

Guidelines for Critical Review Form - Quantitative Studies

© Law, M., Stewart, D., Pollock, N., Letts, L., Bosch, J., & Westmorland, M., 1998

Introduction

- These guidelines accompany the Critical Review Form for Quantitative Studies developed by the McMaster University Occupational Therapy Evidence-Based Practice Research Group (Law et al. 1998). They are written in basic terms that can be understood by clinicians, students and researchers.
- Where appropriate, examples and justification for the guidelines/suggestions are provided to assist the reader in understanding the process of critical review.
- Guidelines are provided for the questions (left hand column) in the form, and the instructions/questions in the Comments column of each component.

Critical Review Components

Citation

- Include full title, all authors (last name, initials), full journal title, year, volume # and page #s.
- This ensures that another person could easily retrieve the same article.

Study Purpose

- Was the purpose stated clearly? - The purpose is usually stated briefly in the abstract of the article, and again in more detail in the introduction. It may be phrased as a research question or hypothesis.
- A clear statement helps you determine if the topic is important, relevant, and of interest to you. Consider how the study can be applied to occupational therapy practice and/or your own situation before you continue. If it is not useful or applicable, go on to the next article.

Literature

- Was relevant background literature reviewed? - A review of the literature should be included in an article describing research to provide some background to the study. It should provide a synthesis of relevant information such as previous work/research, and discussion of the clinical importance of the topic.
- It identifies gaps in current knowledge and research about the topic of interest, and thus justifies the need for the study being reported.

Design

- There are many different types of research designs. The most common types in rehabilitation research are included.
- The essential features of the different types of study designs are outlined, to assist in determining which was used in the study you are reviewing.

- Some of the advantages and disadvantages of the different types of designs are outlined to assist the reader in determining the appropriateness of the design for the study being reported.
- Different terms are used by authors, which can be confusing - alternative terms will be identified where possible.
- Numerous issues can be considered in determining the appropriateness of the methods/design chosen. Some of the key issues are listed in the Comments section, and will be described below. Diagrams of different designs, and examples using the topic of studying the effectiveness of activity programmes for seniors with dementia, are provided.
- Most studies have some problems due to biases that may distort the design, execution or interpretation of the research. The most common biases are described at the end of this section.

Design Types

1. Randomized (RCT)

- Randomized Controlled Trial, or Randomized Clinical Trial: also referred to as Experimental or Type 1 study. RCT's also encompass other different methods, such as cross-over designs.
- The essential feature of an RCT is a set of clients/subjects are identified and then randomly allocated (assigned) to two or more different treatment "groups". One group of clients receives the treatment of interest (often a new treatment) and the other group is the "control" group, which usually receives no treatment or standard practice. Random allocation to different treatment groups allows comparison of the client groups in terms of the outcomes of interest because randomization strongly increases the likelihood of similarity of clients in each group. Thus the chance of another factor (known as a confounding variable or issue) influencing the outcomes is greatly reduced.
- The main disadvantage of RCT's is the expense involved, and in some situations it is not ethical to have "control" groups of clients who do not receive treatment. For example, if you were to study the effectiveness of a multidisciplinary inpatient program for post-surgical patients with chronic low back pain, it may be unethical to withhold treatment in order to have a 'control' group.
- RCT's are often chosen when testing the effectiveness of a treatment, or to compare several different forms of treatment.

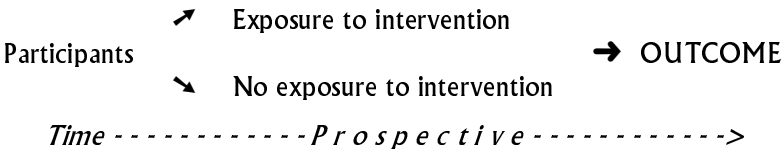


Example: The effects of two different O.T. interventions, functional rehabilitation and reactivation, were evaluated using a randomized controlled trial. 44 patients of a long-term care centre were randomly allocated to one of the two types of intervention. Outcomes were measured using a variety of psychometric tests at 3 different points in time. (Bach et al., 1995).

2. Cohort Design

- A cohort is a group of people (clients) who have been exposed to a similar situation, for example a program, or a diagnosis/disease. Whatever the topic/issue of interest, the groups of clients is identified and followed/observed over time to see what happens.

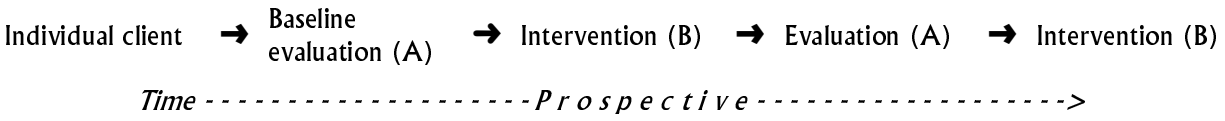
- Cohort designs are “prospective”, meaning that the direction of time is always forwards. Time flows forwards from the point at which the clients are identified. They are sometimes referred to as prospective studies.
- Cohort studies often have a comparison (“control”) group of clients/people who have not been exposed to the situation of interest (eg., they have not received any treatment). One of the main differences between an RCT and a Cohort study is that the allocation of people (clients) to the treatment and control groups is not under the control of the investigator in a Cohort study - the investigator must work with the group of people who have been identified as “exposed” and then find another group of people who are similar in terms of age, gender and other important factors.
- It is difficult to know if the groups are similar in terms of all the important (confounding) factors, and therefore the authors cannot be certain that the treatment (exposure) itself is responsible for the outcomes.
- Advantages of Cohort studies are they are often less expensive and less time-consuming than RCT’s.



Example: Evaluation of a mental stimulation programme used a cohort design to measure changes in mental status in 30 patients over a 2-month time period. The first 15 patients who were admitted to a day care centre received treatment and composed the ‘exposed’ group. The remaining 15 admissions did not receive treatment immediately, and served as a ‘control’ group.(Koh et al., 1994).

3. Single Case Design

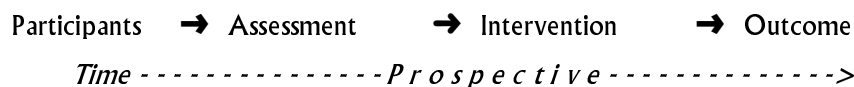
- Single subject/case research involves one client, or a number of clients, followed over time, or evaluated on outcomes of interest.
- There are different types of methods used in single case designs, with different terms used such “n of 1” studies, “before-after trial in the same subject”; or single case “series” involving more than one subject/client.
- The basic feature of any single subject design is the evaluation of clients for the outcome(s) of interest both before (baseline) and after the intervention. This design allows an individual to serve as their own “control”. However, it is difficult to conclude that the treatment alone resulted in any differences as other factors may change over time, for example the disease severity may change.
- It is useful when only a few clients have a particular diagnosis or are involved in a treatment that you want to evaluate. This type of study is easily replicated with more than one client. It’s flexible approach make it particularly appropriate for conducting research in clinical settings.



Example: A study examining the effects of environmental changes during an O.T. intervention on a psychiatric ward used a single case design to observe changes in behaviour in 10 individual patients. Observations of each patient's behaviour were made before, during and after the intervention. (Burton, 1980).

4. Before-After Design

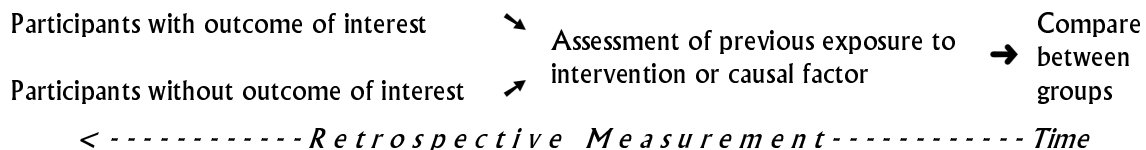
- Before-after design is usually used to evaluate a group of clients involved in a treatment (although as mentioned above, it is a method also used to study single cases/individuals).
- The evaluator collects information about the initial status of a group of clients in terms of the outcomes of interest and then collects information again about the outcomes after treatment is received.
- This is a useful design when you do not wish to withhold treatment from any clients. However, with no "control" group, it is impossible to judge if the treatment alone was responsible for any changes in the outcomes. Changes could be due to other factors such as disease progression, medication use, lifestyle or environmental changes.



Example: The level of caregiver strain following placement of an elderly family member with dementia in adult day care was evaluated using a before-after design. Outcomes of caregiver strain and burden of care were measured in 15 subjects before and after the day care placement. (Graham, 1989).

5. Case Control Design

- Case control studies explore what make a group of individuals different. Other terms used are case-comparison study or retrospective study. Retrospective is the term used to describe how the methods look at an issue after it has happened. The essential feature of a case control study is looking backwards.
- A set of clients/subjects with a defining characteristic or situation, for example a specific diagnosis or involvement in a treatment, are identified. The characteristic or situation of interest is compared with a 'control' group of people who are similar in age, gender and background but who do not have the characteristic or are not involved in the situation of interest. The purpose is to determine differences between these groups.
- It is a relatively inexpensive way to explore an issue, but there are many potential problems (flaws) that make it very difficult to conclude what factor(s) are responsible for the outcomes.



Example: If an occupational therapist wanted to understand why some clients of a day care programme attended a daily activity programme (which was optional) on a regular basis, while other clients did not attend, a case control design could be used to explore differences between the two groups of clients in relation to age, gender, interests, background and current living situation.

6. Cross-Sectional Design

- Involves one group of people, and all the evaluation of the whole group is carried out at the same time.
- This design is often used to explore what factors may have influenced a particular outcome in a group of people. It is useful when relatively little is known about an issue/outcome.
- Surveys, questionnaires and interviews are common methods used in cross-sectional studies. They are relatively inexpensive and easy, as evaluation takes place at one point in time.
- It is impossible to know if all factors have been included in the evaluation, so it is difficult to draw cause-effect conclusions from the results beyond the group of people being studied.

Participants → Measurement of outcomes and other factors at the same time

▲ Time: All done at one point in time ▲

Example: Clients and their families who have been involved in a new activity programme for seniors with dementia can be surveyed or interviewed upon discharge to evaluate the impact of the programme on their quality of life, activity participation and level of satisfaction.

7. Case Study Design

- A case study is carried out in order to provide descriptive information (data) about the relationship between a particular treatment (exposure) and an outcome of interest. It is also called a descriptive study, as that is the primary purpose. There is no control group.
- It is often used to explore a new topic or treatment, when there is little knowledge. However, the results can only be considered in terms of describing a particular situation. It may generate information to support further study of the topic of interest.

Participants with condition of interest → Information about clinical outcome

Example: Twelve patients on a long-stay geriatric ward were observed over a period of time to determine the effectiveness of providing individual and group activities on the ward. Engagement levels were observed and recorded at 10-minute intervals to determine any differences between no intervention, individual activities and group activities. (McCormack & Whitehead, 1981).

Appropriateness of Study Design

- Some of the important issues to consider in determining if the study design is the most appropriate include:
 - Knowledge of the topic/issue: If little is known about an issue, a more exploratory method is appropriate, for example a case study or a cross-sectional design. As our level of knowledge increases, study designs become more rigorous, where most variables that could influence the outcome are understood and can be controlled by the researcher. The most rigorous design is the RCT.
 - Outcomes: If the outcome under study is easily quantified and has well-developed standardized assessment tools available to measure it, a more rigorous design (eg. An RCT) is appropriate. If outcomes are not fully understood yet, such as quality of life, then a design that explores different factors that may involved in the outcomes is appropriate, such as a case control design.

- Ethical issues: It is appropriate to use a research design that uses control groups of people receiving no treatment if there are no ethical issues surrounding the withholding of treatment.
- Study purpose/question: Some designs are well-suited to studying the effectiveness of treatment, including RCT's, before-after designs, and single-case studies. Other designs (eg. case control and cross sectional) are more appropriate if the purpose of the study is to learn more about an issue, or is a pilot study to determine if further treatment and research is warranted.

Biases

- There are many different types of biases described in the research literature. The most common ones that you should check for are described below under 3 main areas:
 1. Sample (subject selection) biases, which may result in the subjects in the sample being unrepresentative of the population which you are interested in;
 2. Measurement (detection) biases, which include issues related to how the outcome of interest was measured; and
 3. Intervention (performance) biases, which involve how the treatment itself was carried out.
- The reader is directed to the bibliography if more detailed information is needed about biases.
- A bias affects the results of a study in one direction - it either "favours" the treatment group or the control group. It is important to be aware of which direction a bias may influence the results.

1. Sample/Selection Biases

- *a.* Volunteer or referral bias:
 - People who volunteer to participate in a study, or who are referred to a study by someone are often different than non-volunteers/non-referrals.
 - This bias usually, but not always, favours the treatment group, as volunteers tend to be more motivated and concerned about their health.
- *b.* Seasonal bias:
 - If all subjects are recruited and thus are evaluated and receive treatment at one time, the results may be influenced by the timing of the subject selection and intervention. For example, seniors tend to be healthier in the summer than the winter, so the results may be more positive if the study takes place only in the summer.
 - This bias could work in either direction, depending on the time of year.
- *c.* Attention bias:
 - People who are evaluated as part of a study are usually aware of the purpose of the study, and as a result of the attention, give more favourable responses or perform better than people who are unaware of the study's intent. This bias is why some studies use an "attention control" group, where the people in the control group receive the same amount of attention as those people in the treatment group, although it is not the same treatment.

2. Measurement/Detection Biases

- *a.* Number of outcome measures used:
 - If only one outcome measure is used, there can be a bias in the way that the measure itself evaluated the outcome. For example, one ADL measure considers dressing, eating, and toileting but does not include personal hygiene and grooming or meal preparation.

- This bias can influence the results in either direction; eg., it can favour the control group if important elements of the outcome that would have responded to the treatment were missed.
- Bias can also be introduced if there are too many outcome measures for the sample size. This is an issue involving statistics, which usually favours the control group because the large number of statistical calculations reduces the ability to find a significant difference between the treatment and control groups.
- *b.* Lack of “masked” or “independent” evaluation:
 - If the evaluators are aware of which group a subject was allocated to, or which treatment a person received, it is possible for the evaluator to influence the results by giving the person, or group of people, a more or less favourable evaluation. It is usually the treatment group that is favoured. This should be considered when the evaluator is part of the research or treatment team.
- *c.* Recall or memory bias:
 - This can be a problem if outcomes are measured using self-report tools, surveys or interviews that are requiring the person to recall past events. Often a person recalls fond or positive memories more than negative ones, and this can favour the results of the study for those people being questioned about an issue or receiving treatment.

3. *Intervention/Performance Biases*

- *a.* Contamination:
 - This occurs when members of the control group inadvertently receive treatment, thus the difference in outcomes between the two groups may be reduced. This favours the control group.
- *b.* Co-intervention:
 - If clients receive another form of treatment at the same time as the study treatment, this can influence the results in either direction. For example, taking medication while receiving or not treatment could favour the results for people in either group. The reader must consider if the other, or additional, treatment could have a positive or negative influence on the results.
- *c.* Timing of intervention:
 - Different issues related to the timing of intervention can introduce a bias.
 - If treatment is provided over an extended period of time to children, maturation alone could be a factor in improvements seen.
 - If treatment is very short in duration, there may not have been sufficient time for a noticeable effect in the outcomes of interest. This would favour the control group.
- *d.* Site of treatment:
 - Where treatment takes place can influence the results - for example, if a treatment programme is carried out in a person’s home, this may result in a higher level of satisfaction that favours the treatment group. The site of treatment should be consistent among all groups.
- *e.* Different therapists:
 - If different therapists are involved in providing the treatment(s) under study to the different groups of clients, the results could be influenced in one direction - for example, one therapist could be more motivating or positive than another, and hence the group that

she worked with could demonstrate more favourable outcomes. Therapist involvement should be equal and consistent between all treatment groups.

Sample

- N = ? The number of subjects/clients involved in the study should be clear.
- Was the sample described in detail? The description of the sample should be detailed enough for you to have a clear picture of who was involved.
- Important characteristics related to the topic of interest should be reported, in order for you to conclude that the study population is similar to your own and that bias was minimized. Important characteristics include:
 - who makes up the sample - are the subjects appropriate for the study question and described in terms of age, gender, duration of a disability/disease and functional status (if applicable)?;
 - how many subjects were involved, and if there are different groups, were the groups relatively equal in size?;
 - how the sampling was done - was it voluntary, by referral? Were inclusion and exclusion criteria described?
 - if there was more than 1 group, was there similarity between the groups on important (confounding) factors.
- Was the sample size justified? The authors should state how they arrived at the sample size, to justify why the number was chosen. Often, justification is based on the population available for study. Some authors provide statistical justification for the sample size, but this is rare.
- Ethics procedures should be described, although they are often left out. At the very least, authors should report if informed consent was obtained at the beginning of the study.

Outcomes

- Outcomes are the variables or issues of interest to the researcher - they represent the product or results of the treatment or exposure.
- Outcomes need to be clearly described in order for you to determine if they were relevant and useful to your situation. Furthermore, the method (the how) of outcome measurement should be described sufficiently for you to be confident that it was conducted in an objective and unbiased manner.
- Determine the frequency of outcome measurement. It is important to note if outcomes were measured pre- and post-treatment, and whether short-term and/or long-term effects were considered.
- Review the outcome measures to determine how they are relevant to occupational therapy practice, ie. - they include areas of occupational performance, performance components and/or environmental components.
- List the measures used and any important information about them for your future reference. Consider if they are well-known measures, or ones developed by the researchers for the specific study being reported. It may be more difficult to replicate the study in the latter situation.
- The authors should report if the outcome measures used had sound (well-established and tested) psychometric properties - most importantly, reliability and validity. This ensures confidence in the measurement of the outcomes of interest.

- Were the outcome measures reliable? - Reliability refers to whether a measure is giving the same information over different situations. The 2 most common forms of reliability are: test-retest reliability - the same observer gets the same information on two occasions separated by a short time interval; and inter-rater reliability - different observers get the same information at the same time.
- Were the outcome measures valid? - Asks whether the measure is assessing what it is intended to measure. Consider if the measure includes all of the relevant concepts and elements of the outcome (content validity), and if the authors report that the measure has been tested in relationship to other measures to determine any relationship (criterion validity). For example, a "valid" ADL measure will include all relevant elements of self-care, and will have been tested with other measures of daily living activities and self-care functioning to determine that the relationship between the measures is as expected.

Intervention

- Intervention described in detail? - there should be sufficient information about the information for you to be able to replicate it.
- In reviewing the intervention, consider important elements such as:
 - The focus of the intervention - is it relevant to occupational therapy practice and your situation;
 - Who delivered it - was it one person or different people, were they trained?;
 - How often the treatment was received - was it sufficient in your opinion to have an impact? was the frequency the same if there were different groups involved?;
 - The setting - was treatment received at home or in an institution? Was it the same for different groups of subjects if there was more than one treatment group?
- These elements need to be addressed if you want to be able to replicate the treatment in your practice.
- Contamination, Co-intervention avoided? - these two factors were described under Biases (see Design section). Were they addressed? If not, consider what possible issues could influence the results of the study, for example, what could happen if some of the clients in the control group received some treatment inadvertently (contamination) or if some subjects were taking medication during the study (co-intervention)? Make note of any potential influences. If there was only one group under study, mark "not applicable (n/a)" on the form.

Results

- Results were reported in terms of statistical significance? Most authors report the results of quantitative research studies in terms of statistical significance, to prove that they are worthy of attention. It is difficult to determine if change in outcomes or differences between groups of people are important or significant if only averages, means or percentages are reported.
- Refer to the bibliography if you wish to review specific statistical methods.
- Outline the results briefly in this section, focusing on those that were statistically significant. If the results were not significant statistically, examine the reasons: was the sample size not large enough to show an important, or significant, difference; or were too many outcome measures used for the number of subjects involved.
- Were the analysis method(s) appropriate? Do the authors justify/explain their choice of analysis methods? Do they appear to be appropriate for the study and the outcomes. You need to consider the following:

- The purpose of the study - is it comparing 2 or more interventions, or examining the correlation between different variables of interest. Different statistical tests are used for comparison and correlation.
- The outcomes - if there is only one outcome measured to compare 2 different treatments, a simple statistical test such as a t-test will probably be sufficient. However, with a larger number of outcomes, involving different types of variables, more complex statistical methods, such as analysis of variance (ANOVA), are usually required.
- Clinical importance was reported? Numbers are often not enough to determine if the results of a study are important clinically. The authors should discuss the relevance of the results to clinical practice and/or to the lives of the people involved. If significant differences were found between treatment groups, are they meaningful in the clinical world? If differences were not statistically significant, are there any clinically important or meaningful issues that you can consider for your practice?

Drop-outs |

- Drop-outs were reported? - The number of subjects/participants who drop out of a study should be reported, as it can influence the results. Reasons for the drop-outs and how the analysis of the findings were handled with the drop-outs taken into account should be reported, to increase your confidence in the results. If there were no drop-outs, consider that as 'reported' and indicate no drop-outs in the Comments section.

Conclusions and Clinical Implications |

- The discussion section of the article should outline clear conclusions from the results. These should be relevant and appropriate given the study methods and results. For example, the investigators of a well-designed RCT study using sound outcome measures could state that the results are conclusive that treatment A is more effective than treatment B for the study population. Other study designs cannot make such strong conclusions, as they likely had methodological limitations or biases, such as a lack of a control group or unreliable measures, that make it difficult to "prove" or conclude that it was the treatment alone that influenced the outcome(s). In these situations, the authors may only conclude that the results demonstrated a difference in the specific outcomes measured in this study for the clients involved. The results may not be generalizable to other populations, including yours. Further study or research should therefore be recommended.
- The discussion should include how the results may influence clinical practice - do they offer useful and relevant information about a client population, or an outcome of interest? Do they warrant further study? Consider the implications of the results, as a whole or in part, for your particular practice and for occupational therapy in general.

Bibliography

Crombie, I.K. (1996). The pocket guide to critical appraisal: A handbook for health care professionals. London: BMJ Publishing Group.

Dept. Of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre (1981). How to read clinical journals: V: To distinguish useful from useless or even harmful therapy. Canadian Medical Association Journal, 124, 1156-1162.

Law, M. (1987). Measurement in occupational therapy: Scientific criteria for evaluation. Canadian Journal of Occupational Therapy, 58, 171-179.

Law, M., King, G. & Pollock, N. (1994). Single Subject Research Design. Research Report #94-2. Hamilton, ON: Neurodevelopmental Clinical Research Unit.

Mulrow, C.D., Oxman, A.D. (Eds.). (1996). Cochrane Collaboration Handbook. Available in The Cochrane Library [database on disk and CD-ROM]. The Cochrane Collaboration: Issue 2. Oxford: Updated Software; 1998.

Norman, G.R. & Streiner, D.L. (1986). PDQ Statistics. Burlington, ON: B.C. Decker Inc.

Sackett, D. L. (1979). Bias in analytic research. Journal of Chronic Disability, 32, 51-63.

Sackett, D.L., Haynes, R.B., Guyatt, G.H. & Tugwell, P. (1991). Clinical Epidemiology. A Basic Science for Clinical Medicine. Second Edition. Toronto, ON: Little, Brown and Co.

Streiner, D.L., Norman, G.R. & Blum, H.M. (1989). PDQ Epidemiology. Toronto, ON: B.C. Decker Inc.

Articles of activity programmes for seniors with dementia (referred to in examples of study designs):

Bach, D., Bach, M., Bohmer, G., Gruhwalk, T. & Grik, B. (1995). Reactivating occupational therapy: a method to improve cognitive performance in geriatric patients. Age and Aging, 24, 222-226.

Burton, M. (1980). Evaluation and change in a psychogeriatric ward through direct observation and feedback. British Journal of Psychiatry, 137, 566-571.

Graham, R.W. (1989). Adult day care: how families of the dementia patient respond. Journal of Gerontological Nursing, 15(3), 27-31, 40-41.

Koh, K., Ray, R., Lee, J., Nair, T., Ho, T. & Ang, P.C. (1994). Dementia in elderly patients: can the 3R mental stimulation programme improve mental status? Age and Aging, 23, 195-199.

McCormack, D. & Whitehead, A. (1981). The effect of providing recreational activities on the engagement level of long-stay geriatric patients. Age and Aging, 10, 287-291.