Understanding Data; An introduction to critical appraisal

GEORGINA GETHIN

Royal College of Surgeons in Ireland, Dublin, Ireland

Some basic information on how to help clinicians in understanding the results of papers are presented. Evidence based medicine is discussed and accepted concept of hierarchy of evidence presented in figure. Randomized control trial (RTC) is stressed as gold standard for evaluating the effectiveness of interventions. The examples of exercises (descriptive and inferential statistics) for understanding results are given. It is concluded that once one has a grasp of the basic steps outlined in presented text, the way to developing knowledge around evidence based practice is achieved.

KEY WORDS: data, statistics, critical appraisal

ADDRESS FOR CORRESPONDENCE: Dr. Georgina Gethin
Director of Research and Professional Development
Lecturer Wound Science and EBP
Royal College of Surgeons in Ireland
Dublin, Ireland
E-mail: ggethin@rcsi.ie

INTRODUCTION

In addressing the many challenges in everyday practice, the clinician relies on a variety of sources including personal experience, communication with other colleagues, knowledge gained via education and training and through a search of the literature. However, the volume of literature published on a daily basis means that the busy clinician must develop some of the skills which are necessary to identify the relevant literature, appraise it, and determine how applicable this is to their own practice setting. Unfortunately, anecdotal evidence would suggest that many people rely on the conclusion and abstract from a paper to inform them of the outcomes of a study as they do not have the necessary skills to interpret the statistical analysis. This paper will provide some basic information which any clinician can apply to a paper and help them in understanding the results. The type of studies used as examples are limited to randomized controlled trials of interventions as these are one of the highest levels of evidence.

WHAT IS EVIDENCE BASED MEDICINE?

Evidence based medicine (EBM) is defined as 'the integration of best research evidence with clinical expertise and patient values'(1). This is a process of life-long self-directed learning and allows the integration of good quality published literature with clinical expertise and the opinions and values of patients and their families or carers (2). Indeed it is argued that EBM should also incorporate expertise in performing clinical history and physical examination (2). EBM is important to improving the quality of patient care, as it contributes to identifying those interventions that work and the elimination of those that are ineffective or do not work. There is an accepted concept of a hierarchy of evidence (Fig. 1). The hierarchy has an order, advancing from the simple case studies and opinions, literature reviews through to more advanced methodologies such as the randomised controlled trial (RCT), systematic reviews and meta-analysis. The RCT is considered the gold standard for evaluating the effectiveness of interventions. It is defined as a quantitative, comparative,

controlled experiment in which a group of investigators study two or more interventions in a series of individuals who receive them in a random order (3) while the intention is to make the research objective, the results will only really apply to the limits set within the trial or to the specific population being studied (4).

The population in a research study is the entire group of people with the specific problem, for example all people with venous leg ulcers. However, it would not be possible to study an entire population unless a condition is extremely rare. Therefore, a study tries to obtain a 'sample' of that population. By stating the inclusion criteria such as those with open venous leg ulcers and the exclusion criteria such as patients taking steroids, the researcher sets the limits to which the results of the study will apply. The essential feature is to make patients in the trial representative of all future patients who are liable to benefit from the trial's findings (5).

In the randomised controlled trial (RCT) the study 'sample' is then randomly allocated to one or other intervention and followed up for a specific period of time. The two groups are usually referred to as the 'intervention' and 'control'. The intervention group receives the treatment or intervention of interest which is being investigated. This is then compared to the control group which is often similar in some way to usual or standard care. For example, in a study by Gethin and Cowman (6) they compared honey (intervention group) with hydrogel (control group) to determine debriding efficacy in venous leg ulcers. The results therefore can be applied to those patients with sloughy venous ulcers.

WHAT DO THE RESULTS SAY?

When reading the trial results there are a few simple steps that the individual can take to gain an understanding of the results. This is not a definitive guide to understanding trial results and there are some excellent text and on-line resources available. There are two types of statistics: descriptive and inferential. Descriptive statistics simply summarise and describe the results. These include the mean; median and mode. Inferential statistics are used to make inferences about a situation that has not yet been observed.

Interpretation; In this example the average age of people attending the clinic is 70 years but the median age is 74 years. The median tells us that half of all the people attending are older than 74 and half are younger. The median tells us more about the people attending and is not as influenced by those which are very different from the rest of the group (called outliers). In this case, one person is 33 years which is much younger than the rest of the group – the inclusion of this person changes the mean but it does not influence the median as much. In wound care studies you should take note of the median wound size and median duration of a wound rather than the mean.

The data call also be displayed in quartiles [quarters]. In this set of data, if we broke up all the ages into quarters we can see that one quarter of all people are less than 63 years; one quarter are over 82 years and half of all people are between the ages of 63 and 82. This type of information helps you to understand the characteristics of your group in greater detail than if you simply said the mean age, which happens in so many reports. This also helps you in interpreting results and you can see if the characteristics of the study group are similar to those in your practice.

APPRAISING AN RCT

The Graphic Appraisal Tool for Epidemiological Studies (GATE) framework is an excellent tool for displaying the results and helping to understand the results of an RCT. Available on-line at www.cebm.net To appraise the results of one RCT we will refer to the study by Gethin and Cowman (6).

Population: in this study the population of interest are persons with sloughy venous ulcers. 108 people were

---

Exercise 1: mean, median and mode:

Sample: age profile of 21 people attending a wound clinic.

86,75,76,75,82,68,85,83,82,75,63,69,54,58,33,70,74,67,63,64,84

Mean: this is the average age of people attending the clinic. Add all 21 ages together and divide by the number of people. Answer: 70 years.

Median: this is the mid-point. In this set of ages, you should arrange all the ages in order from the youngest to the oldest. Answer: 74 years

33,54,58,63,63,64,67,68,69,70,74,75,75,75,76,82,82,83,84,85,86

Mode: this is the most frequently occurring number. Answer:75 years.

Exercise 2: healing outcomes and relative risk

Intervention group outcome: 24 out of 54 people healed: $24 \div 54 = 0.44$ or 44%

Control group outcome: 18 healed out of 54 people: $18 \div 54 = 0.33$ or 33%.

Your first questions should be: is this clinically meaningful; am I impressed with this result? Would this percentage of healing in either group be important to me?

Relative Risk (RR): calculated by dividing the intervention group outcome by the control group outcome: $0.44 \div 0.33 = 1.33$.

Interpretation: How many times more likely it is that healing will occur in the intervention group (honey) relative to the control group (hydrogel). An RR of 1 means that there is no difference between the two groups, thus the intervention had no effect. An RR of < 1 means that the treatment decreases the risk of the outcome (healing). An RR > 1 means that the treatment increased the risk of the outcome (healing). Since the RR is > 1 we can say that honey increased the risk of healing.

included in the study and randomised to two groups of 54 and followed-up for 12 weeks.

The intervention group received Manuka honey (n=54); the control group received Hydrogel therapy (n=54).

There are a few outcomes but for now lets look at healing at 12 weeks as an outcome. This is plotted on a 2 x 2 table. You can use this type of table for each outcome of interest.

Intervention Group	Control Group
A) numbers healed = 24	B) numbers healed = 18
C) numbers not- healed = 30	D)numbers not-healed = 36

P-VALUES AND CONFIDENCE INTERVALS

When we pick a sample, they should be similar in some way to the entire population of people with that condition. By knowing the characteristics of the sample in a study we can see if they would represent those with the condition or if they are representative of those in our practice. For example; are the ages similar to those with venous ulcers; are they being treated in the same setting as I would expect, do they have the same problems such as hypertension. This is important not only for your understanding of the study but also for whether or not you would apply the results to your practice. It is also important for understanding confidence intervals (CI).

When we see the results of a study, these are in fact the results from this sample alone. How confident are we that if we got another sample from the population with the problem that we would get similar results? In the case of Gethin and Cowman (6) we see the confidence interval for healing at 12 weeks is presented as 1.02 to 1.88. This means that we are 95% confident that the true result for the entire population lies somewhere between 1.02 and 1.88. For this study, the CI is narrow

which is good. If we see a study with a very wide CI this means that while the results lies somewhere between these two points it is so wide that it might be meaningless.

In our study the p-value for healing at 12 weeks is p = 0.03. Now let's convert that to percentages; 0.03 = 3%. Thus, we are saying that the likelihood of this result happening by chance is around 3%. In statistical terms, the level of significance is usually set at 0.05 or 5%. By stating the level of significance we are proposing that any result that is higher than that is not statistically meaningful as these results could have occurred by chance. For example: a p value of 0.08 means that there is an 8% likelihood that this result could have occurred by chance. So, in our example of Gethin and Cowman (6) we can interpret p = 0.03 as likelihood of healing at 12 weeks having occurred by chance is only 3%. When the level is 0.05 or less we say that it is statistically significant. Importantly this does not mean it is clinically significant – that is up to you.

FINALLY: INTENTION TO TREAT ANALYSIS

The data analysis section of a study will indicate whether analysis was done on a 'per protocol' or an 'intention to treat' (ITT) basis. The first thing to look for is, are all the people who entered the study accounted for at the end. Do not assume that they are. In the case of ITT, all people who enter the study are analysed at the end, regardless of whether they complied with the treatment regime or not. In that way, it is argued, that this type of analysis represents routine practice whereby some people will not comply with a treatment or simply do not complete the treatment. In per protocol analysis the results are analysed only on those that complete the study. The danger with per protocol analysis is that the treatment effect may be overstated. Some studies will provide both scenarios.

STUDY REPORT

In most cases the only information that the practitioner has to interpret a trial is the published report. To improved the quality of reports of trials and standardise information, an international group of epidemiologists, bio statisticians and journal editors published a statement called CONSORT (Consolidation of the standards of reporting trials), (7). This is updated frequently and the most up to date version can be accessed on-line http://www.consort-statement.org/home/

The CONSORT statement comprises a 25-item checklist and a flow diagram. The intention is to make the study process clearer. This format ensures information regarding all aspects of a trial are reported on. However, this does not overcome publication bias. Some evidence shows a propensity for investigators and sponsors to write and submit and for peer-reviewers and editors to accept, manuscripts for publication, depending on the direction of the findings (3). This tendency, which appears to favour trials with positive results, has been called publication bias.

Publication bias is a major problem in professional literature, positive results being more likely to get submitted and published (8). In the absence of reports of studies with 'negative' results, readers can draw conclusions (often incorrect) from a skewed and incomplete database. Provided the methodology is robust, all research results offer valuable information and knowledge to the field and should be published. In addition, the Declaration of Helsinki (1964) states that negative as well as positive results of research should be published or otherwise publicly available (9).

This short paper should provide enough information for the novice to make some interpretation of the results of an RCT. Other factors to be considered are: how subjects were recruited, how they were allocated, who assessed the outcomes, if the assessor knew which treatment the patient received, how outcomes were assessed, the duration of follow-up of the study, the number and type of adverse events and the number and reasons for withdrawals. It is beyond the scope of this paper to address all of these but they are important factors which should not be ignored.

CONCLUSION

The conduct of simply analysis on the report of an RCT is within the capacity of most clinicians. The advantage of performing these simple tests is that is allows one to made a better interpretation of the results and help determine if these results apply to one's own area
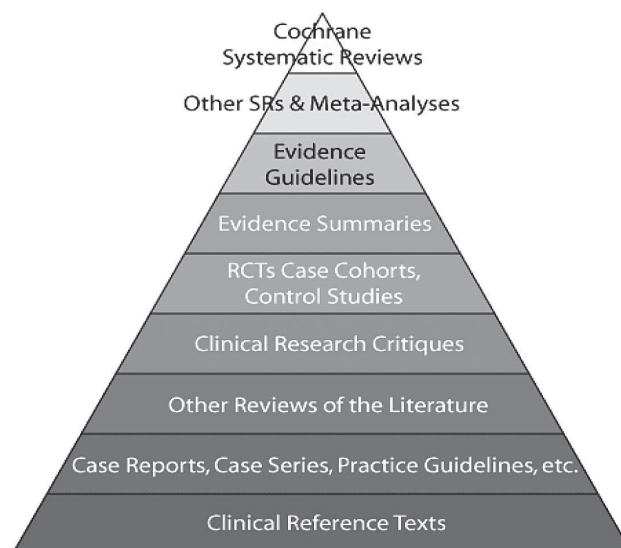

Fig. 1: Hierarchy of evidence

of practice. There are many excellent text and on-line resources that can assist with a more detailed understanding of statistics and once you have a grasp of the basic steps outlined here, you are on the way to developing your knowledge around evidence based practice.

Frequently used Terminology

- Odds Ratio: the odds of the outcome occuring in experimental group compared to the odds of it occuring in the control. (May be used more appropriately in case control and cohort studies.

- Confidence Intervals: 95%; the range in which we can be approximately 95% certain that the true population value lies.

- P Value: probability that a difference between groups would have occured if the null hypothesis was true.

- Relative Risk: RR: How many times more likely it is that an event will occur in the treatment group relative to the control group.

- ARR: Absolute risk reduction; the absolute difference in the rate of events between two groups. Gives an indication of baseline risk and treatment effect. ARR of 0 = no effect.

- RRR: Relative risk reduction. The reduction in the rate of the outcome in the treatment group relative to the control.

- NNT: The number need to treat in the experimental group to prevent 1 bad outcome.

- ITT: A method of analysis in RCT in which all patients randomised are analysed, regardless of whether or not they completed or received the treatment.

- Per-Protocol Analysis: A method of analysis in which only those patienta who completed the treatment to which they were originally allocated are analysed.

REFERENCES

1. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine. How to practice and teach EBM. London: Harcourt, 2000.
2. Akobeng A. Principles of evidence based medicine. Arch Dis Child 2005; 90: 837-40.
3. Jadad A. Randomised Controlled Trials. LondoN :British Medical Journal Books, 1998.
4. Gethin G. Understanding research. Wounds UK 2009; 5: 86.
5. Pocock SJ. Clinical trials: a practical approach. London: John Wiley & sons, 1983.
6. Gethin G, Cowman S. Manuka honey vs. hydrogel - a prospective, open label, multicentre, randomised controlled trial to compare desloughing efficacy and healing outcomes in venous ulcers. J Clin Nursing 2009; 18: 466-74.
7. Begg C, Cho M, Eastwood S. Improving the quality of reporting of randomised controlled trials. JAMA 1996; 276: 637-9.
8. Machin D. General Issues. In Machin D, Day S, Green S, eds.Textbook of Clinical Trials Wiley, England, 2004.
9. WMA. Declaration of Helsinki, ethical principles for medical research involving human subjects. In: 52nd WMA General Assembly (Assembly,W M, ed. Edinburgh, Scotland, 2000.

SAŽETAK

RAZUMIJEVANJE PODATAKA; UVOD U KRITIČKU PROCJENU

G. GETHIN

Kraljevski kirurški zbor u Irskoj, Dublin, Irska

Prikazani su neki osnovni podaci kako pomoći kliničarima u razumijevanju rezultata radova. Raspravlja se o medicini temeljenoj na dokazima i o prihvaćenom konceptu hijerarhije dokaza, što je izneseno u slikovnom prikazu. Naglašeno je da je randomizirani kontrolirani pokus (RTC - *Randomised Control Trial*) zlatni standard za evaluaciju učinkovitosti intervencije. Izneseni su primjeri vježbi deskriptivne i inferentne (prosuđivačke) statistike za razumijevanje rezultata. Naglašeno je da je razumijevanje temeljnih koraka koji su naglašeni u ovom radu put prema stvaranju vlastitog znanja o praktičnom radu temeljenom na dokazu. Na kraju su dane definicije često upotrebljavanih statističkih termina.

KLJUČNE RIJEČI: podaci, prosuđivačka statistika, kritička procjena