: experimental design, data collection, data analysis and data interpretation. They concluded that **misunderstanding and misusing statistical concepts and test methods are significant, recurring problems**, driven by ignoring sample size, ignoring the data distribution, choosing the wrong summary measure, and applying the wrong test for the design. Every one of those is a decision you can get right on purpose, which is precisely why the skill is learnable.

*Yan F, Robert M, Li Y. Statistical methods and common problems in medical or biomedical science research. Int J Physiol Pathophysiol Pharmacol. 2020;9(5):157–163. [PMC5698693](#) · PMID 29209453*

# 2. Where statistics earns its place: thesis, trials, audits, publications

Statistics is not a niche skill for people who like numbers. It is a checkpoint you will pass through in every kind of medical work you do. Here is where it shows up, and why it is not optional in any of them.

## 1

### YOUR THESIS

A dissertation examiner will ask why you chose your analysis before they ask what you found.

*The sample size, the test, and the way you handled missing data are all defended in the viva. A weak justification here sinks otherwise good work.*

## 2

### CLINICAL TRIALS

A trial without a pre-stated analysis plan is not a trial, it is a fishing trip.

*Sample size, primary outcome, and the exact comparison are fixed in advance (Chapter 16, Chapter 17). Regulators and journals demand it.*

## 3

## 4

## AUDITS AND QI

"We improved" needs a number, a denominator, and a way to rule out chance.

*A quality-improvement cycle that reports 8 fewer infections says little until you know out of how many, and whether the drop is real.*

## PUBLICATIONS

Peer reviewers read the methods and results with a statistical eye, every time.

*The wrong test, a missing confidence interval, or a p-value with no effect size are among the first things flagged, and often the reasons for rejection.*

### ⚠ COMMON MISTAKE — TREATING THE ANALYSIS AS AN AFTERTHOUGHT

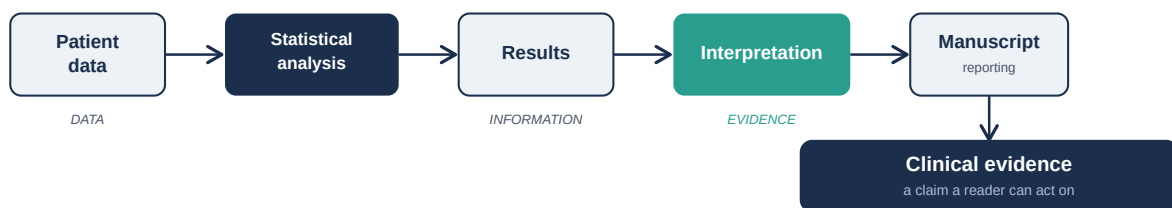
Hesterman and colleagues systematically coded the reasons manuscripts were rejected after peer review at the journal *Headache*. Across submissions from 2014 to 2016, the overall rejection rate was **62.6%**, and problems with study design, analysis and interpretation sat among the leading reasons reviewers turned papers down. Authors who plan the statistics only after the data are collected keep discovering, too late, that the design cannot answer the question they asked.

Hesterman CM, Szperka CL, Turner DP. Reasons for Manuscript Rejection After Peer Review From the Journal *Headache*. *Headache*. 2019;58(10):1511–1518. doi:10.1111/head.13343 · PMID 30011058

## 3. Data, information, evidence: three words that are not the same

People use these three words as if they were interchangeable. They are not, and the difference is the whole job of this book. Each step up the ladder is done by a method, and statistics supplies the methods.

 **FIGURE 1.1 — FROM DATA TO EVIDENCE**



Statistics is the engine at every arrow, not a single step near the end.

*Numbers alone are data. Statistical analysis turns them into information; interpretation turns information into evidence; clear reporting carries that evidence to the reader. Miss any arrow and the chain breaks.*

| LEVEL              | WHAT IT IS                                     | BLOOD-PRESSURE AUDIT EXAMPLE                             | WHAT STATISTICS ADDS                                                           |
|--------------------|------------------------------------------------|----------------------------------------------------------|--------------------------------------------------------------------------------|
| <b>Data</b>        | Raw recorded numbers, no meaning attached yet. | Eighty individual systolic-BP readings, forty per drug.  | Nothing yet — this is the input.                                               |
| <b>Information</b> | Data summarised so a pattern becomes visible.  | Drug A $-12.4 \pm 6.1$ mmHg; Drug B $-8.1 \pm 5.7$ mmHg. | Descriptive statistics: means, SDs, a clear comparison (Chapter 3).            |
| <b>Evidence</b>    | A tested claim with a stated uncertainty.      | Difference 4.3 mmHg, 95% CI 1.6 to 7.0, $p = 0.002$ .    | Inference: a test, a confidence interval, a defensible conclusion (Chapter 5). |

## 4. How statistics is misused, and why that should not scare you off

It would be dishonest to pretend the tool is never abused. It is. But the misuses are a small, repeating set, and knowing them is most of the defence. Prescott and Civil read 100 consecutive papers sent for statistical review at *Injury* and suggested improvements for 90 of them, with an inappropriate analysis identified in 47 and 9 papers reporting a p-value that did not match the method stated. These are not exotic failures. They are the same handful of slips, made again and again.

| COMMON MISUSE                           | WHAT IT LOOKS LIKE IN A PAPER                            | WHERE THE BOOK FIXES IT |
|-----------------------------------------|----------------------------------------------------------|-------------------------|
| <b>Wrong test for the design</b>        | Multiple t-tests across three groups instead of ANOVA.   | Chapter 6, Chapter 9    |
| <b>Ignoring the distribution</b>        | A t-test forced onto clearly skewed data.                | Chapter 4, Chapter 7    |
| <b>p-value with no effect size</b>      | " $p = 0.03$ " reported alone, with no difference or CI. | Chapter 5               |
| <b>Reporting <math>p = 0.000</math></b> | Software output copied verbatim; should be $p < 0.001$ . | Chapter 5, Chapter 21   |
| <b>No sample-size justification</b>     | An underpowered study reported as a negative finding.    | Chapter 16              |
| <b>Counts without denominators</b>      | "8 infections prevented" with no baseline given.         | Chapter 10, Chapter 20  |

### 💡 PRO TIP — BRING THE STATISTICS IN AT THE DESIGN STAGE

The single highest-value habit in medical research is deciding the analysis before you collect a single data point. Altman and colleagues found that statistical input often started too late, at the analysis stage, when the design was already fixed. If Dr. Whitfield had set his sample size and named his test before the audit began, he would have known in advance whether his forty-per-arm design could detect a difference the size of 4.3 mmHg. Planning first is cheaper than fixing later, every time.

## 5. Statistics touches every stage, not just the analysis

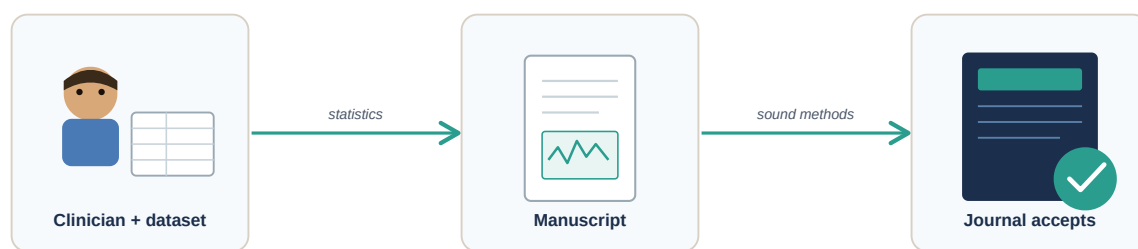
A frequent misconception is that statistics is one box near the end of a study, the bit where you press a button and get a p-value. In reality it runs through the whole project. Get it right early and the later stages take care of themselves; get it wrong early and no clever analysis can rescue you. Worthy's guide for authors is built on the same idea: describe the design and the data first, then choose and report the method that follows.

| STAGE          | THE STATISTICAL QUESTION                                               | IF YOU SKIP IT                                                    |
|----------------|------------------------------------------------------------------------|-------------------------------------------------------------------|
| Design         | How many patients, which outcome, which comparison, what test?         | An underpowered or unanswerable study (Chapter 16).               |
| Analysis       | Is the chosen test right for the data type and distribution?           | A wrong-test result a reviewer will reject (Chapter 6).           |
| Interpretation | What does the effect size and its confidence interval mean clinically? | Confusing statistical with clinical significance (Chapter 5).     |
| Reporting      | Are the numbers, tests and uncertainties all stated clearly?           | An unreproducible, hard-to-review paper (Chapter 19, Chapter 20). |

## 6. Concept in one picture: from clinician to accepted paper

The whole reason to learn statistics is captured in one arc. A clinician holds a dataset. Statistics converts it into a result. Clear reporting turns that result into a manuscript. Sound methods carry that manuscript past peer review into the published record, where the next clinician can use it. Break any link and the patient at the far end never benefits from what you learned.

 FIGURE 1.2 — WHY IT IS WORTH THE EFFORT



The dataset only helps a patient if it survives the whole journey.

*The payoff of statistics is not the p-value itself. It is that a well-analysed, well-reported study is far more likely to be accepted, read, and used in the next patient's care.*

## 7. Worked example: what statistics adds to Drug A vs Drug B

Return to Dr. Whitfield's audit. Two independent groups, roughly forty patients each, systolic blood-pressure reduction in mmHg. Watch the same eighty numbers climb the data-to-evidence ladder.

### STEP 1 — THE DATA, SUMMARISED (INFORMATION)

Descriptive statistics compress eighty readings into a comparison you can see (Chapter 3). Drug A reduced systolic BP by a mean of 12.4 mmHg, Drug B by 8.1 mmHg. The standard deviations, 6.1 and 5.7, tell you how much individual patients scattered around those means. Already the picture is clearer than a raw column, but it is still just a snapshot of these particular patients.

### STEP 2 — THE COMPARISON, TESTED (EVIDENCE)

Because the outcome is continuous, there are two independent groups, and the reductions are roughly normal, the tree in Chapter 6 lands on an independent-samples t-test (Chapter 7). The test asks the honest question: if the two drugs were truly equal, how surprising is a 4.3 mmHg gap in a study this size? The output below shows the answer, together with the confidence interval that matters more than the p-value alone (Chapter 5).

| QUANTITY                              | DRUG A (N = 40)            | DRUG B (N = 40) | COMPARISON                   |
|---------------------------------------|----------------------------|-----------------|------------------------------|
| Mean SBP reduction                    | 12.4 mmHg                  | 8.1 mmHg        | Difference 4.3 mmHg          |
| Standard deviation                    | ± 6.1                      | ± 5.7           | —                            |
| 95% confidence interval of difference | 1.6 to 7.0 mmHg            |                 | Does not include 0           |
| Test                                  | Independent-samples t-test |                 | $t \approx 3.25$ , $df = 78$ |
| p-value                               | $p = 0.002$                |                 | Below 0.05                   |

These are illustrative numbers for a teaching example, computed for a plausible forty-per-arm design; they are not taken from a real trial. What matters is the reasoning. The confidence interval runs from 1.6 to 7.0 mmHg and does not cross zero, so the data are consistent with a real advantage for Drug A of somewhere between about 2 and 7 mmHg. The p-value of 0.002 says a gap this large would be unlikely if the drugs were truly equal. Dr. Whitfield can now write a sentence he can defend, instead of pointing at a spreadsheet.

STEP 3 — HOW TO REPORT IT

 COPY-PASTE REPORTING SENTENCE

"Systolic blood-pressure reduction was greater with Drug A (mean  $12.4 \pm 6.1$  mmHg,  $n = 40$ ) than with Drug B ( $8.1 \pm 5.7$  mmHg,  $n = 40$ ). The difference of 4.3 mmHg (95% CI 1.6 to 7.0) was statistically significant on an independent-samples t-test ( $p = 0.002$ )."

That one sentence states the groups, the effect size, its uncertainty, the test, and the p-value. A reviewer can check every claim in it. That is what statistics adds: not the number, but the accountability behind it.

 Weak claim vs sound claim

REPORTING THE BLOOD-PRESSURE COMPARISON

|  Weak — the reviewer will flag this                                                  |  Sound — statistics doing its job                                                                                             |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| "Drug A worked better than Drug B, lowering blood pressure by more (12.4 vs 8.1 mmHg), so we recommend Drug A." (No test, no uncertainty, a snapshot treated as proof.) | "Drug A reduced systolic BP more than Drug B (difference 4.3 mmHg, 95% CI 1.6 to 7.0, $p = 0.002$ , independent t-test). The effect is modest but consistent, and this audit was not powered for hard outcomes." |

TALKING ABOUT SIGNIFICANCE

|  Weak — confuses the two meanings |  Sound — separates them                                                                                 |
|----------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| "The result was highly significant ( $p = 0.002$ ), so Drug A is a major clinical advance."                          | "The difference was statistically significant ( $p = 0.002$ ). Whether a 4.3 mmHg gain is clinically important is a separate judgement, weighed against guideline thresholds (Chapter 5)." |

8. Questions researchers actually ask

## THE HONEST QUESTION

## THE SHORT ANSWER

**"I am a clinician, not a mathematician. Do I really need this?"**

Yes, but you need the reasoning, not the algebra. Software does the arithmetic; you supply the design decisions and the interpretation.

**"Can I just hand my data to a statistician at the end?"**

You can ask for help, but bring them in at the design stage. Input that starts only at analysis often finds the study cannot answer its own question (Altman et al. 2002).

**"Which test should I use?"**

That is Chapter 6. It falls out of four facts: outcome type, number of groups, pairing, and distribution.

**"My p-value is 0.06. Did my study fail?"**

No. A p-value is not a verdict. Report the effect size and its confidence interval, and interpret honestly (Chapter 5).

**"How many patients do I need?"**

Decide before you start, from the effect you care about detecting. That is sample size and power (Chapter 16).

**"Why do reviewers keep flagging my stats?"**

Usually the same short list: wrong test, missing CI, no effect size, or unclear reporting. This book is organised to remove them one by one.

## Practice tasks (do these on your own project)

**Four tasks.** Outcome: a one-paragraph rationale, in your own words, for why your study needs statistics.

- |                          |                                                                                                                                                            |       |
|--------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
| <input type="checkbox"/> | Write your research question in one sentence, then name the primary outcome you will measure.                                                              | 3 min |
| <input type="checkbox"/> | Take one finding from your data and write it twice: once as a bare "I noticed" impression, once as an "I can show" claim with a number and an uncertainty. | 4 min |
| <input type="checkbox"/> | Place your study on the data-to-evidence ladder (Figure 1.1). Which arrow are you at now, and which step is missing?                                       | 3 min |
| <input type="checkbox"/> | List the four stages (design, analysis, interpretation, reporting) and note the one statistical decision you have not yet made in each.                    | 4 min |

### CHECK YOURSELF

1. In your own words, what is the difference between data, information, and evidence?
2. Dr. Whitfield sees a 4.3 mmHg gap between two drugs. Why can he not conclude Drug A is better from that number alone?

3. Name the four stages of a study where statistics does real work.
4. Give two common ways statistics is misused in medical papers.
5. A study is carried out with no statistical input. According to Altman and colleagues, what happens more often to such papers?
6. A colleague reports " $p = 0.000$ ". What should it read instead, and why?

▼ **SHOW ANSWER KEY**

1. Data are raw recorded numbers with no meaning attached. Information is data summarised so a pattern is visible (means, SDs). Evidence is a tested claim carrying a stated uncertainty (a test, a confidence interval, a defensible conclusion).
2. A single audit is one snapshot; repeated with fresh patients the averages would wobble. He needs a test and a confidence interval to know whether the gap is bigger than the expected scatter (here, 95% CI 1.6 to 7.0,  $p = 0.002$ ).
3. Design, analysis, interpretation, and reporting.
4. Any two of: wrong test for the design, ignoring the distribution, reporting a p-value with no effect size,  $p = 0.000$  instead of  $p < 0.001$ , no sample-size justification, counts without denominators.
5. They were more likely to be rejected without even being sent out for peer review (Altman et al. 2002).
6. It should read  $p < 0.001$ . A p-value is never exactly zero; the software has simply rounded a very small number, and writing 0.000 is imprecise and gets flagged by reviewers.

**CHAPTER SUMMARY · CHAPTER 1**

1. Statistics is the bridge from "I noticed" to "I can show". It is what lets a single dataset become a claim about patients you will never meet.
2. It is not optional in any serious medical work: thesis, clinical trial, audit, and publication all pass through a statistical checkpoint.
3. Data, information and evidence are three different things (Figure 1.1). Each step up the ladder is done by a method, and statistics supplies the methods.
4. The misuses of statistics are a small, repeating set. Nearly half of papers sent for one journal's statistical review used an inappropriate analysis (Prescott & Civil 2013), and the recurring errors are predictable and preventable (Yan et al. 2020).
5. Statistics works across the whole study, not just the analysis: design, analysis, interpretation, and reporting. Bring it in at the design stage, where it is cheapest (Altman et al. 2002; Worthy 2015).

6. Papers without statistical input are more often rejected before review (Altman et al. 2002), and analysis problems sit among the leading reasons for rejection after review (Hesterman et al. 2019).
7. The Drug A vs Drug B example shows the payoff: eighty numbers become one defensible sentence with an effect size, a confidence interval, and a test. Chapter 6 shows you how to choose that test.

---

## REFERENCES · CHAPTER 1

- › Altman DG, Goodman SN, Schroter S. How statistical expertise is used in medical research. *JAMA*. 2002;287(21):2817–20. [doi:10.1001/jama.287.21.2817](https://doi.org/10.1001/jama.287.21.2817) · PMID 12038922
- › Prescott RJ, Civil I. Lies, damn lies and statistics: errors and omission in papers submitted to INJURY 2010-2012. *Injury*. 2013;44(1):6–11. [doi:10.1016/j.injury.2012.11.005](https://doi.org/10.1016/j.injury.2012.11.005) · PMID 23182752
- › Yan F, Robert M, Li Y. Statistical methods and common problems in medical or biomedical science research. *Int J Physiol Pathophysiol Pharmacol*. 2020;9(5):157–163. [PMC5698693](https://pubmed.ncbi.nlm.nih.gov/29209453/) · PMID 29209453
- › Hesterman CM, Szperka CL, Turner DP. Reasons for Manuscript Rejection After Peer Review From the Journal Headache. *Headache*. 2019;58(10):1511–1518. [doi:10.1111/head.13343](https://doi.org/10.1111/head.13343) · PMID 30011058
- › Worthy G. Statistical analysis and reporting: common errors found during peer review and how to avoid them. *Swiss Med Wkly*. 2015;145:w14076. [doi:10.4414/smww.2015.14076](https://doi.org/10.4414/smww.2015.14076) · PMID 25650947