

Designer's Corner

Methodological Issues in Outcomes Research

Souraya Sidani

Introduction

With the increasing societal demand for professional and financial accountability, nurses and other health-care professionals are challenged to demonstrate that the care they provide "makes a difference" – that their interventions achieve positive client outcomes effectively and efficiently. Effective interventions produce the desired responses. In nursing, desired responses refers to measurable changes in the client's health status, condition, or behaviour that indicate the resolution of a presenting problem or diagnosis or the prevention of a condition (Hegyvary, 1993; Lang & Marek, 1990).

Demonstrating that nursing interventions are effective rests on the ability to detect the expected changes. The ability to detect the changes depends, in turn, on selecting outcomes that are attributable to the antecedent care, sensitive to nursing care, and congruent with the unit of analysis, and on assessing the outcomes at the appropriate time (Bond & Thomas, 1991; Griffiths, 1995; Hegyvary, 1991; Jones, 1993; Stewart & Archbold, 1992). Even if the right outcomes are selected and assessed at the right time, detecting the expected changes requires that two methodological issues, selection of outcome measures and implementation of the intervention, be appropriately addressed; otherwise there is the likelihood that "real" intervention effects, when present, will not be detected and, subsequently, that the validity of conclusions regarding effectiveness of the intervention in achieving the desired outcomes will be threatened (Lipsey, 1990; Scheirer & Rezmovic, 1983).

Souraya Sidani, Ph.D., R.N., is an Assistant Professor in the Faculty of Nursing at the University of Toronto.

Selection of Outcome Measures

If they are to detect real intervention effects, outcome measures must be maximally responsive to any changes brought about by the intervention and minimally responsive to anything else (Lipsey, 1990). For the measures to respond to the treatment effects, they must be valid, reliable, and sensitive to change.

To be maximally responsive to the intervention effects, outcome measures must be construct valid. They should represent, accurately, the particular outcome concept or the characteristic that the intervention is expected to affect or change. A measure with established validity represents all the domains of the concept being examined and does not capture variability in the participants' responses associated with the method of data collection (i.e., has minimal method bias) and/or with the influence of distinct outcome concepts that could be either related or unrelated to the concept of interest (Cook & Campbell, 1979). Using instruments with no established construct validity may lead to difficulty in interpreting the findings. Non-statistically significant findings could indicate either that the intervention is not effective in producing the desired outcome or that the measure did not validly capture the intended outcome. Significant findings are questionable, since it is difficult to know exactly what was measured by the instrument – the intended outcome or a different but related concept. Both situations have the potential for incorrect conclusions. For example, symptom retrenchment – defined as a reduction in the frequency, intensity, duration, and intrusiveness of symptoms – is the expected outcome of a psycho-educational intervention directed at instructing clients in management of symptoms associated with their presenting illness and its treatment. It is inappropriate to measure symptom retrenchment with a symptom-transition scale that incorporates items representing both positive and negative changes in symptoms, especially if a total-scale score is used to reflect the expected outcome. The total-scale score captures the negative changes, representing symptom extension, which is another intended outcome and is therefore invalid in reflecting the specific changes expected of this intervention.

Instruments must be reliable in order to be valid and capable of detecting the intended intervention effects. Reliable instruments measure the outcome variable consistently and with minimal error. Error represents fluctuations in the measure scores that are unrelated to the characteristic being measured. These fluctuations are either random – related to chance factors such as the clarity of instructions or the subjects' motivation or fatigue at the time of measurement – or systematic –

related to factors such as the subjects' comprehension, acquiescence, or social desirability (Waltz, Strickland, & Lenz, 1991).

Error of measurement, whether random or systematic, increases variability in the distribution of scores for subjects in the experimental groups, leading to increased within-group variance. Increased within-group variance, in turn, reduces the statistical power to detect significant intervention effects, increasing the potential for erroneous conclusions regarding effectiveness of the intervention in achieving the expected outcomes (Lipsey, 1990; Stuccliffe, 1980).

Valid and reliable measures are sensitive to change. The term "sensitivity to change" refers to two properties of outcome measures: ability to detect differences in the outcome between individuals who received the intervention and those who did not (i.e., inter-individual differences), and ability to detect change in the outcome, within the same individual, over time (i.e., intra-individual differences) (Guyatt, Kirshner, & Jaeschke, 1992; Stewart & Archbold, 1993). It is generally believed that measures sensitive to inter-individual differences are not necessarily responsive to intra-individual change. These measures usually focus on measuring true or stable between-subject differences (Carver, 1974). Thus differences in the outcome observed after implementation of the intervention could be due to the true between-subject differences, rather than to the effects of intervention. True between-subject differences – related to personal characteristics, for example – contribute to the undesirable within-group variance; this variance is not accounted for by the intervention, and is considered error variance, which can obscure the treatment effects (Lipsey, 1990). Furthermore, outcome measures that reflect stable inter-individual differences are less likely to capture clinically important changes expected as a result of the intervention (Stewart & Archbold, 1992). Therefore, outcome measures need not discriminate among participants with different levels on the outcome being measured, but they must be able to capture the intra-individual changes in the outcome resulting from the intervention. Such responsive instruments draw upon salient aspects of the outcome that are likely to undergo changes. They inquire about the extent of change in the outcome, using a Likert-type scale ranging from "no change" to "problem resolved," for example. They are characterized by their ability to detect minimal score variability in the outcome measured before and after treatment, and to show observable difference in the scores between the two occasions of measurement. Low variability is desirable, since it reflects the participants' homogeneity with respect to the characteristic being measured at pre-test (i.e., initial equivalence) and with respect to their response to the treatment (such as whether

they responded in a uniform manner to the intervention) (Nicewander & Price, 1983). Measures with a fine-grained scaling method, representing the range of responses over a continuum, are sensitive to changes on the scores over time. Coarse-grained measures may lead to floor or ceiling effects, thus limiting upward or downward changes in the responses over time (Kirshner & Guyatt, 1985; Lipsey). Examples of such instruments are criterion-referenced measures, such as those designed to detect the participants' acquisition of health-related knowledge or behaviour as a result of a psycho-educational intervention. For a discussion of methods for developing instruments sensitive to intra-individual change and testing their responsiveness, see Deyo and Centor (1986), Deyo and Inui (1984), Guyatt, Deyo, Charlson, Levine, & Mitchell (1989), Guyatt et al. (1992), and Kirshner and Guyatt).

Hegyvary (1993) expresses some concern regarding outcomes measurement when the outcome variable, such as symptoms, is both the indicator for the treatment and the outcome of treatment. In fact, any nursing intervention is a response to a clinical problem, whether actual or potential, with the goal of resolving or preventing it. For instance, diet therapy involving reduced sodium intake is prescribed to lower blood pressure in cardiac patients. Thus hypertension is the indicator for diet therapy, while an improved blood-pressure level is the expected outcome. It is true that, conceptually, the same variable – blood pressure – is considered the indicator for treatment and the outcome of treatment; however, what is actually the indicator is the *level* of the variable, and the expected outcome is the *change* in its level. Researchers deal with this issue by using the indicator as the pre-test measure and the outcome as the post-test measure. Alternatively, goal-attainment – lowered blood pressure – could be used as the outcome for nursing care (Lang & Marek, 1990). Methods for developing goal-attainment scaling techniques are offered by Inzer and Aspinall (1981) and Martin and Scheet (1992). These techniques include assessing the client's progress toward resolution of the presenting problem (i.e., the indicator for treatment). A Likert-type scale, ranging from "no change" to "problem resolved," is used to determine where the client stands. These scales tend to be sensitive to clinically significant intra-individual differences.

Implementation of the Intervention

The method used to implement an intervention influences the ability to detect significant intervention effects. In particular, selecting the control group and maintaining integrity in implementation have a direct impact on the extent of the observed effect.

It is essential that a control group be included in a study evaluating the impact of an intervention on selected outcomes, in order to rule out threats to the internal validity. A control group is needed to enhance the attribution of the observed effects to the *intervention*, and not to *random* or *systematic irrelevancies* correlated with factors other than the treatment (Cook & Campbell, 1979). The contrast, on the outcome variables, between the means of control group and experimental group indicates the effectiveness of the intervention. Selection of the control group influences whether this contrast reflects a significant difference between the experimental groups in the outcome measured at post-test. If the control condition chosen does not differ greatly from the treatment condition, the contrast between the means of the group is small. That is, if treatment-as-usual, which may incorporate some components of the treatment being evaluated, is selected as the control condition, then the contrast is reduced. Similarly, placebo conditions are often similar to the treatment conditions for which they serve as controls, potentially achieving outcomes similar to those expected for the intervention being evaluated; thus the contrast is decreased, manifested by non-significant differences among the groups that might be mistakenly taken for treatment ineffectiveness (Lipsey, 1990). For example, if the intervention being evaluated consists of a comprehensive discharge plan addressing the social needs of elderly patients, in addition to their physical needs, and the control condition is the usual discharge plan limited to meeting their physical needs, the contrast between the two groups may not be statistically significant, because of the overlapping physical component. While there is no well-determined strategy for addressing this issue, the rule is to select controls who would maximize the contrast between the experimental groups.

For the intervention to produce the expected effects, it should be implemented as designed and in a consistent manner across the participants. A treatment protocol developed to guide the intervention delivery specifies the purpose of the intervention; the equipment required; and the activities to be undertaken or the procedures to be performed, as well as their frequency and duration. Professionals are required to follow the protocol faithfully when delivering the intervention, in order to implement it consistently and as designed. While it is possible to maintain the integrity of treatment implementation under controlled laboratory conditions, this is difficult under less controlled field conditions. As a result, the intervention may be delivered inconsistently, thus reducing the power to detect any significant effect (Kirchhoff & Dille, 1994; Scheirer & Rezmovic, 1983; Sechrest, Ametrano, & Ametrano, 1983).

An intervention that is not well defined, specifically described, clearly circumscribed, and carefully operationalized is difficult to deliver, especially if several interveners are responsible for implementing it at different times and in different clinical settings. Interveners who do not have a clear understanding of what activities to perform, when, for how long, and with whom will provide the intervention as they perceive it, and not necessarily as prescribed. This variability in implementation will necessarily result in difficulty determining the exact type, duration, and frequency of care provided, thus making it even more difficult to anticipate the consequences of this intervention and when they can be expected to occur. The end result of this lack of integrity in implementation is an observed variability in outcomes, which obscures the effects of the intervention (Sechrest et al., 1983). For instance, the intervention "provide psychological support" may have different meanings: listen to the client, hold the client's hand, encourage the client to ventilate feelings, or give positive feedback. Consider what the results would be if the intervener provided positive feedback instead of encouraging ventilation of feelings, as designed. The intended outcome, let's say of reduced emotional distress, may not be achieved. The problem of inconsistent implementation becomes more acute and complicated when the client is responsible for or actively involved in the process of care delivery, without the researcher's or the clinician's supervision. For instance, it may be difficult to know what type of exercises a cardiac patient performs at home, how frequently, and for how long, as compared to the prescribed exercise program. Therefore, it becomes important to develop mechanisms for monitoring implementation and collecting relevant data – for instance, the researcher or clinician can keep track of how many educational sessions each client attended. Data representing the intervention-as-delivered could then be used in outcomes analysis, as they are directly linked to, and account for, the variability in the observed outcomes.

Conclusions

Demonstrating the effectiveness of nursing interventions in producing the desired outcome is essential for examining the contribution of nursing care to clinical outcomes, and for maintaining the professional and financial accountability of nursing. The ability to demonstrate the effectiveness of the intervention depends on selecting measures that are maximally responsive to the intervention, and not to other true individual differences; on selecting the appropriate control group, maximizing the contrast between the treatment groups; and on maintaining the integrity of the intervention during implementation. Disregarding

the potential influence of these factors on the intervention outcomes threatens the validity of the conclusions.

Strategies for dealing with these methodological issues have been suggested. These must be further examined in order to determine their feasibility and appropriateness in various situations. Alternative strategies are also needed. Researchers are strongly encouraged to develop mechanisms to deal with these methodological issues appropriately, and a priori, in order to enhance the validity of their findings.

References

- Bond, S., & Thomas, L.H. (1991). Issues in measuring outcomes of nursing. *Journal of Advanced Nursing*, 16, 1492-1502.
- Carver, R.P. (1974). Two dimensions of tests: Psychometric and edumatic. *American Psychologist*, 29, 512-518.
- Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Deyo, R.A., & Centor, R.M. (1986). Assessing the responsiveness of functional scales to clinical change: An analogy to diagnostic test performance. *Journal of Chronic Disease*, 39(11), 897-906.
- Deyo, R.A., & Inui, T.S. (1984). Toward clinical applications of health status measures: Sensitivity of scales to clinically important changes. *Health Services Research*, 19(3), 275-289.
- Griffiths, P. (1995). Progress in measuring nursing outcomes. *Journal of Advanced Nursing*, 21, 1092-1100.
- Guyatt, G.H., Deyo, R.A., Charlson, M., Levine, M.N., & Mitchell, A. (1989). Responsiveness and validity in health status measurement: A clarification. *Journal of Clinical Epidemiology*, 42(5), 403-408.
- Guyatt, G.H., Kirshner, B., & Jaeschke, R. (1992). Measuring health status: What are the necessary measurement properties? *Journal of Clinical Epidemiology*, 45(12), 1341-1345.
- Hegarty, S.T. (1991). Issues in outcome research. *Journal of Nursing Quality Assurance*, 5(2), 1-6.
- Hegarty, S.T. (1993). Patient care outcomes related to management of symptoms. *Annual Review of Nursing Research*, 11, 145-168.
- Inzer, F., & Aspinall, M.J. (1981). Evaluating patient outcomes. *Nursing Outlook*, 29, 178-181.
- Jones, K.R. (1993). Outcome analysis: Methods and issues. *Nursing Economics*, 11(3), 145-151.
- Kirchhoff, K.T., & Dille, C.A. (1994). Issues in intervention research: Maintaining integrity. *Applied Nursing Research*, 7(1), 32-46.
- Kirshner, B., & Guyatt, G. (1985). A methodological framework for assessing health indices. *Journal of Chronic Diseases*, 38(1), 27-36.

- Lang, N.M., & Marek, K.D. (1990). The classification of patient outcomes. *Journal of Professional Nursing*, 6(3), 158-163.
- Lipsey, M.W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.
- Martin, K.S., & Scheet, N.J. (1992). *The Omaha System: Applications for community health nursing*. Philadelphia: W.B. Saunders.
- Nicewander, W.A., & Price, J.M. (1983). Reliability of measurement and the power of statistical tests: Some new results. *Psychological Bulletin*, 94(3), 524-533.
- Scheirer, M.A., & Rezmovic, E.L. (1983). Measuring the degree of program implementation: A methodological review. *Evaluation Review*, 7(5), 599-633.
- Sechrest, L., Ametrano, D., & Ametrano, I.M. (1983). Evaluations of social programs. In C.E. Walker (Ed.), *The handbook of clinical psychology*. Homewood, IL: Dow Jones-Irwin.
- Stewart, B.J., & Archbold, P.G. (1992). Nursing intervention studies require outcome measures that are sensitive to change. Part 1. *Research in Nursing and Health*, 15, 477-481.
- Stewart, B.J., & Archbold, P.G. (1993). Nursing intervention studies require outcome measures that are sensitive to change. Part 2. *Research in Nursing and Health*, 16, 77-81.
- Stucliffe, J.P. (1980). On the relationship of reliability to statistical power. *Psychological Bulletin*, 88(2), 509-515.
- Waltz, C.F., Strickland, O.L., & Lenz, E.R. (1991). *Measurement in nursing research* (2nd ed). Philadelphia: F.A. Davis.

Date accepted: September 1996