

# A practical guide to statistics for general practice

Winnie Chen

*This paper is part of a series of papers on research.*

**EVIDENCE-BASED MEDICINE** is a well-established part of general practice in Australia.<sup>1</sup> Understanding research is embedded within the current curriculum of The Royal Australian College of General Practitioners (RACGP), with the ability to discuss 'scientific and statistical information' for clinical decisions listed as a required skillset for general practitioners (GPs).<sup>2</sup> In the past few years, the COVID-19 pandemic has further highlighted that interpreting epidemiology and statistics is not only relevant for GP academics, but is also an integral part of clinical care.<sup>3</sup> For example, GPs are often the first point of contact for patients asking about the evidence for masks, diagnostic accuracy of COVID-19 tests, vaccine efficacy and effectiveness of new antiviral treatments.

This article aims to provide a practical guide to interpreting common medical statistics encountered in general practice through two case studies. The abstract and results used here are simplified examples. In clinical practice, framing a research question, conducting a database search and critical appraisal of the selected paper are key first steps in interpreting and using research evidence.<sup>4</sup>

## CASE STUDY 1

### Is this medication effective?

A man aged 71 years with symptomatic COVID-19 qualifies for a newly approved medication X. He asks, 'Will this medication work for keeping me out of

hospital?' You discuss with the patient evidence from a randomised controlled trial (RCT) that was conducted in patients with a similar demographic and risk profile to this patient (see below). You focus on interpretation of statistical information presented in the Results section of the abstract.

### Abstract results

In the RCT, patients were randomised to the intervention group receiving medication X (n = 489) and a control group, receiving placebo (n = 485). During the trial period, the incidence of patients in the intervention group who had a primary outcome of COVID-19-related hospitalisations or death by day 28 was 0.82% (4 of 489 patients), compared with an incidence of 5.77% (28 of 485 patients) in the control group. The incidence of primary outcome was 4.96% less in the intervention than control group (95% CI: 2.72, 7.19%;  $P < 0.001$ ; relative risk reduction 85.8%).

A basic clinical interpretation of key statistical information from the abstract is outlined in Table 1. This table provides a concise interpretation of each concept, but is limited in that it does not adequately address the complexities and nuances of each statistical concept. For example,  $P$ -values and statistical significance are commonly misunderstood concepts,<sup>5</sup> and methods for their conceptualisation have been addressed previously in in-depth resources.<sup>4,6</sup>

### To the patient

You explain to the patient that the study looked at whether medication X

is effective for preventing COVID-19-related hospitalisation or death, reducing the risk of this outcome from 5.77% to 0.82%. The study did not address other outcomes, such as effectiveness in improving mild symptoms. There is evidence that medication X reduced the number of people with serious COVID-19, and that the improvements were statistically significant. Twenty-one people need to be treated for one person to benefit. The decision to use medication X for the patient will also depend on relevant Australian guidelines (eg COVID-19 Living Guidelines<sup>7</sup>) and other contextual factors, such as the availability of a Pharmaceutical Benefits Scheme subsidy.

## CASE STUDY 2

### How good is this screening test?

A woman aged 28 years comes to see you for routine antenatal care for a normal-risk pregnancy. She wants to learn more about screening test X for Trisomy 21. A study compared the diagnostic accuracy of screening test X for Trisomy 21 to a gold standard (amniocentesis). You need to explain the following results regarding screening test X for Trisomy 21 to the patient:

- sensitivity = 80%
- specificity = 95%
- positive predictive value (PPV) = 2%
- negative predictive value (NPV) > 99%.

In clinical practice, GPs rarely need to calculate these values from raw research results. However, we use the raw results of the research study on screening test X to explain each of these four test properties (Table 2).

An interpretation of key statistical information relevant to Case study 2 is outlined in Table 3. Sensitivity and specificity are properties of the diagnostic test itself, whereas PPV and NPV are heavily influenced by disease prevalence in the underlying study population.<sup>8</sup> For example, if the patient was from a

population with higher disease prevalence than the research population, the PPV of a positive result using screening test X would be higher than 2%.

#### To the patient

You explain to the patient that screening test X is better for ruling in Trisomy 21

(95% specificity) than ruling it out (80% sensitivity). If she has a positive test, there is a high likelihood of this being a false positive result (2% PPV), but if she has a negative screening test there is a high likelihood that this is a true result (>99% NPV). In practice, given the limited accuracy of screening test X, alternative

**Table 1. Key statistical information in Case study 1**

Key concept	Information in article	Explanation and interpretation
Background	RCT comparing intervention (medication X) with control (placebo)	-
Primary outcome	The composite primary outcome was COVID-19-related hospitalisations or death by day 28	Main outcome the study is comparing between the intervention and control groups A study usually has one primary outcome and several secondary outcomes
Raw results	Incidence: 0.82% in the intervention group versus 5.77% in the control group	Incidence is a measure of frequency. In this case it is the number of people developing the primary outcome (hospitalisation or death) during the study period (eg for the intervention group, the incidence is 4/489 = 0.82%) Interpretation: 0.82% of people in the intervention group developed hospitalisation or death over the study period
Absolute difference (also known as absolute risk reduction)	4.96% less in the intervention than control group	Absolute difference measures the effect size (ie the difference in incidence between the two groups): Absolute difference = 5.77% incidence in the control group - 0.82% incidence in the control group = 4.96% (rounded up) Interpretation: 4.96% fewer people in the intervention group experienced hospitalisation or death
NNT	Not provided in the article (n = 21)	The NNT is calculated by dividing 1 by the absolute difference (in decimals): NTT = 1/0.0496 = 21 (rounded up) Interpretation: 21 people need to be treated with the medication for 1 person to benefit
Relative risk	Relative risk reduction = 85.8% in the intervention group	Relative risk is a measure of the association between the treatment and outcome of interest. Absolute differences should be presented alongside relative differences to accurately interpret study results Relative risk = 0.82% incidence in intervention group/5.77% incidence in the control group = 0.142 Relative risk reduction = 1 - 0.142 (relative risk) = 0.858 Interpretation: people in the intervention group were 0.14-fold as likely (or 0.86-fold less likely) to experience hospitalisation or death
95% CI for absolute difference in incidence of 4.96%	2.72% to 7.19%	CIs estimate the precision of a result Interpretation: the likely range of the true difference in incidence is between 2.72% and 7.19%
P-value	P < 0.001 for absolute difference in incidence of 4.96% (statistical test used not specified)	The P-value is the probability that the difference (or a more extreme one) was found due to random chance, assuming that the intervention has zero effect P < 0.05 (5%) is the commonly used arbitrary cut-off for statistical significance Interpretation: the P-value met the prespecified cut-off, meaning that the RCT found a treatment effect in medication X versus placebo

CI, confidence interval; NNT, number needed to treat; RCT, randomised control trial.

screening tests with higher sensitivity and specificity, such as non-invasive prenatal testing (NIPT), should be considered, but may incur out-of-pocket costs.<sup>9</sup>

## Conclusion

Medical statistics are often taught with a strong research focus. This article provides a practical guide to statistics for busy, predominantly clinically focused GPs and GP registrars. Two case studies have been used to guide the interpretation of common statistical findings relating to clinical effectiveness and diagnostic accuracy. This guide complements existing in-depth resources, and readers are encouraged

to access statistical textbooks or journal articles for a thorough understanding of the concepts described in this article.<sup>4,8,10</sup>

### Author

Winnie Chen MBBS, MPH, FRACGP, Lecturer, Flinders NT, Darwin, NT

Competing interests: None.

Funding: None.

Provenance and peer review: Commissioned, externally peer reviewed.

### Correspondence to:

winnie.chen@menzies.edu.au

### References

1. Margolis SA. Evidence-based medicine. *Aust J Gen Pract* 2018;47(6):325. doi: 10.31128/AJGP-06-18-1234e.
2. Royal Australian College of General Practitioners (RACGP). Research in general practice: East

Melbourne: RACGP, 2022. Available at [www.racgp.org.au/education/education-providers/curriculum/curriculum-and-syllabus/units/research-in-general-practice](http://www.racgp.org.au/education/education-providers/curriculum/curriculum-and-syllabus/units/research-in-general-practice) [Accessed 1 November 2022].

3. Ramsay N, Sidhu H, Waters EK. Mathematics for medical practitioners and patients: Understanding how vaccines work for the community. *Aust J Gen Pract* 2020;49(5):293–94. doi: 10.31128/AJGP-08-19-5038.
4. Greenhalgh T. How to read a paper: The basics of evidence-based medicine and healthcare. 6th edn. Newark: Wiley, 2019.
5. Tam CWM, Khan AH, Knight A, Rhee J, Price K, McLean K. How doctors conceptualise P values: A mixed methods study. *Aust J Gen Pract* 2018;47(10):705–10. doi: 10.31128/AJGP-02-18-4502.
6. Frank O, Tam CM, Rhee J. Is it time to stop using statistical significance? *Aust Prescr* 2021;44(1):16–18. doi: 10.18773/austprescr.2020.074.
7. National COVID-19 Clinical Evidence Taskforce. Australian guidelines - caring for people with COVID-19. Melbourne, Vic: National Clinical Evidence Taskforce, 2023. Available at <https://clinicalevidence.net.au/covid-19/#living-guidelines> [Accessed 26 March 2023].
8. Trevethan R. Sensitivity, specificity, and predictive values: Foundations, plabilities, and pitfalls in research and practice. *Front Public Health* 2017;5:307. doi: 10.3389/fpubh.2017.00307.
9. Harraway J. Non-invasive prenatal testing. *Aust Fam Physician* 2017;46(10):735–39.
10. Campbell MJ. Statistics at square one. 12th edn. Hoboken: Wiley-Blackwell, 2021.

**Table 2. Accuracy results of screening test X compared with the gold standard for Trisomy 21**

	True diagnosis of Trisomy 21 using a gold standard		Total
	Disease +	Disease -	
Screening test X +	8	499	507
Screening test X -	2	9,491	9493
Total	10	9,990	10,000

correspondence [ajgp@racgp.org.au](mailto:ajgp@racgp.org.au)

**Table 3. Key statistical information in Case study 2**

Key concept	Value	Calculation	Description
True positives	8 people	Provided	No. people who are screen positive and disease positive
True negatives	9,491 people	Provided	No. people who are screen negative and disease negative
False positives	499 people	Provided	No. people who are screen positive and disease negative
False negatives	2 people	Provided	No. people who are screen negative and disease positive
Sensitivity	80%	No. true positives/everyone who is disease positive: $8/(8 + 2) \times 100 = 80\%$	How good the test is for detecting Trisomy 21 (ie ability to 'rule out')
Specificity	95%	No. true negatives/everyone who is disease negative: $9,491/(9,491 + 499) \times 100 = 95\%$	How good the test is for identifying people without Trisomy 21 (ie ability to 'rule in')
PPV	2%	No. true positives/everyone who is screen positive: $8/(8 + 499) \times 100 = 2\%$	If you have a positive screening test, there is 2% chance that it is true
NPV	>99%	No. true negatives/everyone who is screen negative: $9,491/(9,491 + 2) \times 100 > 99\%$	If you have a negative screening test, there is a >99% chance that it is true

NPV, negative predictive value; PPV, positive predictive value.