Chapter 12: Research Methods in Psychiatry

Rebecca Koncz

Specialty of Psychiatry, The University of Sydney, Australia, and Sydney Local Health District Mental Health Services, Australia.

and

Nerissa Soh

Specialty of Psychiatry, The University of Sydney, Australia.

Introduction

It is an unfortunate falsehood held by some doctors in training that education in basic research methods may serve little more than a pathway to complete your (potentially tedious) "evidence-based medicine" tasks. To the contrary, research is a fascinating and rewarding aspect of psychiatry that enables us to think outside of the box about complex and interesting diagnostic and management conundrums. If understood as intended, the principles of basic research methodology provide an important foundation for interpreting and delivering sound clinical care to your patients well into the twilight of your medical, psychiatric, or (dare we say) surgical careers.

Moreover, questionable research practices unfortunately run rife throughout the scientific community. Whilst controversial, it is estimated that up to one-third of scientists admit to questionable research practices, and this may well be an underestimate (Fanelli, 2009). Questionable research practices vary in nature, from examples such as selective reporting of results and overstating their implications (Gerrits et al., 2019), to less common although more serious scientific misconduct, which includes fabrication, falsification or plagiarism (Bauchner et al., 2018). Therefore, it is the responsibility of us as clinicians and researchers to uphold rigorous ethical and practical standards when interpreting and conducting research to ensure we do not fall prey to such malpractice.

This chapter will endeavour to provide a brief, but practical, overview of the principles of research methodology used in psychiatry. This includes an introduction to study types, common terminology, as well as basic statistical methods as a foundation to interpreting study results and applying them to your patients.

Study Types

When reading the abstract of a paper, begin by asking yourself "What type of study is this?"

Research studies may broadly be considered qualitative, quantitative, or a mixed method approach – that is, a combination of the two. The most common qualitative methods include participant observation (typically in their natural environment), in-depth or semi-structured interviews and focus groups. These types of studies offer insights into one's experience, perceptions and social concepts (Hammarberg et al., 2016) by describing, interpreting and developing theories from observations as they occur in natural, and not experimental, conditions. Although there is some debate as to the methodological standards to be expected of qualitative research, the current recommendations are outlined in the "Standards for Reporting Qualitative Research" (O'Brien et al., 2014).

Quantitative methods, on the other hand, typically enable a researcher to test a particular hypothesis using numerical or pre-defined categorical data. Quantitative methods describe, explain, predict, or control variables of interest. These studies may be observational – that is, without specific intervention from the researcher – or experimental (which may also be referred to as interventional). Key features to look for when reading a peer-reviewed study, whatever the study type, include:

1. Is there a clear description of the research question, aim, hypotheses or purpose of the study?

2. Is there a clear description of the participants of the study? Depending on the study type, this may be, for example, the profile of a patient (a case report), the demographic profile of a population (a cohort population) or the profile of a defined group (a randomised control trial; RCT, a comparative study, or a single arm study).

3. Is there a clear description of how the participants were recruited? Was ethics approval required and if yes, was it sought?

4. Is there a clear description of the study method? This may range from a very detailed trial protocol for an RCT, to a registered systematic review protocol presenting how a literature search was conducted. Validity and reliability of measurement methods (assays, psychometric tools) would be presented here. If there are statistical results presented in the study, is the statistical methodology used explained?

5. Are the results that are presented relevant to the methods and to the aims of the study?

6. Does the discussion interpret the results to answer the aims of the study? This is where nuance of the study's findings and its relevance to previous research are presented.

7. Does the study explain its limitations? A study's limitations often feed into potential risk of bias. These may include a small sample size, the limited profile or recruitment of a study sample rendering it unsuitable for generalization to the rest of the population, study methods that are not validated or not validated in the study population, low retention rates of study participants, or too few participants to provide sufficient statistical power to make a conclusion. In the case of systematic reviews, limitations may relate to the languages studies were published in, the inability to access unpublished data, or the low quality (low certainty) of the available evidence.

8. Did the authors state their conflicts of interest? This often relates to financial conflicts of interest, or that the study is sponsored by a commercial body.

A brief summary of the types of quantitative studies follows.

Observational Studies: Descriptive and Analytical

Descriptive studies, as the name suggests, systematically describe aspects of a variable of interest in a particular population. Data collection is commonly 'naturalistic' - that is, occurs in a natural environment without interference from the experiment - and is not typically hypothesis driven. Descriptive studies do not attempt to analyse the link (correlational, causal or otherwise) between an exposure and an outcome (Beaglehole et al., 1993), but may define the prevalence. *Prevalence* is the frequency of a particular characteristic, condition or disease process in a

defined population within a specific time period (Adams and McGuire, 2022). It is numerically defined as:

$$Prevalence = \frac{Number\ of\ cases\ in\ a\ population\ at\ one\ time\ point}{Total\ population\ at\ the\ same\ time\ point}$$

An example of a descriptive study in psychiatry could be the prevalence of alcohol use in adults with severe mental illness from 1950 to 2000.

The *case study* is a particular type of descriptive study that provides an in-depth, detailed description about a person, group or condition. This is particularly helpful when there is limited or no pre-existing literature on the given subject, such as when describing rare phenomena.

Analytical observational studies include *cross-sectional, case-control and cohort studies. Cross-sectional studies* investigate the association between a potential risk factor for an illness, and the outcome. In a cross-sectional study, all variables (exposure, outcome and confounding variables) are measured simultaneously. On the one hand, these studies are relatively time-efficient, easy to conduct and like descriptive studies, are useful for analysing the prevalence of a condition. However, it is not possible to make conclusions about the association between the risk factor and outcome.

Case-control studies require a group with the disease (the "cases") within a given population to be identified, and then compares this group to a suitable group without the disease (the "controls"). It is a type of longitudinal study in which the investigator retrospectively assesses the "exposure" (or risk factor) for the given disease between the cases and controls. Case-control studies therefore estimate the strength of an association between a disease condition and the risk factor(s) under question. The control group should be comparable based on other demographic (such as age, sex, source of recruitment) and clinical variables, particularly the presumed distribution of the exposure, to minimise the risk of selection bias (Lewis and Pelosi, 1990).

Case-control studies are commonly analysed using an "odds ratio" (OR), which is the measure of the disease in the exposed compared to the rate of the disease in the unexposed. Mathematically, this looks like:

	Cases	Controls	Total
Exposed	A	B	A+B

Unexposed	C	D	C+D
Total	A+C	B+D	

$$Odds\ ratio = \frac{A/B}{C/D} = \frac{AD}{BC}$$

One advantage of this study type is its use in uncommon illnesses, such as schizophrenia. An example of a large case-control study is to assess parental age and the risk of schizophrenia (Byrne et al., 2003).

Cohort studies begin by defining the population (cohort) without the disease. The cohort is selected based on exposure to a particular risk factor, and then followed up longitudinally to see who develops the disease or outcome. The exposure may be binary (i.e. exposed or not) or graded (e.g. minimal, moderate, extensive), the latter of which may provide evidence for a dose-response relationship (Prince et al., 2003). Given the temporal sequencing of events, and that participants are chosen based on their exposure (not an outcome, like a case-control study), the cohort study is preferential to provide information on causality and/or risk of developing a disease. Other advantages of this study type are its capacity to examine both rare exposures and multiple outcomes in one study. However, cohort studies may require an extensive follow-up duration to obtain data on a given outcome that occurs many years after an exposure, which therefore may be particularly costly. Cohort studies are useful to determine the *incidence* of a condition. *Incidence* is the number of *new* cases of a disease or condition in a defined population within a specific time period (Centers for Disease Control and Prevention, 2012).

Finally, it should be noted that guidelines have been developed to improve the quality of reporting observational trials. The STROBE – Strengthening the Reporting of Observational Studies in Epidemiology – statement is recognised internationally and provides a checklist to this effect (Vandenbroucke et al., 2007).

Interventional Studies – Randomised Controlled Trials

The *randomised controlled trial* (RCT) is widely considered to be the gold standard for evaluating treatment efficacy and safety. Subjects are randomly assigned to one of two groups: the 'experimental' group – that is, the group to receive the treatment under investigation – or the other 'control' (or comparator) group. The randomisation process should distribute any potentially confounding factors equally between groups, to minimise selection and/or observer bias, so that any difference in outcome can be attributed to the treatment itself (Kendall, 2003).

As with all studies, there should be a clearly defined hypothesis *a priori*, which will in turn define the intervention, population and primary outcome of the study. "PICO" is a commonly used acronym to assist in defining the study question. For example:

- Population, for example, adults with schizophrenia
- Intervention, for example, daily treatment with olanzapine
- Comparator, for example, compared to placebo
- Outcome, for example, reduce the positive symptoms of psychosis?

Effective randomisation includes the process of allocation concealment, and the subject and investigator should remain "blind" to group allocation for the duration of the study. If both the investigator and participant do not know who is allocated to each group, it is referred to as "double-blinded". Apart from the different intervention, both groups should be treated identically, so that any difference in outcome can be confidently attributed to the intervention itself (Akobeng, 2005). Quality control in reporting of RCT protocols and results is of utmost importance. 'CONSORT' - Consolidated Standards of Reporting Trials' (Schulz et al., 2010) remains the best practice guideline for reporting RCTs.

Systematic Reviews

A systematic review is a review that uses clear and systematic methods to identify all empirical evidence (primary studies) in order to answer a specific research question. The method aims to reduce bias and thus result in more reliable conclusions for informing decision-making. Systematic reviews commence with a clear and detailed protocol that presents an *a priori* methodological approach (Moher et al., 2015). Protocols for systematic reviews may be registered at the international Prospective Register Of Systematic Reviews PROSPERO.

Systematic review protocols include clear and explicit research questions (see PICO criteria) that the final review aims to answer, proposed search strategies including search terms, the databases and data sources that will be searched, inclusion and exclusion criteria for the empirical studies that are identified, how the risk of bias of empirical studies will be assessed and how the included data will be synthesised. The Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) (Moher et al., 2015) is a checklist which many peer-reviewed journals require a systematic review to meet and include to be accepted for publication.

A published systematic review will clearly articulate the research question it aims to answer, present the search strategy and data sources that were searched, the inclusion/exclusion criteria for empirical studies and the assessment tools used to assess the empirical studies' risk of bias. The inclusion/exclusion criteria will often be detailed for each of the PICO criteria, as well as study type (e.g., RCTs, cohort studies, case series), publication date limits and language limits. The dates that the database searches were conducted should also be stated. The total number of unique studies and publications retrieved is stated along with the number excluded and the reasons for exclusion reflecting the inclusion/exclusion criteria. The number of empirical studies retained is stated and often the number that are included in quantitative analyses (that is, a meta-analysis) and qualitative analyses.

As systematic reviews formally assess the risk of bias of their included empirical studies and as well as their own methodology, they allow for judgements to be made as to how reliable or robust their conclusions are. Risks of bias include publication bias or non-reporting bias, such as non-published studies not being canvassed. Other risks of bias assessed by tools (e.g., Cochrane's Risk of Bias tool for RCTs or for Non-Randomised Studies of Intervention (ROBINS-I)) include whether or not participants and assessors were blinded to the intervention, selective reporting of outcomes and in the case of RCTs, whether or not the participants were truly randomized in the selection process. Authors' conflicts of interest should also be presented and are usually incorporated into studies' risk of bias assessments.

Meta-Analysis

A meta-analysis is when the results of individual empirical studies are combined to give a summary statistic. The basis of a meta-analysis is a systematic review, which identifies the empirical evidence that will be combined to give the overall statistic. Not every empirical study included in a systematic review will be included in a meta-analysis, as meta-analyses require the set of empirical studies to use similar methods for assessing outcomes.

Generally, systematic reviews and meta-analyses do not require ethics approval, as they assess primary studies that have already received approval and are not directly collecting data from participants or using participants' tissues.

Hierarchy of Evidence

Study design influences the quality of a study's evidence (Schünemann, 2013). For studies of interventions, the strongest level of evidence (and thus with the least bias) is a systematic review of RCTs, followed by RCTs, then non-randomised comparative studies (cohort studies, case controls studies) and then case

series or case reports (Coleman et al., 2009). RCTs generally provide stronger evidence than observational studies, which in turn provide stronger evidence than uncontrolled case series (Schünemann, 2013).

However, it should be noted that a well-conducted cohort study may yield better quality evidence than a poorly conducted RCT. Further, some topics do not lend themselves to RCT methodology, including for ethical reasons. The quality of the evidence is reduced if there is bias (see Systematic Reviews). With regard to diagnostic accuracy, high quality evidence is derived from cross-sectional or cohort studies in patients with diagnostic uncertainty and studies that directly compare test results with an appropriate reference standard (Schünemann, 2013).

Measurement Concepts

Reliability and Validity

Reliability and validity are concepts used to evaluate the quality of measurement (Adams and McGuire, 2022). *Reliability* refers to the consistency of measurement across time and between different observers. Reliability can be assessed in several ways:

1. Internal consistency: how consistently each item on a scale measures the same construct.
2. Inter-rater reliability: the extent of agreement between scores provided by different raters.
3. Test-retest reliability: how consistently the scores are measured over time.

Validity refers to the accuracy of findings or measures, that is, how well a tool measures what it is supposed to measure. Validity can be assessed using different dimensions. Some examples of these include:

- Face validity: does the tool appear to measure what it purports to?
- Construct validity: does the tool reflect the characteristics of the hypothetical construct we want to measure? An example may be a scale to measure depression severity, based on relevant existing knowledge and criteria for the concept of depression.
- Content validity: does the tool include all aspects that are relevant to the construct?
- Criterion-related validity: how well does the test sub-score correlate with or perform against an independent measure or standard (concurrent validity) or future standard (predictive validity)?

Diagnostic Studies

A diagnostic test is a procedure used to determine if a patient or study participant is likely to have a particular condition. In psychiatry, diagnostic tests are relatively rare as we are yet to identify specific, quantifiable biomarkers to determine the presence of a primary psychiatric illness. However, in Alzheimer's disease for example, advances in positron emission tomography (PET) imaging technology have enabled us to visualise core pathological features *in vivo*, such as β-amyloid and tau. Such diagnostic tests are usually compared to a 'gold standard', which for Alzheimer's disease has historically only been confirmed at post-mortem (Peacock and Peacock, 2011). For a diagnostic test, four measures are commonly used to describe the adequacy and utility of the test (Trevethan, 2017):

1. Sensitivity: The ability of a test to detect all people who have a condition (i.e. the 'true positives'). Mathematically, this looks like:

$$Sensitivity = \frac{True\ positives}{True\ positives + False\ negatives}$$

2. Specificity: The ability of a test to detect all people who do not have a condition (i.e. the 'true negatives'). This is equivalent to:

$$Specificity = \frac{True\ negatives}{True\ negatives + False\ positives}$$

3. Positive predictive value (PPV): The probability that an individual with a positive test result will truly have the condition. Mathematically, this looks like:

$$Positive\ predictive\ value = \frac{True\ positives}{True\ positives + False\ positives}$$

4. Negative predictive value (NPV): The likelihood that a person with a negative test result will truly not have the condition. Mathematically, this looks like:

$$Negative\ predictive\ value = \frac{True\ negatives}{True\ negatives + False\ negatives}$$

Risky Business

An important concept in psychiatric research is "risk". This may apply to the probability of acquiring a psychiatric disorder following exposure to a particular risk factor (for example, the risk of developing schizophrenia after exposure to regular cannabis use), or the risk of a particular event following exposure to treatment (for example, the risk of elderly adults developing a stroke following the use of antipsychotic medication). Statistically speaking, the two commonly cited types of risk include:

1. Absolute risk: This is the probability of an event occurring over a period of time (Boyce and Hadzi-Pavlovic, 2011). An example may be the risk of developing late life Alzheimer's disease following treatment for mid-life hypertension.
2. Relative risk: This is a proportional measure of risk between two groups which differ by some factor, such as a demographic or clinical factors, including exposure to a particular risk or treatment. For example, if the lifetime risk of developing autism spectrum disorder is 43 per 1000 males and 11 per 1000 females (Maenner, 2021), the relative risk is $0.043/0.011 = 3.9$ in males compared to females.

Number Needed to Treat and Number Needed to Harm

Number needed to treat (NNT) is defined as the expected number of people who need to receive an intervention instead of a comparator therapy in order for one additional person to achieve or avoid an event or outcome within a specified timeframe. For example, a NNT of 10 indicates for every 10 participants given the intervention treatment, one additional person will achieve or avoid an event compared to when the comparator treatment is given. Mathematically, NNT is calculated from risk differences (RDs) and is represented by the formula below (Schünemann et al., 2022).

$$NNT = \frac{1}{absolute\ value\ of\ risk\ difference} = \frac{1}{|RD|}$$

NNTs can also be calculated from risk ratios (RR) and from odds ratios and assumed control risks (ORs and ACRs). NNTs are expressed as positive whole numbers and if the calculated values have decimals, they are always rounded up.

You may also come across the term "number needed to harm" (NNH), which implies that the intervention treatment results in an unfavourable event. It is not recommended that this term be used and instead, be replaced with the terms "number needed to treat for an additional beneficial outcome" or "number needed to treat for an additional harmful outcome" (Schünemann et al., 2022).

Other Concepts in Basic Statistics

Effect size refers to the magnitude of the effect of a variable. This is commonly reported using the statistic, Cohen's d, whereby it is widely accepted that a value of 0.2 indicates a small effect size, 0.5 indicates a medium effect size, and 0.8 and above represents a strong effect size (Adams and McGuire, 2022).

A *confidence interval* describes the degree of certainty with which the true value of the statistical test lies within a given range. A 95% confidence interval indicates that we are 95% certain that the true value lies within the range. Smaller confidence intervals therefore indicate a more precise estimate of the mean value (Riffenburgh, 2012).

Statistical significance is the likelihood that the relationship between two variables is not due to chance. This is typically determined by hypothesis testing. The null hypothesis states that any observed differences between groups is entirely due to chance (Beaglehole et al., 1993). The *p-value* is then used to calculate the likelihood that the null hypothesis is either accepted or rejected; the larger the p-value, the greater the probability that the null hypothesis is accepted – that is, observed differences between groups are likely due to chance and not true between-group differences. It is customary practice to accept 0.05 or 0.01 as significance levels.

Although widely used, much confusion and debate shrouds the misinterpretation and over-reliance on p-values. The p-value itself is a 'probability statement about the observed sample in the context of a hypothesis' and should be augmented with additional statistical information (such as effect size) to avoid misinterpretation (Altman and Krzywinski, 2017). There has even been a push from within the scientific community for the 'entire concept of statistical significance to be abandoned' - not an eradication of p-values of confidence intervals, but no longer treated categorically as either proving or disproving the null hypothesis (Amrhein et

al., 2019). As Amrhein et al. (2019) argue, this oversimplification can lead to brash assumptions, and incorrect interpretations, including that non-significance equates to no effect. Caution is warranted when interpreting a p-value, large or small, and weight given to other factors such as study design and data quality.

Finally, it should be noted that statistical significance does not necessarily equate to clinical significance. For example, a statistically significant 2-point increase on a scale measuring depressive symptoms may not be clinically relevant. One must always consider the clinical application, which includes the generalizability and limitations of research to relevant patient populations when interpreting the results of any study.

Ethics

Research ethics may be defined as 'the ethics of the planning, conduct, and reporting of research' and should at a minimum protect interests of the public, the research participants (both human and animal) and the researchers conducting the studies (Resources for Research Ethics Education, 1999-2016). The World Medical Association's Declaration of Helsinki (The World Medical Association, 2013) is 'a statement of ethical principles for medical research involving human subjects, including research on identifiable human material and data'. It comprises general principles that places the patient's (participant's) health and wellbeing at the forefront when conducting research, as well as principles related to risks, burdens and benefits of the research, groups and individuals that are vulnerable, the privacy and confidentiality of research participants, informed consent, post-trial provisions for participants and communication of results. Particularly pertinent to research in psychiatry are the principles relating to research where participants are mentally incapable of giving informed consent, and the principle that research in a vulnerable group is only justified if it serves the needs or priorities of the vulnerable group and benefits them, and thus the research cannot be conducted in a non-vulnerable group.

In order to be published in peer reviewed journals, researchers usually must declare that the study was conducted in line with ethics approval, or that ethics approval was not required, for example in a literature review or a quality assurance project. For example, the Australian and New Zealand Journal of Psychiatry (ANZJP) requires research conducted in human participants to adhere to the Declaration of Helsinki. Further, ANZJP requires that studies with human (or animal) participants must state in the methods section that approval was provided by the relevant Ethics Committee or Institutional Review Board, or that the approval was waived.

The values and principles of ethical research in humans comprise (The Australian Research Council and Universities Australia, Updated 2018):

1. Respect for human beings and valuing their autonomy, including protecting and empowering those with reduced or no autonomy.
2. That the research has merit and integrity, to ethically justify being conducted with human participants.
3. That the research is just, where the benefits and burdens of research are fairly distributed and there is 'fair treatment' in participant recruitment.
4. Beneficence is practised by researchers, by considering and assessing risks of harm and potential benefits of the research to participants and the wider community.

Research with merit is characterized by its capacity to enhance knowledge, social welfare, and individual wellbeing, while also advancing researchers' proficiency. Integrity in research is upheld by scholars dedicated to the pursuit of knowledge, adhering to ethical principles, and conducting research with honesty. When research maintains integrity, both favourable and unfavourable outcomes are transparently communicated, facilitating scrutiny, and enabling the findings to enrich public knowledge and comprehension. Human research ethics protects both the researcher and the participant, ensuring that the research benefits outweigh any negative effects of the study, that the research does not waste resources. A study should not deviate from the ethically approved protocol unless approval is sought and granted by the ethics committee for the amendments. In addition, ethics approval usually requires researchers to submit regular progress and safety reports.

Concluding Remarks

Whilst this chapter is not an exhaustive deep dive into research methodology, it provided an overview of common study types and statistical concepts. Further reading is recommended. Nevertheless, this has hopefully imparted a foundation to understand and define basic principles in psychiatric research to enable critical evaluation of the literature, to ultimately improve the standard of clinical care provided to patients.

Further Reading

Hammarberg, K., Kirkman, M., & De Lacey, S. (2016). Qualitative research methods: when to use them and how to judge them. *Human Reproduction*, *31*(3), 498-501.

National Health and Medical Research Council (2009). NMHRC levels of evidence and grades for recommendations for developers of guidelines. Available at: **https://edit.nhmrc.gov.au/sites/default/files/images/NHMRC%20Levels%20and%20Grades%20(2009).pdf**

Cite as:

Koncz, R., Soh, N. (2024). Research methods in psychiatry. In Boyce, P., Harris, A., and Malhi, G.S. (Eds.), *The Sydney textbook of psychiatry* (pp. 153–166). The University of Sydney.

References

Adams KA and McGuire EK (2022) *Research methods, statistics, and applications.* Sage Publications.

Akobeng AK (2005) Understanding randomised controlled trials. *Archives of disease in childhood* 90(8): 840-844.

Altman N and Krzywinski M (2017) Points of significance: interpreting P values. *Nature methods* 14(3): 213-215.

Amrhein V, Greenland S and McShane B (2019) Scientists rise up against statistical significance. *Nature* 567(7748): 305-307.

Bauchner H, Fontanarosa PB, Flanagin A, et al. (2018) Scientific misconduct and medical journals. *Jama* 320(19): 1985-1987.

Beaglehole R, Bonita R and Kjellström T (1993) *Basic epidemiology.* World Health Organization Geneva.

Boyce P and Hadzi-Pavlovic D (2011) Risks risky and not so risky. *Acta Neuropsychiatrica* 23(1): 43-45.

Byrne M, Agerbo E, Ewald H, et al. (2003) Parental age and risk of schizophrenia: a case-control study. *Archives of general psychiatry* 60(7): 673-678.

Centers for Disease Control and Prevention (2012) Principles of Epidemiology in Public Health Practice. Atlanta, GA: United States Department of Health and Human Services.

Coleman K, Norris S and Weston A (2009) National Health and Medical Research Council. *NHMRC Additional Levels of Evidence and Grades for Recommendations for Developers of Guidelines: Stage 2 Consultation; Early 2008–End Jun 2009*.

Fanelli D (2009) How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PloS one* 4(5): e5738.

Gerrits RG, Jansen T, Mulyanto J, et al. (2019) Occurrence and nature of questionable research practices in the reporting of messages and conclusions in international scientific Health Services Research publications: a structured assessment of publications authored by researchers in the Netherlands. *BMJ open* 9(5): e027903.

Hammarberg K, Kirkman M and De Lacey S (2016) Qualitative research methods: when to use them and how to judge them. *Human reproduction* 31(3): 498-501.

Kendall J (2003) Designing a research project: randomised controlled trials and their principles. *Emergency medicine journal* 20(2): 164-168.

Lewis G and Pelosi AJ (1990) The case-control study in psychiatry. *British journal of psychiatry* 157(2): 197-207.

Maenner MJ (2021) Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2018. *MMWR. Surveillance Summaries* 70.

Moher D, Shamseer L, Clarke M, et al. (2015) Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews* 4: 1-9.

O'Brien BC, Harris IB, Beckman TJ, et al. (2014) Standards for reporting qualitative research: a synthesis of recommendations. *Academic medicine* 89(9): 1245-1251.

Peacock J and Peacock P (2011) *Oxford handbook of medical statistics.* Oxford university press.

Prince M, Stewart R, Ford T, et al. (2003) *Practical psychiatric epidemiology.* Oxford University Press Oxford.

Resources for Research Ethics Education (1999-2016) *What is Research Ethics?* Available at: **http://research-ethics.org/introduction/what/**.

Riffenburgh RH (2012) *Statistics in medicine.* Academic press.

Schulz KF, Altman DG and Moher D (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Journal of Pharmacology and pharmacotherapeutics* 1(2): 100-107.

Schünemann H (2013) Bro ek J, Guyatt G, Oxman A, editors. GRADE Working Group. GRADE handbook for grading quality of evidence and strength of recommendations.

Schünemann H, Vist G, Higgins J, et al. (2022) Chapter 15: Interpreting results and drawing conclusions. In: Higgins J, Thomas J, Chandler J, et al. (eds) *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane.

The Australian Research Council and Universities Australia (Updated 2018) National Statement on Ethical Conduct in Human Research 2007. Commonwealth of Australia, Canberra: The National Health and Medical Research Council.

The World Medical Association (2013) World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. World Medical Association.

Trevethan R (2017) Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. *Frontiers in public health* 5: 308890.

Vandenbroucke JP, Elm Ev, Altman DG, et al. (2007) Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Annals of internal medicine* 147(8): W-163-W-194.