# Chapter 11: Psychiatric Epidemiology

Brian O'Toole

*Brain and Mind Centre, The University of Sydney, Australia, and Specialty of Psychiatry, The University of Sydney, Australia.*

## Introduction

According to the Dictionary of Epidemiology (Porta, 2014) the science of epidemiology is the study of the distribution and determinants of disease in populations. The word itself comes from the Greek words "epi", which means *upon* and "demos", which means *the people*. The term was first used by Homer but its medical meaning was conferred by Hippocrates (Martin and Martin-Granel, 2006). It invoked the old superstitious notion of the things that vengeful gods send to afflict the people, like fires, flood, famine, frogs, plagues, and pestilence. In the context of mental health, psychiatric epidemiology refers to psychiatric illnesses in humans and in a treatment context the term "clinical epidemiology" has been coined to refer to the study of the distribution and determinants of diseases encountered in the clinic, but the essential scientific underpinning of this variant differs only from population epidemiology in its focus; the scientific methodology remains the same although the techniques may vary with each research question. And the principles apply not only to noxious exposures: for example, risk and protective factors, such as having a positive and warm emotional environment while growing up, can ameliorate a good deal of later emergent psychiatric disorders.

Humans are possibly the most complex and least understood of scientific fields of enquiry, particularly our psyche. As we know now, the complexity of this field encompasses genes, upbringing, emotions, cognitions, personalities, and behaviour, and is an exciting and developing field in Australia. Epidemiology is devoted to uncovering the causal relations between *exposure* to something and some consequent *outcome*, for example cannabis consumption and psychotic disorder, or interpersonal domestic violence and depression. If a causal pathway from a particular exposure to a particular disorder can be shown, then this opens up the possibility of

interventions that may prevent the disorder from occurring. Epidemiology has evolved from roots in sociology, psychology, statistics and medicine to develop reliable methods to pursue an understanding of the human condition. Any practical scientific study of these things has to rely on data that come from smaller samples that, hopefully, represent their parent populations. And this is where the trouble starts. There are many, many threats to the validity of clinical epidemiological studies, and each has a bearing on the process of statistical inference; that is, the process of using statistics to make inferences about human or biological characteristics that apply beyond the sample of data results to hand. Statistical inference is the bedrock of epidemiological research.

The statistic that tells us whether we have a significant association between the exposure of interest and the outcome of interest rests on a set of assumptions that must stringently apply. These rest on only two fundamental features: the assumptions that are necessary about the population members (and hence the sample) and the assumptions that are necessary about the parameter (statistic) of interest. These twin features lead to the source of the three possible threats to causal inference. Before outlining these, it is important to understand how epidemiological studies are designed, because the analysis – and therefore the conclusion – depends on how the study is constructed.

## Study Designs

The main goal of medical research is to explore the possible causal associations between an Exposure (**E**) of some kind of agent and the subsequent development of a Disease (**D**) of some kind. In the simplest case, we can think of two basic groups of people, in one of whom some people are **E**xposed to a causal agent and some people are not **E**xposed ($\bar{\mathbf{E}}$), and some who develop **D**isease (**D+ve**) and some who do not (**D-ve**).

A simple example would be a study where a group of people who have been diagnosed with major depressive disorder and a group of people who have not been so diagnosed are compared to see whether they were admitted to the emergency department (ED) with suspected suicidal attempt. In this simple scenario we can construct a 2x2 table that shows this relationship:

| | **D +ve** | **D -ve** | *Total* |
|---|---|---|---|
| **E** | *a* | *b* | $a + b$ |
| $\bar{\mathbf{E}}$ | *c* | *d* | $c + d$ |
| *Total* | $a + c$ | $b + d$ | $N = a+b+c+d$ |

**E** = Depressed, $\bar{\mathbf{E}}$ = Not depressed, **D +ve** = Admitted to ED, **D –ve** = Not admitted to ED. (Those who remember their Algebra from high school will recognise that *a*, *b*, *c*, and *d* are the numbers of people in each cell of the table, and N is the total number of people in the study.)

There are not too many ways to design an epidemiological study. Life can be simple. The most obvious way to begin is to select a group of people who are already exposed to something and a group of people who are not exposed. Broadly speaking, this strategy of *selection based on exposure* gives a *Cohort Study* design. Under these circumstances, selection sweeps across the rows of the table above, where a total of ($a + b$) people have been diagnosed with major depressive disorder and a total of ($c + d$**)** people are not so diagnosed; ($a + c$) are admitted to ED and ($b + d$) have not been admitted. The numbers ($a + b$) and ($c + d$) are fixed by design, while the numbers ($a + c$) and ($b + d$) depend on the outcome of the study assessments.

Of course, it is sometimes feasible to conduct an *experiment*, where an experimenter actually *causes* the exposure. This is common in lab-based and therapeutic or drug evaluation studies. When the exposure is *randomly allocated* to the exposed and unexposed groups, this broadly defines a *Randomised Controlled Trial* (RCT). Randomisation is used to distribute any important errors randomly across the two groups in order to control for potential confounding. It also gives control of exposure to the experimenter. Of course, it is not always possible to conduct an RCT, perhaps because of legal or moral or practical considerations, in which case we have to resort to other types of studies to provide the best available evidence.

Difficulties that adhere to cohort studies are often because it is hard to assemble a comparable cohort, or perhaps because the disorder is rare and thus less likely to be detected in a cohort study (unless the study was huge). Another consideration is the location of the researcher – a clinical epidemiologist needs access to a clinic and clinicians (and clinicians need access to an epidemiologist and statistician!). It is understandable that much of medical research is carried out on patients in tertiary referral teaching hospitals, often with captive subjects/patients. This is often the case with diseases such as the variety of cancers that are individually rare (such as cervical or prostate or pancreatic) or people with schizophrenia because it would be tiring and fruitless to assemble a community-based cohort to try and detect them efficiently. In this instance, study designs can begin by selecting people (cases) who already have the disorder and then select a set of *control* people who do not. This strategy of *selection based on disease outcome* broadly defines a *Case-Control Study*. Under these circumstances selection begins from the top of the table and sweeps downwards, where a total of ($a + c$) cases has been selected and a total of ($b + d$) controls has been selected to compare with them. (Importantly, in this case, the marginal totals ($a + b$) and ($c + d$) have no meaning.)

Another strategy for conducting epidemiological studies that does not rely on selection by exposure or selection by disease, is simply to select subjects based on some other characteristic. An example of this strategy design would be a study that took all consecutive patients admitted to Westmead Hospital and then determine how many of them had the exposure of interest (e.g., how many were victims of interpersonal violence) and how many developed the outcome of interest (e.g., major depressive illness). This general strategy broadly defines a *Cross-Sectional Study*, and is the method commonly used for community and social surveys. Cross sectional studies are very useful in determining the extent of a particular problem, for example the Australian Bureau of Statistics' National Surveys of Mental Health and Wellbeing, or the WHO World Mental Health Surveys that describe the prevalence in the community of a range of non-psychotic mental health diagnoses.

Finally, some studies in clinical epidemiology restrict themselves to studies of cases only; in this situation, *selection is based on disease outcome* but no control group enters the picture. This design is only powerful if the exposure and the outcome are so clearly related that anybody could see them (for example, prescribing the drug thalidomide for morning sickness and seeing the incidence of birth defects).

This simple explication of available study designs reveals only a few possible designs, but the world is more complicated than that since there are many other more subtle types of designs that expand this rubric (Catts et al., 2010). Moreover, hybrid designs can be used, for example a longitudinal study of cases with a comparison control group drawn from the community if, when followed up over time, would constitute a *case-cohort* study, incorporating features of both case-control and cohort studies.

In the accepted wisdom that prevails in Epi-Land, study designs can be ranked based on the quality of the scientific evidence; for example, the Cochrane Collaboration limits its attention to reviews of RCTs as the only reliable solid evidence acceptable to the Evidence-Based Medicine devotees. This is, of course, controversial. Each of the possible study designs are useful for different research questions. A case series is useful to define a particular disease and may involve qualitative analysis to discern the essential features of a particular disorder and discriminate it from other variants or subtypes. These studies are the means of differentiating diagnoses, where, for example, schizophrenia can be distinguishable from schizophreniform disorder, anxiety can be distinguished from depression in spite of symptom overlap, and Bipolar I can be distinguished from Bipolar II. Once the disorder has been defined, cross-sectional studies are useful to estimate the size of the problem. For example, the regular National Surveys of Mental Health and Wellbeing (NSMHWB) that are conducted by the Australian Bureau of Statistics are

there precisely to determine the prevalence of psychiatric disorders in the Australian population. But these do not provide a basis for analysis of causes or risk factors, except in a limited descriptive sense. Case-control studies are very useful for assessing the risk factors for disorders, particularly disorders that have low prevalence in population studies like the NSMHWB. A prime example of this are the psychotic disorders, since a very large population sample would be needed to provide a large enough sample to give accurate estimates for such low prevalence disorders (O'Toole, 2000).

Cohort studies permit the examination of outcomes over time, for example the effectiveness of treatment of first episode psychosis, where people who present to hospital are followed to determine the proportion whose condition improves over time. Some cohort studies can continue for many years, such as the famous Dunedin Study (Koenen et al., 2007) that enrolled all children born in a one-year period 1972-73 in the New Zealand city of Dunedin and followed them for more than 36 years – with an enviable 96% response rate!

Finally, experimental studies such as randomised controlled trials are useful for evaluating treatments that are randomly allocated to individuals to compare treated and untreated individuals. While it is often cited that RCTs provide the best evidence – for treatments – they cannot be used where, for example, the exposure itself cannot be administered at random, or indeed at all, in situations where the exposure is toxic (think child sexual abuse, combat trauma). A well-conducted cohort study has the potential to provide excellent evidence, as does a well-conducted case-control study. But the physical abnormalities associated with Thalidomide during pregnancy were discovered from a case series. And Semmelweiss discovered the cause of puerperal fever in childbirth by washing his hands a few times and not washing his hands a few times and noting the death rate! His colleagues treated him like a germ despite the fact that he had just put his finger on a prime mechanism of disease transmission. But remember, there has never been an RCT reported of the effectiveness of parachutes (Smith and Pell, 2003).

## Incidence and Prevalence

All illness starts with an event, like the lighting of a fuse to a bomb. The event might be the experience of interpersonal violence, or it could be just being conceived (in cases such as genetic disorders). At some point in time, the individual is disease-free, but after that time they are not (i.e., they have been struck down with or acquired the illness or condition). When cases appear for the first time, they are called *incident* cases. For example, the number of new cases of severe depression presenting at Emergency Department in a specified time period would constitute a class of incident cases of depression. The concept behind incidence is that incident

cases are *new* cases of disease. In epidemiologic studies it is often preferable to study incident cases because there is no other factor such as prior treatment or diagnoses to complicate matters.

*Prevalence*, on the other hand, refers to the number of existing cases of a particular disease at a particular point in time. For example, the number of patients who are known to have anxiety disorder in New South Wales in June 2009 would constitute a prevalence. A special example is *lifetime prevalences* (also described by epidemiologists as cumulative incidence), which is the number of people who experience the disease over their lifetime up to the present. Psychiatric epidemiological assessment routinely records the onset and recency of symptoms to give prevalence over a lifetime (or cumulative incidence), one-year prevalence, six-and- one month prevalence, and so on. The important feature of the term prevalence is that it designates a state of nature at a particular point in time and place. In contrast, incidence is a change in the state of nature (Miettinen, 2007).

## Errors in Studies

There are only three fundamental errors that can occur when using samples to draw inferences about larger populations: one can either study a biased set of subjects, or use measures that are biased or inaccurate, or something else creeps in to bias the observed relationships. These errors are termed selection errors, measurement errors, and confounding errors. Although there are only three fundamental types of errors there are many, many ways to make them.

The first error that can vitiate the results of an otherwise well-conducted study, is the issue of the unit on which the data are based. Assembling a set of subjects that do not allow a fair test of the hypothesis or theory that a researcher is trying to grapple with will lead to *selection errors*. Using only people who come to Emergency Department with first episode psychosis will not guarantee a random sample of people with psychosis. A bias is automatically introduced if there are discrepancies between the population of people that you want to make inferences about and the sample of people that you actually collect data from. Every study needs to ask fundamental questions: *Whom do I want to talk about?* That is, in statistical parlance, *what is the unit of observation?* And, as well, there is the critical supplementary question: *How will I gather a suitable sample of them to study?*

## The Process of Subject Selection

Conceptually, selection begins with specification of the people you want to draw inferences about; they are conceived in broad terms like “doctors”, “women”, “people with major depressive illness”. At this level, they are termed the *reference*

*population*. Variation away from the reference population results in a threat to the external validity of the study, where the results of the study may not be applicable (generalisable) to the parent population that the researchers intended it to be. For example, a survey of AMA members cannot be representative of the total population of doctors simply because not all doctors belong to the AMA, so that findings from a survey of AMA members might not be applicable to – i.e. generalisable to – all doctors.

Next, the study must then proceed to determine who is available and eligible for selection. This requires the construction of the *sample frame*. It sets the eligibility criteria for inclusion and exclusion. Examples of sample frames could be things like a register of GPs in the Nepean Division of General Practice, or patients at Orygen in Melbourne, or patients in specific psychiatric wards. Sample frames are tools that allow study subjects or units of observation to be accessed and selected for study and are a practical way of determining who is eligible for study; it is critical that they be representative of the reference population, otherwise bias creeps in.

Once the sample frame has been identified and constructed there has to be a mechanism for selection. Strictly speaking, to be able to use inferential statistics the sample must be selected as a *probability sample*, where the probability of selection is known. Just taking any higgledy-piggledy convenience sample that was assembled and close by (e.g., captive patients in a ward) and expecting to apply the results to a general population is sadly misguided. It may be appropriate if you only want to confine your conclusions to the ward but not if you want to go beyond and generalise to other people. The determination of who is actually selected and with what probability is a critical element in the chain of decisions from study conception to publication of scientifically sound findings. Probability sampling requires that each potential individual for study has a known and non-zero chance of selection. Non-probability samples include haphazard methods like volunteering, purposive sampling, internet surveys, and general convenience samples. All of these other methods have no probability basis and therefore confer no validity in statistical analysis. Beware.

Once the sample frame and the sampling scheme have been determined, the last step is to ensure study subjects respond. In human studies, subjects can be lost to follow-up by, for example, dropping out of studies or failing to comply with medication regimens after selection and data collection. Examples abound where, for instance, people who have depression in the first wave of a longitudinal study fail to respond in a second wave, or even fail to respond in a first wave. In all types of study, it should be necessary to follow and document the path from the identification and selection of study subjects through to the actual gathering of complete data from them. Any deviation in representativeness of subjects from the reference population

along this path will inevitably introduce errors and consequent bias when inferences are drawn from the data analysis. The Australian Bureau of Statistics and the World Health Organisation do not use non-probability methods and consequently their data is much more accurate. In the face of non-response from the selected subjects, sometimes it is possible to quantify and adjust for the differences between the selected and responding subjects (Lynch et al., 1993; O'Toole et al., 1996) in cases where information is available on the whole sample.

## Measurement Errors

It is axiomatic that every measure taken on an individual carries with it an element of uncertainty because classification is never perfect. Thus, the second error that can creep into epidemiological studies is that of *measurement error*. The American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM), now in its fifth edition (published in 2013 with a text revision in 2022), was developed to standardise psychiatric diagnostic criteria across the United States (and the world), alongside the World Health Organisation's International Classification of Disease (ICD), now in its 11th edition. Unfortunately, DSM and ICD do not always agree on their definitions; while DSM-III introduced the diagnosis of posttraumatic stress disorder (PTSD) in 1980 it took ICD a further 12 years to add it to their lexicon, and the diagnostic criteria for PTSD have changed five times (Spitzer et al., 2007).

Since the 1980s clinician-scientists have developed standardised interviews for assessing population mental health; such interviews as the Diagnostic Interview Schedule (DIS) (Robins, 1981), the Structured Clinical Interview for DSM (SCID) (Spitzer et al 1987) now in its 5th edition, and the World Health Organisation's Composite International Diagnostic Interview (CIDI) (WHO 1997) have been developed to provide standardised diagnoses in epidemiological studies; the CIDI in particular has been developed as a computerised interview that can be administered by lay interviewers and indeed by computer assisted self-completion without a clinician or interviewer present. The CIDI algorithms yield diagnoses that are mapped both onto DSM and ICD systems and are the bedrock of the WHO World Mental Health Surveys. Further developments have led to a plethora of standardised instruments that focus on specific disorders, such as alcohol use disorders, sleep disorders (Taylor et al., 2018), PTSD (Weathers et al., 2018), and many others. If psychiatric studies use these standardised methods, then this directly permits comparison with other studies. In Australia, using the same CIDI that the ABS used in the National Survey of Mental Health and Wellbeing has enabled studies comparing the relative prevalence of disorders in peoples as diverse as Australian Vietnam veterans (O'Toole et al., 2009) and Aboriginal and Torres Strait Islander Australians (Nasir et al., 2018).

Errors in the diagnostic rubric and changes in diagnostic measurement can be disastrous in clinical studies. Accurate diagnosis is essential for studies that seek to find risk factors or markers for disease, for example, genomic studies of schizophrenia must rely on accurate diagnosis. Unfortunately, medical diagnosis is rarely 100% accurate, so errors in diagnostic status have to be accepted as part of the territory for a clinical epidemiologist (and statistician). In some cases, these errors can be specific to the population of study subjects assembled where, for example, cultural or language features may lead to misclassification of study subjects. For example, application of diagnostic and symptom criteria developed for white, western nations, may misclassify Aboriginal and Torres Strait Islander Australians at-risk for psychosis (Hamilton, 2008).

## Confounding

The third error that operates to cast doubt on research findings is that of *confounding*, from the Latin *confundere*, to mix together. Literally, when this is operating, the statistical results that have been observed could be accounted for by some other, perhaps undetected or unmeasured variable or relationship. There is an example in a study of cancer and schizophrenia. It is well known that a diagnosis of schizophrenia increases the risk for a number of physical ailments, such as obesity (with its accompanying increased cancer and mortality risk), poorer access to medical care, increased risk of suicide, and lower life expectancy. However, studies of cancer incidence in schizophrenia have been highly various, with some studies (Lichtermann et al., 2001) showing increased rates of cancer and others (Dalton et al., 2005) showing no increased risk and others still (Grinshpoon et al., 2005) showing a reduced risk. Given the increased risk factor load, such as very high smoking rates, the discrepancy between observed and expected cancer rates was acknowledged as an "epidemiological puzzle" (Jablensky and Lawrence, 2001). However, many previous studies comparing people with schizophrenia with the general population had failed to account for a prime risk factor for cancer: smoking. In a highly cited meta-analysis (Catts et al., 2008) of studies that used population register-based studies where the incidence of cancer in patients and their first-degree relatives was reported, when background population rates of smoking were adjusted for it was found that schizophrenia carried with it a *reduced* risk of cancer, not only in the primary sufferers but also in their first degree relatives. This has implications for genetic theories of cancer vulnerability in people with schizophrenia. It appears that smoking may have been an unacknowledged and unmeasured confounder of the relationship between schizophrenia and the incidence of cancer in previous studies.

Confounding occurs where a variable that is associated with exposure groups is also associated with the outcome. Unfortunately, confounding is not

detectable using statistical significance testing, or by obtaining ever larger samples. This is because statistical significance is a phenomenon determined by sample size - the larger the *n* the more likely there is to be a significant result – but confounding is a *state of nature*, a property of the subjects assembled for study, and independent of sample size. In strictly statistical terms, confounding is the detection of interaction among study variables, and is assessed in terms of multicollinearity, that is, the extent to which two variables share statistical variance.

Identification of potential confounders is totally dependent upon the state of knowledge at the time. Confounding appears wherever it is not wanted and may be mistaken for something else, such as an effect modifier, like where age modifies the prevalence of dementia. It may also be evidence of a variable that is on the causal pathway between an exposure and a disease, such as a poorer environment while growing up. From a statistical viewpoint, a significant interaction effect between an exposure variable, a potentially confounding third variable, and an outcome would be evidence of any of these situations: the statistics cannot discriminate. This puts the onus firmly on the researcher to interpret the finding in light of the known and available evidence: the statistic itself is neutral. The history of the advancement of knowledge in epidemiology is isomorphic with the discovery, recognition and control of confounding in causal inference. And it is comforting to know that there are only three kinds of errors or biases that can creep into clinical epidemiological studies.

## Critically Judging Research

What follows in Table 11.1 is a guide to appraising studies that is the culmination of the information contained above. Using this guide will help to appraise research papers and research studies – maybe even your own – in terms of the underlying basis of the statistical inferences that are used to support (or demolish) causal inferences drawn from research. It's a complex business. But it is the framework used by journal editors and scientific peer reviewers. And commended to those who want to ensure their practice is on safe grounds.

**Table 11.1.** Critical Appraisal in Clinical Epidemiology

| Critical Question | Is There a Problem? | Does it Threaten Validity? |
|---|---|---|
| 1. What is the research question? | Is it the impact of an intervention, a question of causality, or the magnitude of a problem? | Is it really worth doing? |
| 2. What is the study type? | Is the type of study appropriate to the research question? | If not, how useful are the results going to be? |
| 3. What are the outcome factors (dependent variables) and how are they measured? | Are all relevant outcomes assessed? Is there measurement error? Or differential measurement error? | How important are the omitted outcomes? Is differential or non-differential measurement error an important source of bias? |
| 4. What are the study factors (exposures, or independent variables) and how are they measured? | Is there measurement error? Or differential measurement error? | What effect does measurement error have on the outcome of the study? |
| 5. What is the reference population? The source population? The actual sample? Where did they come from? | Are there selection errors? Was the sample selected as a probability sample? What is the response rate? Is drop-out a problem? | Does this threaten the generalisability (external validity) of the study? |
| 6. In an experiment, how were subjects assigned to treatment? In a follow up study, how many were lost to follow up? | Is this explicit? Were any adjustments made for potential response bias? | Does this threaten the internal validity of the study? |
| 7. What important confounding variables were considered? | Are potential confounders identified and controlled for? | Is confounding an important source of bias? |
| 8. Are statistical tests used? | Are they appropriate to the level of measurement and the number of tests used? | Is it reasonable that significant results have not been obtained by chance? |
| 9. Are confidence intervals given for all major outcomes? | Are they appropriate to the level of measurement and the number of tests used? | Are the results clinically significant? |
| 10. If the results are negative, is a power analysis available? | Was the sample size sufficient to detect a clinically significant result? | Is the study useful, or is the result inconclusive? |
| 11. What conclusions did the author (or you) draw about the research question? Are new hypotheses generated from the study? | | Do you believe them? If so, why? If not, why not? |

## Further Reading

Das-Munshi, J., Ford, T., & Hotopf, M. (2020). Psychiatric epidemiology: Looking to the future. *Practical psychiatric epidemiology*, 425.
Lash, T.L., VanderWeele, T.J., Haneouse S., & Rothman, K. (2021) *Modern Epidemiology*, Wolters Kluwer Health.

## Cite as:

O'Toole, B. (2024). Psychiatric Epidemiology. In Boyce, P., Harris, A., and Malhi, G.S. (Eds.), *The Sydney textbook of psychiatry* (pp. 141–152). The University of Sydney.

# References

Catts SV, O'Toole BI, Carr VJ, et al. (2010) Appraising evidence for intervention effectiveness in early psychosis: conceptual framework and review of evaluation approaches. *Australian & New Zealand Journal of Psychiatry* 44(3): 195-219.

Catts V, Catts S, O'Toole B, et al. (2008) Cancer incidence in patients with schizophrenia and their first-degree relatives–a meta-analysis. *Acta Psychiatrica Scandinavica* 117(5): 323-336.

Dalton SO, Mellemkjær L, Thomassen L, et al. (2005) Risk for cancer in a cohort of patients hospitalized for schizophrenia in Denmark, 1969–1993. *Schizophrenia research* 75(2-3): 315-324.

Grinshpoon A, Barchana M, Ponizovsky A, et al. (2005) Cancer in schizophrenia: is the risk higher or lower? *Schizophrenia research* 73(2-3): 333-341.

Hamilton BA (2008) Assessment of at-risk mental states for psychosis in young aboriginal and non-aboriginal people using the CAARMS. Unpublished Ph.D. Thesis: University of Sydney.

Jablensky A and Lawrence D (2001) Schizophrenia and cancer: is there a need to invoke a protective gene? *Archives of general psychiatry* 58(6): 579-580.

Koenen KC, Moffitt TE, Poulton R, et al. (2007) Early childhood factors associated with the development of post-traumatic stress disorder: results from a longitudinal birth cohort. *Psychological medicine* 37(2): 181-192.

Lichtermann D, Ekelund J, Pukkala E, et al. (2001) Incidence of cancer among persons with schizophrenia and their relatives. *Archives of general psychiatry* 58(6): 573-578.

Lynch DL, Stern AE, Kim Oates R, et al. (1993) Who participates in child sexual abuse research? *Journal of Child Psychology and Psychiatry* 34(6): 935-944.

Martin PM and Martin-Granel E (2006) 2,500-year evolution of the term epidemic. *Emerging infectious diseases* 12(6): 976.

Miettinen OS (2007) Theoretical developments. *The development of modern epidemiology: Personal reports from those who were there*. 231-340.

Nasir BF, Toombs MR, Kondalsamy-Chennakesavan S, et al. (2018) Common mental disorders among Indigenous people living in regional, remote and metropolitan Australia: a cross-sectional study. *BMJ open* 8(6): e020196.

O'Toole BI (2000) Screening for low prevalence disorders. *Australian & New Zealand Journal of Psychiatry* 34(1_suppl): A39-A46.

O'Toole BI, Catts SV, Outram S, et al. (2009) The physical and mental health of Australian Vietnam veterans 3 decades after the war and its relation to military service, combat, and post-traumatic stress disorder. *American journal of epidemiology* 170(3): 318-330.

O'Toole BI, Marshall RP, Grayson DA, et al. (1996) The Australian Vietnam veterans health study: I. Study design and response bias. *International Journal of Epidemiology* 25(2): 307-318.

Porta M (2014) *A dictionary of epidemiology*. Oxford university press.

Robins LN (1981) National Institute of Mental Health Diagnostic Interview Schedule. *Arch Gen Psychiatry* 38: 381-389.

Smith GC and Pell JP (2003) Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *Bmj* 327(7429): 1459-1461.

Spitzer RL, First MB and Wakefield JC (2007) Saving PTSD from itself in DSM-V. *Journal of anxiety disorders* 21(2): 233-241.

Taylor DJ, Wilkerson AK, Pruiksma KE, et al. (2018) Reliability of the structured clinical interview for DSM-5 sleep disorders module. *Journal of Clinical Sleep Medicine* 14(3): 459-464.

Weathers FW, Bovin MJ, Lee DJ, et al. (2018) The Clinician-Administered PTSD Scale for DSM–5 (CAPS-5): Development and initial psychometric evaluation in military veterans. *Psychological assessment* 30(3): 383.