Finding Your Way Through the Forest – A TCM Practitioner's Guide to Evaluating Research: Part 1

Tony Reid

Abstract

Evidence based medicine is the prevailing paradigm of modern healthcare. However, practitioners of traditional Chinese medicine (TCM) vary significantly in their ability to appraise and understand modern research. This is the first in a series of articles that introduce the basic knowledge and skills needed to understand academic research and increase awareness of the limitations and problems associated with research methodology and statistical methods. This paper discusses evidence based medicine, bias, the peer review process, abstracts, and takes a detailed look at poor handling of statistical methodology using examples of relative risk, absolute risk and number needed to treat in the context of modern vaccine research.

Keywords
Evidence based medicine, research, clinical trials, TCM, Chinese medicine, bias, statistical methodology

In his work to amend various aspects of TCM in light of new scientific knowledge from the West, the eminent Qing dynasty physician Wang Qing-ren compared medicine to a forest in which one may easily lose one's way.[1] The situation is no less dire in today's world of intersecting paradigms, where we are not only engaged in the ongoing process of deepening our understanding of TCM, but also trying to make sense of an ever-expanding body of medical research and incorporate this into our daily clinical practice. This voluminous body of scientific medical research is no less a forest - in which abound dangerous beasts, vast areas of wilderness in which we may become lost, and thickets that can severely limit our vision.

This is the first part of a series of articles that discuss commonly occurring sources of error in clinical trial literature. The material is, of necessity, rather lengthy and dense. However, in the final part, a simplified table will be presented that can be used as a checklist when interpreting and evaluating clinical studies. In preparing this material I have assumed very little prior knowledge of statistical methodology and the principles of evidence-based medicine on the part of readers.

The human mind is not a blank slate, and we are all subject to multiple inherent biases – including our accumulated knowledge, experiences and personal prejudices. As with the 'observer effect' in physics, the preconceived notion of the observer (researcher) exerts a profound influence on that which is observed. In the context of academic research, this influence is mainly due to the researcher's own thoughts in the form of pre-conceived ideas that, when fuelled by self-interest, may dominate the mind and affect both perception and interpretation of the events under observation. While the current practice of declaring conflicts of interest is an important step to allow detection of potential biases, I believe that it does not go far enough. What I propose is that research authors declare their biases up front, including the basic

premises upon which research questions are framed. This is as much an exercise in self-awareness as it is in scientific integrity, and I would like to see it become standard practice. The following is a brief statement of my own biases:

1. Evidence based medicine, including clinical research methodology, is a work in progress, hindered mainly by the forces of inertia (ie conservatism within medicine) and vested interests (ie those who stand to gain in terms of money, prestige or political power from the results of medical research).
2. Bias, imprecision and uncertainty are inevitable facts of life, and these also exist within medicine and clinical research. The best we can do is to find better ways to control and reduce them - one of which is for authors and researchers to state their underlying assumptions up front.
3. There is value in all of the different forms of clinical research, which may in varying degrees be applied within clinical practice. This paper will focus on clinical trials, as these, together with systematic reviews and meta-analyses of the same, have been placed at the top of the evidence hierarchy.
4. Medicine is an art as much as a science, and the clinical encounter is always much more than the application of statistical knowledge to an over-simplified clinical problem.

Introduction: history of statistical methodology

Since the early 1960s statistical studies in medicine have moved from being a newly introduced innovation to the most acceptable way to verify medical theories and practices. However, the critical limitations of this methodology are often ignored and this, along with poor handling of appropriate statistical methods, has resulted in many false positive and non-replicable results in clinical research to date.[2] Statistical studies have played an increasingly important role within medicine since the 1960s through the pioneering work of Hill, based on methodology developed by Fisher in the 1930s. This types of study (ie statistical) underpinned advances in diagnosis and treatment during the 1970s. Unfortunately, the derailing influence of vested interests soon became increasingly apparent, beginning in the1980s and extending to the present day. These issues have been well documented and discussed elsewhere.[3,4,5]

In the 1990s the medical profession adopted a system for ranking clinical studies and other sources of medical knowledge, including the development and implementation of methodologies to reduce bias in medical research. This initiative was spearheaded by a group of Canadian epidemiologists headed by Sackett. Thus was born what is now termed evidence based medicine and evidence based practice. Concurrently, since the mid 1990s the CONSORT group (focused on clinical trials) and the PRISMA group (focused on systematic reviews and meta-analyses of trials) have periodically issued statements in an effort to develop and disseminate international standards for transparent reporting of medical research.[6]

Guidelines for assessing clinical trials have been developed and refined since the 1960s. Currently accepted models include the Jadad scale, the Physiotherapy Evidence Database scale (PEDro) and the Cochrane Collaboration's risk of bias tool.[7] These, together with the CONSORT and PRISMA statements mentioned above have been widely published, with the aim of improving and standardising both the design and reporting of clinical trials. They have been incorporated into university curriculums and are generally accepted within the medical profession. However, at the time of writing there are still many clinical trials and reports of trials that either fall short of these standards or introduce unacknowledged sources of error.[2]

How to read an academic paper

When you read an academic paper in which other works are cited, these other works are usually research studies, review papers, official clinical guidelines or pages from an authoritative textbook. The purpose behind including citations is twofold. Firstly, to provide a source from which to verify the facts that are being used in the discussion. Secondly, some references may direct readers to an article or a chapter in a book in which there is a more comprehensive and detailed discussion of issues that have only been mentioned quite briefly. Therefore, an important part of reading an academic paper is to include at least a cursory glance at the references and then, if necessary, to access a particular source, either to check the facts or to gain a deeper understanding of the issues involved by reading what others have to say on the subject. You should not always accept a fact or a viewpoint simply because a reference has been cited. The following discussion aims to illustrate why this process of additional scrutiny is absolutely necessary. You may have presumed that the process of scrutinising source material has already been done for readers through peer review. The inconvenient truth is that the process of peer review, in which qualified experts read and critically evaluate a paper before publication, is subject to a considerably greater degree of inconsistency and inaccuracy than we would expect from normal human error. Let us take a small but important example - research paper abstracts. A survey conducted in 1999 found that up to 68 per cent of abstracts in papers from medical journals are either false or misleading, and the situation remained quite

poor over the following ten years.[8,9] With the subsequent publication of the CONSORT statements in 2001 and then in 2010, which included guidelines for proper reporting of abstracts, it was hoped that the situation may have improved somewhat.[10] However, this appears not to be the case.[11] The 'abstract' is a short summary of findings at the beginning of a paper that is freely accessible through an online search for scientific papers on a particular subject. While many academic papers are available free of charge, most are not, and journals may charge 30 to 50 US dollars for access to a complete paper. This can add up quite quickly when you are searching for information on a particular topic. It is not surprising, therefore, that the abstract is the most frequently read portion of scientific papers available on PubMed.[12] The purpose of the abstract is to provide a concise and accurate summary of the paper, highlighting the main content, the purpose of the research, the relevance or importance of the work, as well as the main outcomes together with sufficient supporting data.[13] Unfortunately, quite a large number of abstracts are inaccurate or misleading, in spite of peer review. There are several reasons for this, and these also apply to research papers in general:[7,8,10]

- 'Publication bias': This leads to pressure on researchers to come up with something positive, as negative trials tend not to be published.
- Vested interests: Sometimes those conducting a study have a vested interest in the drug or treatment protocol appearing to be more effective and safer than it really is. The fact that most clinical trials on drug treatments are funded - and often conducted - by pharmaceutical companies, is pertinent in this regard.
- Poor understanding of statistics by researchers and reviewers.
- Sloppy reporting, which may have been ignored because of a reported positive outcome.

Please, take a moment to let this fact sink in - around half of the abstracts that appear in medical research papers are not supported by the actual findings in the paper. This is a clear sign that the peer review process is deeply flawed, and you should not therefore fully place your trust in the information gained from the abstract of an academic paper. Unfortunately, this is just the tip of the iceberg. The situation only gets worse when we examine the other parts of an academic paper, specifically papers within the clinical trial literature.[14] Indeed, the situation has become so dire that some believe that 'the time may have come to stop assuming that research actually happened and is honestly reported, and assume that the research is fraudulent until there is some evidence to support it having happened and been honestly reported'.[15] These and related issues form the body of this series of articles and will be discussed at greater length in subsequent instalments.

Poor handling of statistical methodology

In a recent review which discusses the widespread poor handling of statistical methodology in medical research literature, we find the following bold, but comforting, statement: 'the most important thing clinicians should know about statistics, are not formulas but basic concepts.' The author proposes that the best way for medical researchers to avoid the common statistical pitfalls is to design and analyse their studies in consultation with a qualified statistician.[16] The implication here is that medical researchers, who may have only studied a single unit of medical statistics as undergraduates, should not attempt to apply statistical science in their research, but rather leave it to the experts. In the same way that a physician will refer patients to a surgeon when there is a surgical problem, a specialist statistician should be given charge of this component of a medical study. Otherwise, problems can, and frequently do, arise. Before discussing basic statistical concepts, I would like to begin with an example that is relevant to the global situation in which we now find ourselves that pertains directly to statements about important public health initiatives, such as COVID-19 vaccination.

Relative risk, absolute risk and number needed to treat

When looking at new treatments or strategies that are aimed at avoiding bad health outcomes, the results of such a study may be expressed in three ways: the relative risk reduction (RRR – often shortened to 'risk reduction', RR), the absolute risk reduction (ARR) and number needed to treat (NNT). Only the last two give a concrete and practical picture of the real-world application of the intervention in question. The RRR is routinely used by government health authorities and agencies, as well as doctors, when they want to encourage the general public or a patient to take up a particular intervention that is aimed at reducing a particular health outcome.[17,18] As an example, the results of the same study are presented below according to these three different ways of describing the outcome:

- RRR: If you have this test every two years, it will reduce your chance of dying from this cancer by around one-third over the next 10 years. (RRR = 1–2/3, i.e. 33.3 per cent)
- ARR: If you have this test every two years it will reduce your chance of dying from this cancer from around three in a 1000 to around two in a 1000 over the next 10 years.
- NNT: If around 1000 people have this test every two years, one person will be saved from dying from this cancer every 10 years.

Not surprisingly, it has been demonstrated that patients were more likely to take this test when it was presented in RRR terms, and least likely to take it when presented in NNT terms.[16] In this way, the RRR, while being grossly misleading, but nevertheless a 'statistically valid' measurement, has become the measurement of choice to use in the context of convincing a layperson to follow medical advice. This is justified by being regarded as a 'win-win' tactic by government agencies, pharmaceutical companies, as well as the medical profession. Driven by such factors, a statistical calculation which is misleading and inappropriate has found its way into the front line of descriptors that are used in presenting the results of clinical studies.[16,17]

The RRR tells you how the results compare between the two groups in a trial. The number of people receiving the intervention who had the bad outcome (eg contracting a COVID-19 infection) is looked at in comparison to the number of people who had the same outcome in the control group (ie those not receiving any treatment). The RRR is simply the two numbers divided, and then expressed as the percentage reduction in risk of obtaining the bad outcome when the treatment in question is applied (eg a new vaccine). It is calculated by taking the fraction obtained by dividing the two results, subtracting it from one, and then expressing this number as a percentage. The RRR is therefore simply a way of comparing the two groups in a trial and does not tell you anything about how the trial results can be applied to a larger community. Referring to the previous example, the original trial data compared two groups of women, 1000 in each group, over a 10-year period, in which one group underwent screening for breast cancer every two years while the other did not. There were two cancers in the screened group and three in the unscreened group within the 10 year study period. From these results, the above values for RRR, ARR and NNT were obtained. In this case the RRR is calculated as follows: one minus two thirds, which equals one third, which becomes 33.3 per cent. As you can see, if a person is told that an intervention will reduce the risk of cancer by 33.3 per cent, this may be easily misinterpreted to mean that out of 100 women who would normally have got breast cancer, about 33 of them could have avoided it if they had undergone breast screening every two years. Unfortunately, this is not what the RRR is telling us. Moreover, in terms of the numerical value of the RRR, its inapplicability as a measure of clinical effectiveness stems from the fact that the same result is obtained every time you have three in one group and two in the other – regardless of how many subjects were in each of the two groups. You will get the same RRR value with groups of 100, 1000 or one million. Obviously, with smaller numbers of subjects, the results become increasingly more compelling – and vice versa. This is indeed a blunt instrument: it is both misleading and inaccurate, and as such, erroneous.

The ARR, on the other hand, while still being a mathematical expression, brings us much closer to a clear expression of what the results of a trial are really telling us. The ARR is meant to give an indication of the net effectiveness of an intervention compared to no intervention, and is the difference between the results of the no-treatment group (ie the control group) and the treatment group, all expressed as percentages. Thus, it is calculated according to the normal guiding principle of subtracting the placebo effect from the treatment effect; it is a valid and significant metric. In the above example of breast screening the two cancer deaths in the screened group (two out of 1000 equals 0.2 per cent) are subtracted from the three cancer deaths in the non-screened group (three out of 1000 equals 0.3 per cent) to give the real (or 'absolute') reduction in risk that is associated with bi-annual breast screening over a 10-year period (0.1 per cent).

Finally, the NNT spells out the data in the clearest fashion and tells you how many people need to receive a treatment, undergo screening, modify their lifestyle, etc. in order to enable one person to avoid a specific bad health outcome.[16,17,19]

A current example of the way in which the RRR is being used to sway public opinion and encourage people to receive a treatment relates to the newly developed agents to combat SARS CoV-2 infection. Although they are still in the experimental stage - and according to expert opinion, several steps are required before full authorisation for their use is justified[20] - the efficacy of these agents has been universally expressed and promoted using the RRR - the 'best' one having an RRR of 95 per cent.[21] Put another way, according to the trial results of this particular agent, the NNT is 119: so out of 119 people treated, one will be protected.[22] Or, with an ARR of 0.84 per cent, the real reduction in risk of contracting SARS CoV-2 infection in those that have received 'the jab' is less than one per cent. Moving on to the real-world scenario in Israel, where a large portion of the population have been vaccinated, an analysis of the outcomes gives an RRR of 94 per cent and a NNT of 217.[20] This equates to one person protected out of 217 that have been innoculated. This is another example of how misleading the RRR can be: a one per cent decrease in the RRR (as is widely reported in the media), translates to approximately half of the original degree of protection that was found in the original trial (i.e. one person out of 217 versus one out of 119). This latest analysis of vaccine effectiveness from academics in the UK and Luxemburg has the fact checkers - even those from reputable news sources - scrambling to reassure their readers that somehow the RRR

is much more valid and meaningful than the ARR.[23] None of the fact-checkers mentions the NNT.

The fact that we are often exposed to both media reports and abstracts that are misleading speaks to the ongoing need to be wary of the aforementioned derailing factors: publication bias, vested interests, poor understanding of statistics by researchers and reviewers, and sloppy reporting. There is not a lot we can do about the first two items on this list, but now we will never be fooled again by the relative risk reduction metric. There are many other aspects of 'creative' statistical reporting against which we can also be armed. These will be discussed in our next instalment.

Tony Reid is a graduate of the Sydney Institute of Traditional Chinese Medicine and holds master's degrees in acupuncture and TCM from the University of Western Sydney. He has contributed to TCM as a clinician, lecturer, administrator, course designer and industry consultant since the early 1980s.

Endnotes

1. Wang, Q., transl. Chung, Y., Oving, H., Becker, S. (2007). *Yi Lin Gai Cuo (Correcting the Errors in the Forest of Medicine)*. Blue Poppy Press: Boulder, CO.
2. Evans, S. (2010). Common Statistical Concerns in Clinical Trials. *J Exp Stroke Transl Med,* 3(1)1-7.
3. Healey, D. (2004). *Let Them Eat Prozac – the Unhealthy Relationship Between the Pharmaceutical Industry and Depression*. NYU Press: New York.
4. Healey, D. (2012). *Pharmageddon*. University of California Press.
5. Gotzsche, P. (2012). *Deadly Medicines and Organised Crime – How Big Pharma Has Corrupted Healthcare*. CRC Press: Florida.
6. CONSORT stands for Consolidated Standards of Reporting Trials, and encompasses various initiatives developed to alleviate the problems arising from inadequate reporting of randomised controlled trials. For more details, go to <http://www.consort-statement.org/> PRISMA stands for Preferred Reporting Items for Systematic Reviews and Meta-Analyses - see <http://www.prisma-statement.org>.
7. Duke University website, Medical Center Library & Archives (N.D.). Systematic Reviews: the process: Appraisal & Analysis from <https://guides.mclibrary.duke.edu/sysreview/analysis> [retrieved 11-05-2020].
8. Berwanger, O., Ribeiro, R., Finkelsztejn, A. et al. (2009). The quality of reporting of trial abstracts is suboptimal: survey of major general medical journals, *J Clin Epidemiol,* 62(4):387-92.
9. Pitkin, R., Branagan, M., Burmeister, L. (1999). Accuracy of data in abstracts of published research articles, *JAMA,* 31;281(12):1110-1.
10. Consort 2010 Statement, Extensions – Abstracts, from: <http://www.consort-statement.org/extensions?ContentWidgetId=562> [Retrieved 19-04-2020].
11. Choi, S., Cheung, C. (2016). Don't judge a book by its cover, don't judge a study by its abstract. Common statistical errors seen in medical papers, *Anaesthesia,* 71. 10.1111/anae.13506.
12. Islamaj Dogan, R., Murray, G., Névéol, A., Lu, Z. (2009). Understanding PubMed user search behavior through log analysis. Database (Oxfrod): 2009, bap018.
13. University of Melbourne. (2020). Academic Skills, Writing an Abstract. Retrieved 17th April, 2020 from: www.services.unimelb.edu.au/academicskills
14. Ioannidis JP. (2005). Why most published research findings are false. PLoS Med. 2(8):e124.
15. Smith, R. (2021). Time to assume that health research is fraudulent until proven otherwise?. *BMJ Opinion*, from <https://tinyurl.com/28cufk3p> [retrieved 10-08-2021]
16. Evans, S. (2010). Common Statistical Concerns in Clinical Trials. J Exp Stroke Transl Med. 3(1)1-7.
17. Naik, G., Ahmed, H., Edwards, A. (2012). Communicating risk to patients and the public. Br J Gen Pract. 62(597):213-6.
18. McCormack, L., Sheridan, S., Lewis, M., et al. Communication and Dissemination Strategies to Facilitate the Use of Health-Related Evidence. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 Nov. (Evidence Reports/Technology Assessments, No. 213.), from <https://www.ncbi.nlm.nih.gov/books/NBK179104/> [retrieved 19-08-2021]
19. Irwig, L., Irwig, J., Trevena, L., et al. (2008). Smart Health Choices: Making Sense of Health Advice. London: Hammersmith Press. Chapter 18, Relative risk, relative and absolute risk reduction, number needed to treat and confidence intervals, from <https://www.ncbi.nlm.nih.gov/books/NBK63647/> [retrieved 19-08-2021]
20. The criteria by which the new types of COVID-19 vaccines can be regarded as having passed through the experimental stage include:
    - Completing at least two years of follow-up of participants originally enrolled in pivotal clinical trials, even if the trials were unblinded and now lack a placebo control. All vaccine manufacturer phase three trials were already designed with this planned duration.
    - Ensuring, prior to including in the list of populations for which a vaccine is approved, that there is substantial evidence of clinical effectiveness that outweighs harms in special populations such as: infants, children and adolescents; those with past SARS-CoV-2 infection; the immunocompromised; pregnant women; nursing women; frail older adults; and individuals with cancer, autoimmune disorders, and haematological conditions.
    - Requiring thorough safety assessment of spike proteins being produced in-situ by the body tissues following administration, and spike proteins' full biodistribution, pharmacokinetics and tissue specific toxicity.
    - Completion of vaccine biodistribution studies from administration site and safety implications of mRNA translation in distant tissues.
    - Thorough investigation of all severe adverse reactions reported following COVID-19 vaccination, such as deaths, reported in the United States and global pharmacovigilance systems.
    - Assessment of safety in individuals receiving more than two doses.
    - Inclusion of gene delivery and therapy experts in the Vaccines and Related Biological Products Advisory Committee, in recognition of the fact that the novel COVID vaccines work on the premise of gene delivery, in contrast to conventional vaccines.
    - Enforcing stringent conflict of interest requirements to ensure individuals involved in data analysis and Biologics License Application-related decision-making processes have no conflict of interest with vaccine manufacturers.
    See Tanveer, S., Rowhani-Farid, A., Hong, K., et al. (2021). Transparency of COVID-19 vaccine trials: decisions without data, BMJ Evidence-Based Medicine Published Online, doi: 10.1136/bmjebm-2021-111735 ; also Wastila, L., Doshi, P., Merchant, H., et al. (2021). Citizen Petition from Linda Wastila, Document ID FDA-2021-P-0521-0001, from: <https://tinyurl.com/k95v6kmv> [retrieved 19-08-2021]; see also Burns, P., Rohrich, R., Chung, K. (2011). The levels of evidence and their role in evidence-based medicine, *Plast Reconstruct Surg,* 128(1), 305–310.
21. Polack, F., Thomas, S., Kitchin, N., et al. (2020). Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine, *N Engl J Med,* 383(27), 2603–2615.
22. Olliaro, P., Torreele, E., Vaillant, M. (2021). COVID-19 vaccine efficacy and effectiveness-the elephant (not) in the room, *Lancet Microbe,* 2(7):e279-e280.
23. Reuters Fact Check (2021). Why Relative Risk Reduction, not Absolute Risk Reduction, is most often used in calculating vaccine efficacy, from <https://www.reuters.com/article/factcheck-thelancet-riskreduction-idUSL2N2NK1XA> [retrieved 31-07-2021].