CJNR 2009, Vol. 41 N° 4, 32–50

Résumé

L'identification de biais dans les recherches sur les interventions infirmières complexes : une liste de vérification

Marsha Campbell-Yeo, Manon Ranger, Celeste Johnston et Dean Fergusson

Un biais est défini comme étant toute erreur systématique faussant une estimation des résultats de recherche. Dans des études portant sur des interventions infirmières complexes, les biais sont particulièrement difficiles à repérer en raison de problématiques liées à l'anonymat et au choix des outils d'évaluation. Les auteurs identifient des stratégies de dépistage de biais dans les recherches sur les interventions. Une analyse documentaire et une consultation auprès d'experts révèlent six volets liés au développement de protocoles de recherche qui offrent des possibilités quant à la réduction de biais : le concept de recherche; la définition de l'intervention; le choix des comparaisons; la randomisation/l'allocation; l'intégrité de l'intervention; et la détermination des résultats. Les auteurs proposent une liste de vérification qui aidera les chercheurs à réduire le risque de biais dans le cadre de la préparation de protocoles d'essais portant sur des interventions infirmières complexes. Le recours à une telle liste peut bonifier la rigueur scientifique et assurer aux cliniciens l'accès à une information fiable.

Mots clés : biais, intervention complexe

CJNR 2009, Vol. 41 Nº 4, 32–50

Controlling Bias in Complex Nursing Intervention Studies: A Checklist

Marsha Campbell-Yeo, Manon Ranger, Celeste Johnston, and Dean Fergusson

Bias is defined as any systematic error resulting in an inaccurate estimate of the outcome of a study. In studies of complex nursing interventions, bias is particularly difficult to control because of issues related to blinding and choice of controls. The authors identify strategies to control bias in intervention studies. They conduct a literature review and consult expert opinion to identify 6 areas of study protocol development that have potential for reducing bias: study concept, definition of intervention, selection of comparisons, randomization/allocation, integrity of intervention, and ascertainment of outcomes. They provide a checklist to help researchers reduce the potential for bias in preparing protocols for complex nursing intervention trials. Use of the checklist can enhance scientific rigour and thus help to ensure that clinicians are ultimately provided with reliable information.

Keywords: bias, complex intervention, nursing research

Study bias can be defined as any design error that results in an over- or under-estimation of the effect of an intervention, thus threatening the validity of the findings (Norman & Streiner, 2000). When properly conducted, the double-blind, randomized, placebo-controlled trial (RCT) has been identified as the gold standard for controlling bias in health research (Rees, Wade, Levy, Colford, & Hilton, 2005). However, this type of trial faces challenges for determining the effectiveness of complex nursing intervention trials, for three reasons. First, in intervention trials it can be difficult to blind investigators and participants to the intervention (Bang, Ni, & Davis, 2004). Second, in these trials, significant issues related to the appropriate choice of a comparison group can arise, given that the use of placebos is often impossible or unethical (Mann, 2007). Third, RCTs are usually limited to a single intervention, such as a drug, and are not designed to address issues that arise with more complex interventions, such as nursing interventions composed of interrelated elements (Hawe, Shiell, & Riley, 2008).

While controlling bias is only one of many aspects that must be considered in conducting research, it merits special attention in the context of complex nursing intervention trials, given the above-mentioned limitations of the classical RCT. It is therefore necessary to identify methods

for eliminating study bias. One such method is the development of carefully planned study protocols.

There are no concise guidelines for designing protocols that minimize the potential for bias in intervention studies. Therefore, the aim of this article is to develop a bias-control checklist to aid nurse researchers and health-care professionals in the planning of study protocols.

A computerized search of CINAHL, Pubmed, PsycINFO, Web of Science, and the Cochrane Collaboration Libraries was conducted for the years 1990 to 2009. Keywords included bias, complex interventions, clinical trials, and nursing research. The search was limited to English-language articles. Articles were excluded if they did not discuss at least one concept/issue that can contribute to the generation of study bias in intervention research. Relevant articles were reviewed. Expert opinion was elicited from clinical intervention researchers from three universities who came together to discuss the issue. Six areas of concern were identified and formed the basis for the checklist. The literature was searched based on keywords related to each topic.

The initial search found 38 articles in CINHAL, 14 in Pubmed, 14 in PsychInfo, 18 in Web of Science, and 18 in the Cochrane Libraries. Numerous articles overlapped among databases and for several articles the primary focus was not controlling bias in an intervention trial. A total of 10 articles were retained. Using this literature and expert input, we identified six primary points to be addressed in order to reduce the potential for bias during protocol development: (1) study concept, (2) definition of the intervention, (3) selection of comparisons, (4) randomization/allocation, (5) integrity of the intervention, and (6) ascertainment of outcomes. Each primary point is summarized in the checklist (Figure 1) and described below.

Study Concept

Examination of the study concept presents the first real opportunity for researchers to identify and control bias. We define study concept as the issues and ideas that need to be considered, weighed, defined, and formalized in developing and justifying a study protocol. These include: determining the study topic, purpose, and hypothesis; determining the need for such a study; justifying its need; and seeking input, feedback, and buy-in from all study stakeholders. While an examination of the study concept will not target a specific form of bias per se, it is a platform from which investigators can both identify and minimize the potential for a multitude of biases.

The study concept should be based on a thorough literature review, at which time constructs related to the topic of interest may be identified as potential sources of bias. For example, if post-operative pain is the study

Figure 1 ***Controlling Bias in Complex Intervention Trials: A Checklist***

Key Points	Strategies	Purpose
Study Concept	☐ Conduct systematic literature review/ meta-analysis ☐ Survey experts and colleagues; administer usual care ☐ Conduct pilot test(s)	• Creates basis for hypothesis • Predicts possible contaminates, co-founders, and co-interventions • Provides link between clinical practice and existing evidence • Establishes feasibility
Definition of Intervention	☐ Clearly define intervention a. *Timing, duration, interval between each exposure* b. *Resources required and setting* c. *Characteristics and education of the provider*	• Ascertains replicability, generalizability, and uniformity of delivery
Selection of Controls	☐ Differentiate between best and usual care a. *Conduct a survey and document usual care* b. *Conduct a pilot study* c. *Apply protocol-driven controls* d. *Consider a range of control arms*	• Allows for isolation of direct versus indirect effects of intervention • Ensures greater consistency and comparability between groups
Randomization and Allocation *(continued)*	☐ Ensure true randomization a. *Use opaque, sequentially numbered envelopes* b. *Use off-site randomization, by phone or computer*	• Creates balance between the groups

Key Points	Strategies	Purpose
Randomization and Allocation	☐ Conceal allocation a. *Permuted variable blocks* ☐ Ensure unpredictability of groups a. *Identify the means of recruitment and the recruiters* b. *Maintain records for all those eligible to participate: those approached, those enrolled, and those refused* ☐ Clearly outline all steps in study protocol manual	• Limits differences in potentially confounding variables at baseline • Decreases subversion bias • Decreases research-staff manipulation of group assignment
Integrity of Intervention	☐ Monitor intervention fidelity a. *Address design, training, delivery, receipt, and enactment of intervention(s)* b. *Use intervention-monitoring tools/ inter- and intra-rater coder reliability* ☐ Identify potential co-interventions a. *Conduct random surveys to monitor standard care* ☐ Monitor for contamination a. *Conduct random checks to determine whether participants have discussed the study with other participants or with nurses* b. *Consider cluster randomization* c. *Compensate for intra- or within-cluster correlation*	• Ensures that the study intervention remains stable across time, places, and persons • Limits participant awareness of the treatment being received by the other group

Analysis	☐ Limit loss to follow-up a. *Identify potential participant burden* ☐ Conduct intention-to-treat analysis	• Limits attrition, maintains sample size, and maintains statistical power • Maintains group allocation, allowing for the balancing of confounding factors across groups
Ascertainment of Outcome	☐ Choose a blind design a. *Blind participants, care providers, and investigators where possible* b. *Blind evaluator(s) of outcomes* c. *Consider videotaped participant responses or direct entry of responses into computerized databanks* ☐ Choose objective outcomes ☐ Provide clear and precise written procedures for data collection ☐ Monitor adherence to study interventions	• Minimizes variations in interpretation of data and outcomes • Controls for Hawthorne effect

topic, pre-operative anxiety may be a related construct. If participants receive pre-operative care for their anxiety, this may affect post-operative pain, and it therefore becomes a source of bias. The identification of this source of bias in advance allows the researcher to incorporate measures to reduce bias during the development of the protocol. In the same way, the researcher can predict possible contaminates, confounders, and co-interventions that have been identified or summarized in previous studies (Blair, 2004).

Definition of the Intervention

The potential for exposure bias can arise when researchers fail to adequately define or fully describe interventions being examined (Campbell et al., 2000). Complex clinical interventions are particularly vulnerable to this type of bias, given their multi-faceted nature (Glasziou, Meats, Heneghan, & Shepperd, 2008). A precise definition of the intervention can ensure uniform delivery of the intervention, thus reducing the chance of exposure bias (Lindsay, 2004). A clear and complete definition of the intervention also allows for easy replication of the study, improves generalizability, and enhances the clinical utility of the findings (Campbell et al., 2000). In order to allow access to the definition of the intervention by granting agencies, those providing the study intervention, and those interested in utilizing the results, a precise definition should be included in the study protocol, included in a study manual for research personnel and staff, and reviewed during dissemination of the results.

In a review of 47 RCTs of nursing interventions published in 2000–01, inadequate definition was identified as the most common source of bias (Lindsay, 2004). Reflecting the significance of this issue, these trials originated in eight countries, focused on 14 different health fields (hospital and community populations), and included nursing, specialty, and high-impact general medical journals (e.g., *Lancet, British Medical Journal*). Similarly, a recent review including 27 nursing journals found that 141 research articles published in 2005 reported suboptimal definition of interventions (Conn, Cooper, Ruppar, & Russell, 2008). While the intervention definition accounted for an average 7.3% of article space, the space given to methodological descriptions accounted for over 20.7%. Moreover, only 38 articles (27.0%) reported sufficient detail about the intervention to allow for replication of the study or for translation of the intervention into practice.

Complete definition of the intervention should include not only details on the nature of the treatment but also information about its delivery (i.e., timing, duration, and interval of each exposure); materials needed (such as patient handouts or devices); the setting; and the characteristics and education of the provider (Glasziou, Meats, Henehan, &

Sheppard, 2008). For instance, in a recent study examining the efficacy of maternal skin-to-skin care during heelstick in very preterm neonates, infant condition, position, duration of the intervention, and maternal interaction were clearly defined and the five phases of blood collection were delineated (Johnston et al., 2008). Additionally, measures were taken to control potential sources of bias such as the technical skills and education of the provider, the setting, and the urgency of blood work.

Selection of Appropriate Comparisons

In the evaluation of interventions, the primary purpose of a comparison group is to distinguish between the direct effects of the intervention and the indirect effects of participation in the study (Paterson & Dieppe, 2005). The lack of an appropriate comparison group increases the potential for bias. For instance, participants' symptoms may improve merely due to the passage of time, regression of an acute flare-up, or altered perception because they have been cared for or have been told that they should feel better. Therefore, the use of a comparison group that differs, ideally, only in that it does not receive the intervention is important in order to control for these confounding variables.

Two issues are important in choosing an appropriate comparison group: knowledge of the evidence-based recommendations and guidelines for treating a condition (best care), and current practice (usual care) in the clinical setting (Mann & Djulbegovic, 2003).

Best Versus Usual Care

In intervention studies, the researcher's primary aim is to determine whether a treatment or intervention improves outcomes. Thus, if the intervention is being compared to usual care, it is important to determine whether usual care is reflective of the most recent findings in the literature. If usual care deviates significantly from best care, or if there is inconsistency in the usual care that is provided, then protocol-driven control treatments can ensure greater consistency and improved comparability between the groups (Silverman & Miller, 2004). If protocolized care is used as a comparison, measures are required to determine and ensure protocol compliance of both participants and care providers, in the same manner as in the intervention arm. A pre-trial observational survey, pilot study, or run-in phase could determine the feasibility of the protocolized group and the acceptability of the proposed intervention to staff and participants at all potential sites.

In cases where there is a lack of sufficient evidence to define best practice, great diversity of care, or significant staff reluctance to support new interventions, protocolized comparison alone may not be sufficient.

In such circumstances, the investigator may choose to consider a three-arm trial: a treatment group, a protocolized group reflective of one accepted form of care, and a comparison or usual-care arm reflective of current practice on the unit. Using this design, the researcher compares the efficacy of the new intervention to two alternatives rather than one. Every attempt should be made to ensure that each group is matched with respect to the experience of the health-care provider and number of interactions, thus ensuring that the intervention is the only difference between them (Silverman & Miller, 2004). Protocolized comparison groups enhance scientific validity because they limit inconsistencies between groups. However, if they do not adequately represent current practice, they are less generalizable and may be of little clinical value. The choice of losing generalizability in order to increase the scientific validity of a study should depend on the research purpose and question.

Randomization and Allocation

Selection bias can occur if comparison groups are not considered equal at baseline, prior to the commencement of the intervention. Bias-reducing strategies such as randomization and allocation concealment are important because their exclusion has been associated with amplified treatment outcomes of 20–45% (Balk et al., 2002; Kunz, Vist, & Oxman, 2007).

Randomization

Randomization is considered an optimal method for ensuring balance between groups because it limits differences in potentially confounding variables at baseline (Kunz et al., 2007), enhances the validity of statistical methods of analysis (Bridgman et al., 2003), and reduces the chance of mal-distribution of key predictors (Blair, 2004). True randomization occurs when participants have an equal chance of being assigned to the intervention or the comparison group, without interference from the investigators. Pseudo-randomization, or systematic assignment, has been mistakenly referred to as true randomization in some clinical research trials (Bridgman et al., 2003). In this instance, group assignment may be dictated by factors such as birth date, day of clinic visit, or room assignment. This type of allocation, which is easily predicted by investigators and participants, can lead to potential tampering with participant assignment.

There are several acceptable methods of randomization. One of the simplest, most straightforward, and least expensive is the use of sequentially numbered, opaque, sealed envelopes (SNOSE). This is a reasonable choice, especially for smaller single-centre trials (Doig & Simpson, 2005),

while pharmacy-controlled randomization and 24-hour central randomization by phone-in or Internet have been particularly useful in larger trials or in trials with more than one centre (Schulz & Grimes, 2002).

Allocation Concealment

The most important aspect of randomization is the unpredictability of group allocation, which is referred to as allocation concealment. Inadequate concealment has been associated with an increase of up to 40% in effect sizes (Juni, Altman, & Egger, 2001). Unlike blinding, which controls bias during the course of the study, allocation concealment prevents selection bias and preserves allocation sequence before and until group assignment. Therefore, allocation concealment must be a priority in all studies where participants are randomized (Forder, Gebski, & Keech, 2005).

For the majority of randomization methods, large sample sizes are needed to ensure groups of equal size and of evenly distributed participants. However, in the case of trials with smaller sample sizes, such as many nursing intervention trials, the use of block randomization is helpful. Blocking is used to ensure that, at specific points of enrolment, equal numbers of participants have been assigned to receive either treatment. In unblinded studies it is vital that more than one block size be used, to prevent the anticipation of allocation sequence by the investigators (Schulz & Grimes, 2002). Permuted block randomization is a variation that alters the allocation sequence of specific blocks sizes. For example, blocks of four might consist of AABB, ABAB, ABBA, BABA, and so forth.

Subversion bias, a type of selection bias, occurs when research staff manipulate recruitment to enable the enrolment of specific participants in either the comparison or the experimental group. To avoid this, the randomization sequence should be prepared and conducted by an independent person preferably not linked with the field of study. Thus, telephone or Web-based sequence generation is an excellent choice, especially for multi-centre studies.

In keeping with the CONSORT guidelines (http://www.consort-statement.org), investigators are obliged to give details of all aspects of randomization and allocation concealment, including the individuals responsible for group assignment. Study protocols should include a process for accurately recording all eligible participants, those who are enrolled and those who refuse, and any participants who withdraw during the course of the study. Data on reasons for refusal or withdrawal and missed eligible participants should also be systematically collected.

Integrity of Intervention

Throughout the course of a study, researchers must constantly verify that the intervention remains stable over time as well as from place to place and person to person (participants, caregivers, researchers, etc.). When trials involve human subjects or complex interventions, many uncontrollable psychological effects and non-specific treatment effects can occur. To minimize these unintended effects, one must first be aware of them. The following section describes biases that could affect the integrity of interventions and how they can be minimized.

Intervention Fidelity

Intervention fidelity can be defined as the extent to which an intervention is carried out consistently, as planned, throughout all stages of the study (Bellg et al., 2004). It is considered central to the evaluation, comparison, and dissemination of all intervention research (Horner, Rew, & Torres, 2006). The effects of even the most well-defined intervention cannot be fully interpreted unless specific processes to ensure receipt and evaluation of the intervention have been put in place. Consistency in intervention delivery is often directly correlated with the complexity of the intervention, the number of sites, and the duration of the study. Several aspects of an intervention can affect fidelity. These include design, training, delivery, receipt, and enactment (Dumas, Lynch, Laughlin, Phillips Smith, & Prinz, 2001). To ensure fidelity, researchers should ask themselves: Have I provided a detailed definition using a combination of verbal, written, and electronic means that convey all aspects of the intervention to those providing the intervention? Can I guarantee that all the providers will be trained in a consistent manner? Have I included specific criteria to assess delivery outcomes? Have I incorporated ways to maintain provider competence and consistency over time by including an evaluation and feedback process? How will I know if the participants received the appropriate intervention?

Intervention-monitoring tools can be quantitative and/or qualitative in nature. They may consist of simple questions (*yes/no*) or be more descriptive (*none, adequate, excellent*) (Dumas et al., 2001). Providers and participants may simply be asked on a random basis about the delivery of the intervention (Orwin, 2000), or there may be a more sophisticated system. For example, in a large study examining methods for improving diabetes management in the community, a virtual networking system was used (Minnick, Catrambone, Halstead, Rothschild, & Lapidos, 2008). Similarly, in a large-scale prevention trial testing the effectiveness of family, peer, and school interventions for conduct disorder, substance abuse, and school failure, researchers incorporated an extensive fidelity

check that included a review of randomly selected videotaped sessions. Trained coders recorded adherence to the intervention protocol and evaluated the delivery technique of the provider, to ensure that the fidelity of both content and process was evaluated (Dumas et al., 2001). If coders are to be used in this manner, the researcher must also incorporate inter- and intra-rater coder reliability checks into the proposal (Santacroce, Maccarelli, & Grey, 2004).

Co-interventions

The addition of other treatments that are not included in the study protocol could influence the study's outcomes. These are known as co-interventions. In general, a balance of co-intervention use across study groups will dilute the observed treatment effect and an imbalance will introduce bias. Take the example of a study that evaluated the effect of kangaroo care as non-pharmacological pain relief for painful procedures in preterm infants. The introduction of a practice policy that allowed the administration of a 24% sucrose solution to the infants prior to such procedures was a co-intervention that could possibly have interacted with the study outcome (Johnston et al., 2008). Specifically, if the possibility of a co-intervention was not recognized prior to study commencement, the researcher could not incorporate methods to monitor for or prevent its use during the study. A pilot test is a valuable means of identifying such co-interventions and can help researchers in controlling this type of bias throughout the trial. Once it is identified, the researcher may choose to measure and control for the co-intervention (a priori) in the analysis, or may include its use as part of the intervention definition to ensure a balance between groups.

Contamination

Contamination can occur when participants in either group become aware of the treatment that the other group is receiving (Torgerson, 2001). This is especially relevant in trials where the intervention cannot be blinded, and if it occurs more than minimally it can destroy the internal validity of the trial. Consider a trial in a postpartum unit where some mothers are in the experimental group and others are in the comparison group. Bearing in mind that most mothers do not have a private room, contamination between these participants could occur when they talk among themselves or witness differences between their treatments. One method for minimizing this type of contamination is cluster randomization. In a cluster trial, groups of participants, rather than individuals, are randomized to the intervention or comparison group (Torgerson, 2001). Cluster allocation is not without its drawbacks. The randomization of groups requires much larger sample sizes, which could increase the length

and complexity of the trial, as well as its costs (Torgerson, 2001). Torgerson (2001) argues that unless the anticipated contamination rate is greater than 30%, contamination is more efficiently dealt with by individual randomization of an increased sample size — thus avoiding the above-mentioned disadvantages. Researchers should thoroughly reflect on the pros and cons of the cluster approach before applying it to their study design. Alternatively, researchers may include qualitative analysis to assess participants' views on treatment credibility as a means to quantify the effect of potential contamination influence on outcomes (Licciardone & Russo, 2006).

Analysis

Attrition

When participants drop out before the end of a trial or before the end of the experimental phase, bias can occur. This type of bias is known as sample attrition, and it may affect both the external and the internal validity of a trial (Barry, 2005). Attrition rates that are well balanced between groups can contribute to reduced statistical power and generalizability of outcomes (Leon et al., 2006). However, imbalanced attrition is more problematic. When this happens, the characteristics of the remaining participants, both within and between groups, differ significantly from those of the participants who have dropped out. This creates difficulties in determining whether outcomes are related to the intervention or to attrition. Prevention of attrition is a key factor in all studies and is especially important in studies where participants are not blinded to the intervention being tested (Leon et al., 2009).

Qualitative research is an excellent way to determine the potential for attrition, because it is suited to studying the variations of complex human behaviour. The use of interview or focus groups with potential participants prior to the study, or with those who have failed to complete the study, offers valuable insights into why people do not wish to enrol in a study or why participants choose to drop out (Lewin, Glenton, & Oxman, 2009). For example, if used prior to the study, qualitative research methods can determine the degree of participants' preconceived likes and dislikes regarding the intervention. Additional incentives or an alteration in the protocol can then be used to reduce potential attrition. Alternatively, qualitative methods can be used to explain the specific reasons for dropping out and possible variations between those who continue with the study and those who do not.

Attrition bias can also result from missing and incomplete data. For example, it may result from participants failing to answer all questions on a questionnaire. A pilot study to pre-test the questionnaire or the inclusion of follow-up phone calls may prevent this type of bias (Hayward

et al., 2007). Nonetheless, when it occurs, researchers may deal with it by using a variety of approaches, including creating an imputed data set (Donders, van der Heijden, Stijnen, & Moons, 2006). Although this approach can introduce additional bias, if missing items make up less than 5% of data this method has a minimal effect on overall results.

Intention to Treat

According to Eysenbach (2005), intention to treat (ITT) analysis is the only reliable way to avoid attrition bias. However, ITT analysis is not a perfect solution, since it significantly decreases a study's power to identify differences between groups. It is used to analyze all patients assigned to a study group, regardless of whether there was contamination, whether they complied with treatment, or whether they completed the trial (Fergusson, Aaron, Guyatt, & Hebert, 2002). There are various definitions of ITT, and there is no consensus among researchers on when it should or should not be applied. The benefit of using ITT analysis is that it maintains group allocation, allowing confounding factors to be balanced across groups. For example, if someone who was allocated to receive the treatment intervention was missed and received usual care, he or she would still be included in the analysis as part of the treatment group. This method ensures that although the oversight may have been random, any unknown sources of bias, such as timing of the intervention or differences in care providers, that could falsely influence the outcomes are controlled for. In their survey of published RCTs, Hollis and Campbell (1999) found that about half of these had used ITT analysis but had applied it in various ways. In addition, several studies used inadequate methods to deal with missing data on the primary outcome variable. Intention to treat analysis is best applied when complete outcome data are available for all randomized participants. The authors recommend that researchers make every attempt to follow up on all participants who have abandoned the trial, in order to decrease the rate of missing data for the primary outcome.

Ascertainment of Outcome

Ascertainment bias occurs when outcomes are erroneously attributed to the phenomenon under study. It can be introduced by the people who deliver the intervention, the participants, or the people collecting and analyzing the data. Ascertainment bias may be introduced if the research assistant or principal investigator has certain beliefs about the study or is not blinded to the allocation of participants to the different treatment groups. In addition, participants who know which group they have been allocated to could influence the outcomes of the study (non-blinded allocation).

Various strategies are recommended at different points within the trial to minimize ascertainment bias. During the data-collection phase, the best approach is to blind the investigator and participants. This method is known as double-blinding. While double-blinding is standard practice for most experimental trials, it is often unfeasible in nursing intervention trials because both the use of a placebo and the blinding of the caregiver can rarely be achieved. Thus, single-blinded designs, with the investigator or, more importantly, the evaluator of the outcome remaining blinded, are more common. Examples include videotaping of participant responses that are objectively scored by blinded and trained coders (Johnston et al., 2009) and direct entry of patient responses into computerized databanks followed by analysis, without knowledge of group allocation.

Hawthorne Effect

It is recognized that merely participating in a study can influence a participant's behaviour, thereby affecting the outcome. This phenomenon, known as the Hawthorne effect, is a result of the increased attention and support that participants receive with trial participation. Given the complexity of nursing and therapeutic relationships, nursing intervention trials are prone to the Hawthorne effect. However, this effect can be limited by means of a protocol design that allows for equal time spent with participants in the two groups (McCarney et al., 2007).

Summary

We have argued that the internal validity of any study is dependent on the level of bias that is introduced. Our checklist provides an overview to assist researchers in anticipating and controlling bias during the design and conduct of intervention trials. This checklist is intended for nurse researchers who wish to use a systematic approach in the preparation of research protocols, so that potential sources of bias can be avoided. Despite considerable progress in reducing bias in clinical trials, current tools focus on double-blind RCTs and on reporting rather than trial planning. While our checklist does include similar concepts, such as sequence generation, allocation, blinding, and incomplete outcome data (www.consort-statement.org), we have included additional concepts, in particular study design, selection of comparisons, and definition and integrity of interventions. These concepts are especially relevant for preparing and conducting complex intervention studies. Their inclusion complements the work of other authors who have highlighted differences in the reporting of non-pharmacological trials (Boutron & Ravaud, 2009) and pragmatic trials (Zwarenstein et al., 2008) when compared to a

double-blind randomized controlled drug trial. In a recent review highlighting strategies for improving the quality and explanatory power of nursing science, Borglin and Richards (in press) found that randomization alone was a necessary but insufficient method for reducing bias in intervention trials. They argue the importance of careful participant selection, consistent performance of the intervention, reduction of attrition, and blinding of assessors. We anticipate that adherence to this checklist will lead to improvements in the scientific rigour of intervention trials. We hope it will also serve to strengthen the impact of complex nursing intervention studies in the wider field of medicine.

Although this article focuses on nursing interventions, it is also relevant for other disciplines conducting similar types of complex intervention research — for example, surgery, complementary medicine, physical therapy, or the social and health sciences, all of which can present similar challenges.

References

Balk, E. M., Bonis, P. A., Moskowitz, H., Schmid, C. H., Ioannidis, J. P., Wang, C., et al. (2002). Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *Journal of the American Medical Association, 287*(22), 2973–2982.

Bang, H., Ni, L., & Davis, C. E. (2004). Assessment of blinding in clinical trials. *Controlled Clinical Trials, 25*(2), 143–156.

Barry, A. E. (2005). How attrition impacts the internal and external validity of longitudinal research. *Journal of School Health, 75*(7), 267–270.

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., et al. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology, 23*(5), 443–451.

Blair, E. (2004). Gold is not always good enough: The shortcomings of randomization when evaluating interventions in small heterogeneous samples. *Journal of Clinical Epidemiology, 57*(12), 1219–1222.

Borglin, G., & Richards, D. (in press). Bias in experimental nursing research: Strategies to improve the quality and explanatory power of nursing science. *International Journal of Nursing Studies.*

Boutron, I., & Ravaud, P. (2009). Improving the reporting of clinical research. *British Journal of Surgery, 96*(5), 449–450.

Bridgman, S., Engebretsen, L., Dainty, K., Kirkley, A., Maffulli, N., & ISAKOS Scientific Committee. (2003). Practical aspects of randomization and blinding in randomized clinical trials. *Arthroscopy: Journal of Arthroscopic and Related Surgery, 19*(9), 1000–1006.

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., et al. (2000). Framework for design and evaluation of complex interventions to improve health. *British Medical Journal, 321*(7262), 694–696.

Conn, V. S., Cooper, P. S., Ruppar, T. M., & Russell, C. L. (2008). Searching for the intervention in intervention research reports. *Journal of Nursing Scholarship, 40*(1), 52–59.

Doig, G. S., & Simpson, F. (2005). Randomization and allocation concealment: A practical guide for researchers. *Journal of Critical Care, 20*(2), 187–193.

Donders, A. R., van der Heijden, G. J., Stijnen, T., & Moons, K. G. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology, 59*(10), 1087–1091.

Dumas, J. E., Lynch, A. M., Laughlin, J. E., Phillips Smith, E., & Prinz, R. J. (2001). Promoting intervention fidelity: Conceptual issues, methods, and preliminary results from the EARLY ALLIANCE prevention trial. *American Journal of Preventive Medicine, 20*(Suppl 1), 38–47.

Eysenbach, G. (2005). The law of attrition. *Journal of Medical Internet Research, 7*(1), e11.

Fergusson, D., Aaron, S. D., Guyatt, G., & Hebert, P. (2002). Post-randomization exclusions: The intention to treat principle and excluding patients from analysis. *British Medical Journal, 325*(7365), 652–654.

Forder, P. M., Gebski, V. J., & Keech, A. C. (2005). Allocation concealment and blinding: When ignorance is bliss. *Medical Journal of Australia, 182*(2), 87–89.

Glasziou, P., Meats, E., Heneghan, C., & Shepperd, S. (2008). What is missing from descriptions of treatment in trials and reviews? *British Medical Journal, 336*(7659), 1472–1474.

Hawe, P., Shiell, A., & Riley, T. (2008). In response to Spillane V., Byrne M.C., Byrne M., Leathem C.S., O'Malley M. & Cupples M.E. (2007). Monitoring treatment fidelity in a randomized trial of a complex intervention. *Journal of Advanced Nursing, 60*(3), 343–352. Important considerations for standardizing complex interventions. *Journal of Advanced Nursing, 62*(2), 267.

Hayward, K., Campbell-Yeo, M., Price, S., Morrison, D., Whyte, R., Cake, H., et al. (2007). Co-bedding twins: How pilot study findings guided improvements in planning a larger multicenter trial. *Nursing Research, 56*(2), 137–143.

Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomized controlled trials. *British Medical Journal, 319*(7211), 670–674.

Horner, S., Rew, L., & Torres, R. (2006). Enhancing intervention fidelity: A means of strengthening study impact. *Journal for Specialists in Pediatric Nursing, 11*(2), 80–89.

Johnston, C. C., Filion, F., Campbell-Yeo, M., Goulet, C., Bell, L., McNaughton, K., et al. (2008). Kangaroo mother care diminishes pain from heel lance in very preterm neonates: A crossover trial. *BMC Pediatrics, 8*(13). Online: http://www.biomedcentral.com/1471-2431/8/13.

Johnston, C. C., Filion, F., Campbell-Yeo, M., Goulet, C., Bell, L., McNaughton, K., et al. (2009). Enhanced kangaroo mother care for heel lance in preterm neonates: A crossover trial. *Journal of Perinatology, 29*(1), 51–56.

Juni, P., Altman, D. G., & Egger, M. (2001). Systematic reviews in health care: Assessing the quality of controlled clinical trials. *British Medical Journal, 323*(7303), 42–46.

Kunz, R., Vist, G., & Oxman, A. D. (2007). Randomization to protect against selection bias in healthcare trials. *Cochrane Database of Systematic Reviews* (Online), *2*(2), MR000012.

Leon, A. C., Mallinckrodt, C. H., Chuang-Stein, C., Archibald, D. G., Archer, G. E., & Chartier, K. (2006). Attrition in randomized controlled clinical trials: Methodological issues in psychopharmacology. *Biological Psychiatry, 59*(11), 1001–1005.

Lewin, S., Glenton, C., & Oxman, A. D. (2009). Use of qualitative methods alongside randomized controlled trials of complex healthcare interventions: Methodological study. *British Medical Journal, 339*(3496), 732–735.

Licciardone, J. C., & Russo, D. P. (2006). Blinding protocols, treatment credibility, and expectancy: Methodologic issues in clinical trials of osteopathic manipulative treatment. *Journal of the American Osteopathic Association, 106*(8), 457–463.

Lindsay, B. (2004). Randomized controlled trials of socially complex nursing interventions: Creating bias and unreliability? *Journal of Advanced Nursing, 45*(1), 84–94.

Mann, H. (2007). Deception in the single-blind run-in phase of clinical trials. *IRB: Ethics and Human Research, 29*(2), 14–17.

Mann, H., & Djulbegovic, B. (2003). Choosing a control intervention for a randomized clinical trial. *BMC Medical Research Methodology, 3*(7). Online: www.biomedcentral.com/1471-2288/3/7.

McCarney, R., Warner, J., Iliffe, S., van Haselen, R., Griffin, M., & Fisher, P. (2007). The Hawthorne effect: A randomized controlled trial. *BMC Medical Research Methodology,* 7(30). Online: www.biomedcentral.com/1471-2288/7/30.

Minnick, A., Catrambone, C. D., Halstead, L., Rothschild, S., & Lapidos, S. (2008). A nurse coach quality improvement intervention: Feasibility and treatment fidelity. *Western Journal of Nursing Research, 30*(6), 690–703.

Norman, G. R., & Streiner, D. L. (Eds.). (2000). *Biostatistics: The bare essentials,* 2nd ed. London: B. C. Decker.

Orwin, R. G. (2000). Assessing program fidelity in substance abuse health services research. *Addiction, 95*(Suppl 3), S309–S327.

Paterson, C., & Dieppe, P. (2005). Characteristic and incidental (placebo) effects in complex interventions such as acupuncture. *British Medical Journal, 330*(7501), 1202–1205.

Rees, J. R., Wade, T. J., Levy, D. A., Colford, J. M. Jr., & Hilton, J. F. (2005). Changes in beliefs identify unblinding in randomized controlled trials: A method to meet CONSORT guidelines. *Contemporary Clinical Trials, 26*(1), 25–37.

Santacroce, S. J., Maccarelli, L. M., & Grey, M. (2004). Intervention fidelity. *Nursing Research, 53*(1), 63–66.

Schulz, K. F., & Grimes, D. A. (2002). Allocation concealment in randomized trials: Defending against deciphering. *Lancet, 359*(9306), 614–618.

Silverman, H. J., & Miller, F. G. (2004). Control group selection in critical care randomized controlled trials evaluating interventional strategies: An ethical assessment. *Critical Care Medicine, 32*(3), 852–857.

Torgerson, D. J. (2001). Contamination in trials: Is cluster randomization the answer? *British Medical Journal, 322*(7282), 355–357.

Zwarenstein, M., Treweek, S., Gagnier, J. J., Altman, D., Tunis, S., Haynes, B., et al. (2008). Improving the reporting of pragmatic trials: An extension of the CONSORT statement. *British Medical Journal, 337*(2390).

Acknowledgements

The authors thank Kate Johnson for editorial assistance.

Marsha Campbell-Yeo is supported by a Clinical Initiative Fellowship from the Canadian Institutes of Health Research (CIHR). Manon Ranger is supported by Fonds de la recherche en santé du Québec (FRSQ). Both of these authors are supported by the Quebec Inter-university Nursing Intervention Research Group (GRIISIQ). Dean Fergusson is the recipient of a CIHR New Investigator Award.

Conflict of interest and financial disclosure: none to declare.

Marsha Campbell-Yeo, RN, MN, NNP, is Neonatal Nurse Practitioner, IWK Health Centre, Halifax, Nova Scotia, Canada; and a PhD candidate at the School of Nursing, McGill University, Montreal, Quebec, Canada. Manon Ranger, RN, MSc, is a PhD candidate at the School of Nursing, McGill University. Celeste Johnston, RN, DEd, is Professor, School of Nursing, McGill University. Dean Fergusson, MHA, PhD, is Senior Scientist and Deputy Director, Ottawa Health Research Institute, and Associate Professor, Faculty of Medicine and Department of Epidemiology and Community Medicine, University of Ottawa, Ontario, Canada.