Research 101:
Developing Critical Evaluation Skills

BY M. Gail Woodbury

Clinicians are aware that wound-care practice must be evidence-based, but often it is difficult to distinguish the quality of scientific evidence. In a recent article, Ryan et al.[1] provided an excellent outline of a strategy to search the literature for evidence on a particular topic in the management of wound patients. Other sources of evidence include colleagues, wound-care "experts," and the cursory perusal of the journal that arrives monthly. Sometimes, wound-care-company representatives will give clinicians a research article that promotes a particular product or intervention. In all of these situations, clinicians must decide if they believe that the product or intervention has merit. This requires a critical evaluation of the article.

One important aspect of critical appraisal is recognizing the research study design and appreciating its pros and cons. The appropriate study design is the one that provides a valid (true) answer to the research question—i.e., a design that minimizes bias to the greatest extent. The process of critically evaluating an article involves finding the answers to essential methodological and reporting questions. The study design and methodological quality dictate the assignment of literature to levels of evidence and, from these, the grade of recommendations for practice is determined. The overall goal of this paper is to help clinicians recognize good evidence so that they can apply it to clinical practice. The specific purposes are to present information about research study designs; present the link between study design, levels of evidence, and recommendations for practice; and discuss how to critically appraise individual articles about interventions using examples.

A goal in conducting research is avoidance of bias. Readers of research articles look for the perception of bias affecting research methods in choice of study design, selection of subjects, creation of treatment groups, application of experimental and control interventions, measurement of outcome, statistical analysis, interpretation and reporting of results. It should be noted that bias may be conscious or unconscious.

Clinical Study Designs

Numerous types of clinical studies are reported in the health-care literature to answer different types of research questions. There are two main classifications of study designs: observational and experimental, as shown in Figure 1. Observational studies are those in which a researcher documents naturally occurring events; in other words, no intervention is introduced. Experimental studies are those in which the researcher introduces an intervention (a program or therapy) and documents the effect.

Observational and experimental studies that involve one group—i.e., that do not have a comparison—are descriptive or uncontrolled studies. The lack of a comparison group means that conclusions cannot be made about the intervention being responsible for the outcome. Any observed change in the subjects' status could have occurred for other unknown and unmeasured reasons, such as the effect of time or the weather. Only two-group studies with an appropriate comparison group permit discussion of causal relationships. Examples of two-group observational studies are case-control and cohort studies, which are described below. A two-group experimental study is

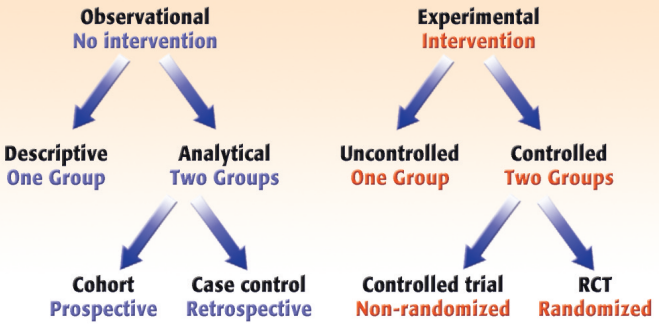
M. Gail Woodbury, BScPT, MSc, PhD, is an Investigator, Program in Rehabilitation and Geriatric Care, Lawson Health Research Institute; and an Adjunct Assistant Professor, Department of Epidemiology and Biostatistics, University of Western Ontario, London, Ontario

referred to as either a controlled trial or a randomized controlled trial (RCT) when subjects are randomly allocated to treatment groups.

Clinical study designs are described in this paper in relation to their potential use to evaluate an intervention, their ability to avoid bias, and the type of investigation for which they are best suited.

FIGURE 1

Types of Clinical Studies



A case study describes a single case and a case series describes several cases used to present, in detail, subject characteristics, an intervention and/or an outcome (something novel). They may be descriptive observational studies, but often they are uncontrolled experimental studies, published to describe an intervention. They are often done as an introductory step to establish an intervention. Intervention effectiveness cannot be determined using a case study or a case series.

Cross-sectional studies may be observational or experimental. They provide a snapshot of a sample where the intervention and outcome are determined at one point in time. Also, they are used to describe subject characteristics, an intervention, and/or an outcome—e.g., a prevalence study. Intervention effectiveness cannot be determined using a cross-sectional study.

Case-control studies are those in which two groups are identified: one group of subjects with an outcome of interest (cases), and an appropriate comparison group of subjects without the outcome of interest (controls). Both groups are reviewed retrospectively to determine the relationship of the outcome to intervention. These studies are considered to be retrospective because subjects are identified on the basis of an outcome. Then, preceding patient and care characteristics are accumulated, usually from records, to explain the outcome. One limitation of this type of study is the inadequacy of patient records, as the information contained therein has been obtained for a purpose unrelated to the research. Intervention effectiveness can be evaluated using a case-control study because there is a comparison group, but this is not the most rigorous study design for this purpose.

In a cohort study, a sample of subjects with a similar disorder is followed prospectively—e.g., patients with heel ulcers, who may be treated to differing extents or not treated with an intervention of interest. Outcomes are measured and compared in groups of treated and not-treated subjects. These studies are useful for identifying uncommon or adverse effects of treatments, or for assessing different approaches or changes in service delivery—e.g., an incidence study. They are useful to evaluate the outcome of treatment when an RCT is not possible. One limitation of this type of study is the lack of equivalence between the naturally occurring treatment groups, which might result in one group being favoured over the other in relation to the outcome. In other words, intervention effectiveness can be evaluated using a cohort study, because there is a comparison group, but this is not the most rigorous study design.

The RCT is considered the gold-standard study design for determining the value of an intervention by comparing it with a placebo or another therapy. A study population is identified; subjects are randomly allocated to intervention or control groups; appropriate blinding is used; the outcome of interest is measured in both groups and compared. Three of its features—

FIGURE 2

Randomized Controlled Trial (RCT)

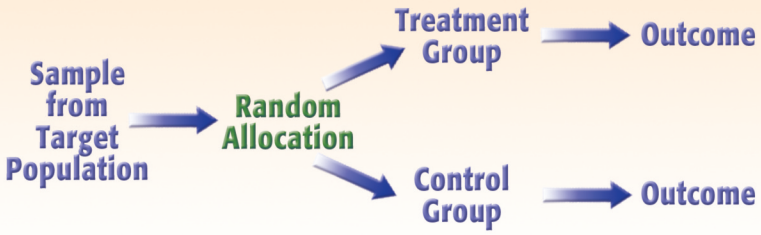
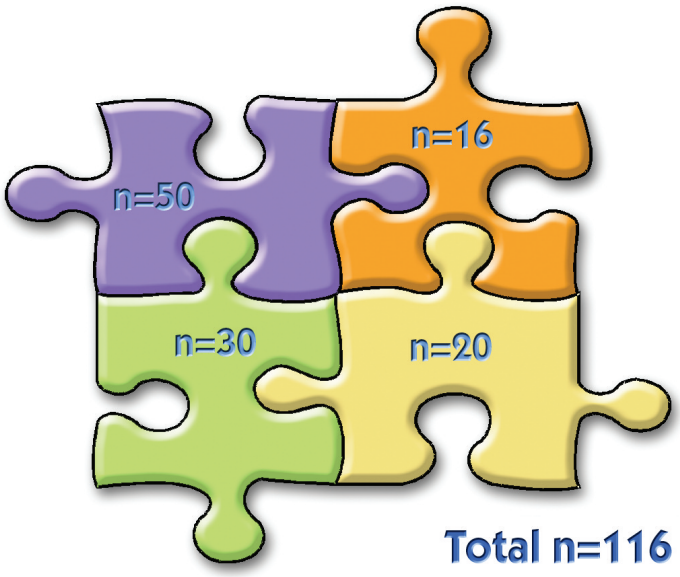


FIGURE 3

RCTs Collated into a Systematic Review Provide More Information

Each puzzle piece represents an individual RCT.



randomization, blinding, and tracking of all participants to the end of the study—help to make it the design that is most effective in minimizing bias. Prior to conducting the study, the required sample size that will detect a meaningful clinical difference and show statistical significance is determined. Random allocation to treatment groups is done to achieve a similar distribution of measured and unmeasured baseline characteristics in both groups.

Some people mistake systematic allocation (e.g., placing patients alternately into groups) as random. There are, in fact, specific requirements that must be met to refer to a study as random—i.e., randomization generated by computer, with the resulting individual subject allocation placed in opaque, sealed envelopes to ensure that no one on the clinical or research team could interfere in any way with the placement of individual subjects into groups. Blinding is used in RCTs to reduce bias; double-blind means that both patient and researcher are blind to (unaware of) treatment grouping. If it is not feasible to achieve double blinding, the next best alternative is blind assessment—i.e., the person who assesses outcome is unaware of the treatment allocation.

Systematic reviews of RCTs are considered to be even more important than individual or multiple RCTs because they provide a meticulously determined summary of information about the value of an intervention. They are distinguished from narrative reviews because they involve an explicit, formal method for locating, evaluating, selecting, assembling, synthesizing, analyzing and interpreting a body of research. If the studies are similar, the results can be combined statistically. This is called a meta-analysis. It may be useful to think of a systematic review as a puzzle in which each individual study is a puzzle piece. The complete review provides a clearer picture because the increased sample size provides more statistical power to detect a difference between groups (see Figure 3). Systematic reviews are powerful tools, but in many health-care areas, including wound care, there are few RCTs.

Levels of Evidence

Levels of evidence based on study design and methodological quality summarize the overall strength of the body of literature on a specific topic. The level of evidence determines the strength of recommendation for a clinical practice guideline or best practice statement as shown in Table 1.

TABLE 1

Relationship Between Levels of Evidence and Grades of Recommendation

Level of Evidence	Grade of Recommendation
Level 1: Large randomized trials with clear-cut results	A
Level 2: Small randomized trials with uncertain results	B
Level 3: Non-randomized trial, with controls	C
Level 4: Case series, no controls	C
Level 5: Expert opinion without critical appraisal	D

Adapted from Sackett[2]

Critical Appraisal of Individual Intervention or Therapy Studies

Suggested questions called "users' guides" have been developed to help clinicians critically appraise different types of articles. These have been published in *The Journal of the American Medical Association (JAMA).* Users' guides to articles about therapy [3,4] are the focus of this paper (as opposed to users' guides that address other issues such as prognosis, harm, diagnostic tests). The guides and discussion of the issues that they address are available at www.cche.net/usersguides/therapy.asp.

There are three general questions that are applied to all types of studies:

1. Are the results of the study valid?
2. What are the results?
3. Will the results help me in caring for my patients?

When appraising a study in which an intervention is evaluated, consider the questions listed in Table 2. The first six questions address the extent to which bias might have crept into the study, consciously or unconsciously, or if the results of the study reflect truth. The next two questions deal with the results and their variability. The final three questions guide us to reflect upon application of the results to our own practices.

A common reason to read research articles is to be aware of recent evidence to support a clinical intervention decision, such as the appropriate type of dressing to use. Two examples of recent evidence published in *Ostomy/Wound Management* in 2003 are presented. You may find it helpful to have a copy of these articles on hand while reading the following appraisals. The first, "A study to compare a new self-adherent soft silicone dressing with a self-adherent polymer dressing in Stage II pressure ulcers," described in its abstract as a randomized, multicentre,

controlled study,[5] illustrates the critical appraisal process using the users' guides. The purpose of the study was "to compare the new self-adherent soft silicone dressing to a commonly used hydropolymer dressing in the treatment of Stage II pressure ulcers."[5:p.45] The first thing to notice is that the purpose does not state the outcomes that are to be compared. The purpose states "compare the effects." It is important for

TABLE 2

Critical Appraisal Questions to Consider

Users' Guides for an Article about Therapy	Responses based on the article: *A study to compare a new self-adherent soft silicone dressing with a self-adherent polymer dressing in Stage II pressure ulcers*[5]
A. Are the results of the study valid?	
1. Was assignment of patients to treatments randomized?	Yes, "Participants were randomly assigned to one of the two treatment options by a predetermined, computer-generated, randomization list stratified by study, and the block size was unknown to the investigators. Each centre received numbered, sealed envelopes to be opened in consecutive order."[5: p.45] This is a very good description of appropriate random allocation.
2. Were all patients who entered the trial properly accounted for? Was follow-up complete?	Yes. Soft silicone group started with n=18; one died after baseline assessment. Hydropolymer group started with n=20; one died of broncho-pneumonia. Adverse devise effects and serious adverse events were reported for both groups. Aside from the two patients who died, there is no indication that any subjects were lost during follow-up.
3. Were patients analyzed in the groups to which they were randomized?	It appears that all subjects were analyzed in the groups to which they were randomized.
4. Were patients, health workers, and study personnel "blind" to treatment?	No, it is stated, "the study was not blinded because dressing differences make blinding difficult to achieve."[5:p.45] This is true and one of the potential sources of bias in this type of study.
5. Were the groups similar at the start of the trial?	Yes in Table 1[5:p.48] we are looking for differences that look important. Mean duration of ulcer, 13 weeks (maximum 52), was longer in the hydropolymer group than 8.3 weeks (maximum 24) in the soft silicone group. Since one of the outcomes measures (wound size) is related to healing, the longer ulcer duration in one group could provide a bias favouring the soft silicone group. However, in Table 25 (p.48) more subjects in hydropolymer group have granulation at baseline (19/20 versus 13/18). This observation favours the hydropolymer group. Is it possible these two observations balance each other?
6. Aside from the experimental intervention, were the groups treated equally?	Yes, we would assume so. There are not many details about the rest of the care except for pressure relief. "All patients except one had pressure relief for more than two hours per day."[5:p.46]
B. What are the results?	
1. How large was the treatment effect?	Tissue damage occurred significantly more often in the hydropolymer group. Only descriptive statistics (mean wound size, numbers of subjects) for the various wound characteristics were provided, so the treatment effects are not known. This means that you have to look at the numbers and somehow decide if they are different in the two groups.
2. How precise was the estimate of the treatment effect?	There was no discussion of precision (no 95% confidence intervals provided).
C. Will the results help me in caring for my patients?	
1. Can the results be applied to my patients?	You would need to decide if the sample was representative of your patients.
2. Were all clinically important outcomes considered?	Cost is an outcome that could have been investigated. Most of the other important outcomes were considered: wound size was obtained by tracing, and other wound variables were either present or absent. Although it was stated that "all wound assessments were made by the same health-care professional (a physician or nurse) throughout the study,"[5:p.45] the subjective nature of the wound outcome assessments is a potential source of bias.
3. Are the likely treatment benefits worth the potential harms and costs?	This is difficult to answer because the only apparent difference was in terms of tissue damage with the hydropolymer dressing. The cost of the dressings was not addressed.

you, the appraiser, to know what you are looking for when the outcome measures and results are described in the paper. For critical appraisal of this article, see Table 2 on page 36.

This study would have been improved (potential bias reduced) by

- incorporating an assessor who was blind to treatment allocation (often the logistics of doing this are prohibitive)
- the use of less subjective measurement of the outcomes
- sample size determination at the beginning to ensure sufficient statistical power
- statistical analysis of the outcomes

Nevertheless, this study has an appropriate study design—RCT—to evaluate an intervention. The critical appraisal shows that most of the methodological issues were handled well. Therefore, we should have some confidence in the conclusions reached by the authors. This author's rating: level 2 evidence, grade B recommendation.

The second example of recent evidence, "An evaluation of an adhesive hydrocellular foam dressing and a self-adherent soft silicone foam dressing in a nursing home setting"[6] is not an RCT. The study was described as a retrospective descriptive study.[6] Therefore, you will recognize that this is not the best design for evaluating these interventions. Subjects were not randomly allocated to achieve group equivalence. The study design is difficult to classify. Because it was described as retrospective, one expects it to be a case-control study, but comparison groups were determined based on intervention not on outcome. In fact, it is a cohort study in which retrospective review of charts was done to determine comparison groups based on dressing application and outcomes. Since data were retrieved from May 27, 1997, until June 18, 2002, and since the soft silicone dressing became available only after June 2001, the hydrocellular foam dressing group was much larger—n=1,643—versus the soft silicone dressing—n=162. In general it is preferable to have groups that are similar in size. In addition, when data for comparison groups are collected over the same time period, potential biases, such as changes in other aspects of wound care over time, may be avoided. In the study, 86 patients were treated with both dressings, which provided a situation in which patient variables were controlled and the dressings could have been compared. However, no comparisons within this group were presented.

In wound-care research it is important to note if patient or wound outcomes are reported. In this article, there were 1,891 patients but 4,200 wounds. The results were presented in terms of wounds. The best way to approach the problem of multiple wounds per patient is to select one wound per patient for reporting. It is obvious that factors that affect one wound of a patient will affect multiple wounds; therefore, the use of multiple wounds per patient introduces a bias and is against the rules for applying statistical tests. This is a fatal methodological flaw in the opinion of this author.

Critical appraisal of this article yields the responses indicated in Table 3 (see page 38).

The critical appraisal indicates that the results of the study are not valid, if we consider the article as a comparison between groups. Therefore, there is no reason

TABLE 3

Critical Appraisal Conclusions

Users' Guides for an Article about Therapy	Responses based on the article: *An evaluation of an adhesive hydrocellular foam dressing and a self-adherent soft silicone foam dressing in a nursing home setting.*[6]
A . Are the results of the study valid?	
1. Was assignment of patients to treatments randomized?	No
2. Were all patients who entered the trial properly accounted for? Was follow-up complete?	No (incomplete data: note differences in wound numbers within tables)
3 . Were patients analyzed in the groups to which they were randomized?	Wounds, not patients, were analyzed.
4. Were patients, health workers, and study personnel "blind" to treatment?	No
5. Were the groups similar at the start of the trial?	No
6. Aside from the experimental intervention, were the groups treated equally?	No
B. What are the results?	Not applicable
C. Will the results help me in caring for my patients?	Not applicable

to consider the size of the results or if they will help in caring for patients.

Potential conflict of interest must always be considered when research has been sponsored by a company, as the research has the potential of being biased.

This study has numerous issues that limit confidence in the results and conclusions. This author's rating: level 4 evidence, grade C recommendation.

Conclusions

Evidence-based practice results from consideration of the combination of

- ✔ good evidence
- ✔ clinical judgement
- ✔ patient values

Good evidence from studies that evaluate therapy is

- derived from clinical studies with appropriate study designs
- with clearly described methods
- with bias minimized or avoided, controlled, or acknowledged in study limitations
- with clear avoidance of conflict of interest

When good evidence is available, it should guide clinical practice. When only lower levels of evidence are available, clinicians need to be aware of this and adjust their clinical decision-making accordingly.

With practice, clinicians can improve their critical appraisal skills and ability to recognize good evidence. Setting up a study club with like-minded clinicians to discuss articles of common interest is one approach to making learning more fun and easier.

Editor's note: The levels of evidence and grades of recommendations in this article have been simplified for clarity. More detailed information that the author has used for this article can be found at www.eboncall.org/content/levels.html.

References

1. Ryan S, Perrier L, Sibbald RG. Searching for evidence-based medicine in wound care: An introduction. *Ostomy/Wound Management.* 2003;49(11):67-75.
2. Sackett DL, Rosenberg WS, Gray JAM, Haynes RB, Richardson. Evidence-based medicine: What it is and what it isn't. *BMJ.* 1996;312:71-2.
3. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA.* 1993 Dec 1;270(21):2598-601.
4. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. *JAMA.* 1994 Jan 5; 271(1):59-63.
5. Meaume S, Van De Looverbosch D, Heyman H, Romanelli M, Ciangherotti A, Charpin S. A study to compare a new self-adherent soft silicone dressing with a self-adherent polymer dressing in Stage II pressure ulcers. *Ostomy/Wound Management.* 2003;49(9):44-51.
6. Viamontes L, Temple D, Wytall D, Walker A. An evaluation of an adhesive hydrocellular foam dressing and a self-adherent soft silicone foam dressing in a nursing home setting. *Ostomy/Wound Management.* 2003;49(8):46-58.

ACTISORB Silver 220

A unique wound care dressing indicated to manage the infection and odour associated with chronic wounds.

Learn more about the benefits of **ACTISORB Silver 220.** Contact your Johnson & Johnson Wound Management Product Specialist at 1-800-668-9045.

ACTISORB Silver 220

ACTIVATED CHARCOAL CLOTH WITH SILVER

Endotoxin/Bacterial Balance

Johnson & Johnson Wound Management

Johnson & Johnson Medical Products
200 Whitehall Drive, Markham (ON) L3R 0T5
Tel: (English) 1 800 668-9045 Tél. : (Français) 1 800 668-9067
Website : www.jnjgateway.com