13

PILOT STUDIES FOR RANDOMIZED CLINICAL TRIALS

Nancy Feeley and Sylvie Cossette

A pilot study is a small-scale study conducted to test the plan and method of a research study (Ballman, 2008). A pilot investigation is conducted preparatory to a subsequent, adequately powered study (Conn, Algase, Rawl, Zerwic, & Wyman, 2010) and is designed to try out, evaluate, and select the methods to be used in the larger study (Polit & Beck, 2012). Pilot studies are increasingly important in nursing research because they contribute knowledge about the feasibility and acceptability of research methods and because pilot data are needed to justify use of specific methods proposed in applications for full-scale investigations.

A pilot study may be conducted prior to any type of major study. This chapter will focus on use of pilot studies to inform planning for efficacy trials of nursing interventions. Efficacy trials (Phase III trials) estimate intervention effects under ideal conditions (Campbell et al., 2000) and are commonly called randomized clinical trials (RCTs). The RCT design is a critical tool for creating evidence for practice because it is the most rigorous way to assess causality (Brown, 2002; Rubin & Zell, this volume; Sidani & Braden, 2011). Piloting is particularly important before a full-scale RCT due to their complexities, challenges, and expense. Both the United Kingdom's Medical Research Council guidelines for developing, evaluating, and implementing complex interventions (Craig et al., 2008) and the U.S. National Institutes of Health clinical trial phases adapted for nursing intervention development and testing (Whittemore & Grey, 2002) highlight the importance of testing the protocol of an RCT before proceeding to an evaluation of intervention effects. Thus, the primary purpose of a pilot study is to refine the protocol for the full-scale study by shedding light on strengths, inadequacies, or omissions of the preliminary plan (Conn et al., 2010; Feeley et al., 2009; Loscalzo, 2009; Polit & Beck, 2012; Shanyinde, Pickering, & Weatherall, 2011; Thabane et al., 2010). Findings from the pilot study are utilized to optimize the protocol for the full-scale study, so that it can be successfully and efficiently executed to generate the highest quality evidence for nursing practice.

Pilot study data are not used to estimate intervention efficacy (Polit & Beck, 2012). Reviews of published pilot studies suggest that this is often misunderstood. For example, 81% of "pilot" studies published in seven major medical journals from 2007–2008 inappropriately included hypothesis-testing (Arain, Campbell, Cooper, & Lancaster, 2010). Another review of published pilot study findings indexed in MEDLINE and EMBASE over the past decade found that only 56% addressed methodological issues (Shanyinde et al., 2011). These findings suggest that there is a need to clarify and emphasize what are the appropriate objectives of pilot RCTs.

Aims

The aims of this chapter are to: (a) define feasibility and acceptability, (b) identify design elements that should be assessed for feasibility and acceptability prior to conducting a full-scale trial, (c) describe the decisions that can be taken based on the findings, and (d) summarize current controversies in the use of pilot studies. The focus is on feasibility and acceptability questions emanating from study design. Assessment of feasibility and acceptability of interventions has been described elsewhere (Sidani & Braden, 2011).

Feasibility and acceptability

Both feasibility and acceptability of design elements should be assessed in pilot RCTs. Feasibility is concerned with the researcher's ability to provide the intervention and complete the study as planned (Feeley et al., 2009). For example, the research team's ability to execute the data collection plan should be examined. Acceptability refers to the suitability of the methods or the intervention from the perspective of the study population, the intervention providers, or health care professionals (Feeley et al., 2009). For example, the researcher can explore participants' perceptions and responses to the data collection procedures. Do they find the procedures too time-consuming, inappropriate, or irrelevant? Feasibility and acceptability are often interrelated. If study participants find the data collection too time consuming (e.g., low acceptability), the feasibility of the data collection plan will also be poor.

Feasibility and acceptability of both the study design and the intervention should be examined. The specific feasibility and acceptability questions examined will depend on the particular methodological challenges anticipated in the full-scale trial, as well as challenges related to the provision of the intervention. It is imperative that specific, clear objectives for the pilot study be articulated, and rigorously evaluated. Furthermore, indicators for assessing each study question should also be identified a priori. Explicit criteria should be identified for determining whether to proceed with a specific design element in the full-scale trial (Arain, Campbell, Cooper, & Lancaster, 2010).

Assessing design elements

Overall design

Various study designs can be used in a pilot study preceding an RCT, including experimental or quasi-experimental designs. There are advantages and disadvantages of each of these options. Nonetheless, it is usually optimal to pilot the same design planned for the full-scale study. The number of groups to be included is an important decision. Choices for comparison groups include: no intervention (control), usual care, an alternative intervention, or varying doses of the intervention. A three group design is useful to compare the experimental intervention to both an attention control and usual care group, as it controls for the effects of the passage of time and attention. However, this design requires a greater number of participants. In the two group design, the researcher needs to consider which control condition to utilize, and there are limitations to either choice. If the experimental group is compared to usual care any effects observed might be attributed to the extra time and attention provided, whereas when the comparison group is an attention control condition, the intervention effect might not be detectable. To assess the efficacy of different components of complex interventions, factorial designs with two or more crossed components should be used (Maxwell & Delaney, 2003).

In a pilot study, numerous design elements can be examined to inform the design of the full-scale study as virtually every element can be assessed. The researcher decides which elements are most crucial to assess based on their understanding of the inherent methodological and feasibility challenges of the design and procedures. Specific research questions are determined, along with the indicators that will be utilized to answer these questions. Once these pilot data have been collected, decisions are taken based on how to modify the design and procedures for the full-scale RCT that will follow.

For example, in a pilot study of an intervention to reduce mother's anxiety and enhance her ability to interact with her very low birth weight infant in neonatal intensive care (Feeley et al., 2008), one of the research questions was: Is it feasible to observe and measure mother-infant interaction in the neonatal intensive care unit (NICU) prior to intervention? The research staff documented all the challenges and issues arising during their observations of mother–infant interaction, as well as those related to scoring the measure. These data were analyzed and the findings indicated that it was not feasible to measure mother–infant interaction in this context for a number of reasons including not being able to hear mothers' verbalizations, and too little interaction between the mother and her newborn. The decision was made to omit this pre-intervention observation from the full-scale RCT that followed (Feeley et al., 2012). This is just one way in which the pilot study findings shaped the design of the RCT. Furthermore, the findings are also useful to other researchers planning to measure mother–infant interaction in this setting.

Tables 13.1 and 13.2 outline important design elements and procedures that can be assessed for feasibility and acceptability in a pilot prior to an RCT, but this list is not exhaustive. Examples of feasibility and acceptability questions that can guide the pilot study, along with examples of the indicators to answer these questions are included. Decisions concerning the design of the subsequent trial will follow naturally from the findings of the pilot study.

The Consolidated Standards of Reporting RCTs (CONSORT) statement is an evidence-based set of guidelines now widely adopted by journals in many disciplines to improve the reporting of RCTs, and enable readers to evaluate the methodological rigor of these studies (Moher et al., 2010). The CONSORT guidelines can also assist investigators to enhance the design of RCTs and pilot studies by alerting them to the methodological issues that should be addressed (Schulz, Altman, Moher, & CONSORT Group, 2010). At the pilot stage, the researcher should be cognizant of the standards and design the pilot study and the RCT with these guidelines in mind. The following section will discuss some of the key design elements that should be assessed, including those that arise from the CONSORT requirements.

Participant identification and screening

The success of the eventual full-scale RCT depends on enrolling a sufficient number of participants. A well-developed approach to screening potential participants is needed to avoid recruiting and randomizing ineligible persons (Polit & Gillespie, 2010). Thus, developing procedures to identify and assess the eligibility of potential participants should be explored (Conn et al., 2010). Further, the feasibility and acceptability of these procedures should be assessed.

Recruitment and consent

Questions related to recruitment are the focus of a pilot study when the researcher is uncertain about whether sufficient numbers of eligible participants can be accessed at designated study sites. Estimates of the proportion of eligible participants who will agree to participate are

Table 13.1 Design elements and procedures: assessment for feasibility in a pilot RCT

Element/Procedure	*Questions*	*Indicators*
Participants	What are the most effective methods for identifying and accessing the population of interest? (e.g., list of appointments, while hospitalized)	• Number of potential participants identified using various methods • % screened at various sites
Screening	How much time is required to screen potential participants?	• Record time needed to screen
Recruitment	What is the number of participants recruited of those screened? (information needed to assess generalizability and provides information about eligibility criteria)	• Document reasons for exclusion (e.g., type of illness, language barriers, severity of the illness) • % of persons screened who enroll
	Are recruitment strategies effective?	• Characteristics of participants
	What are the characteristics of those who agree to participate? Are these the participants desired?	• Length of time to recruit, and time of recruitment
	How much time is required to recruit participants?	
	When is the best time to recruit participants?	
	Can an adequate number of participants be recruited at the study sites? (assess the adequacy of settings for recruitment)	
Consent	Can eligible participants provide consent?	• Difficulties with consent arising from challenges such as language, literacy, or cognitive ability
	Is it feasible to provide potential participants with the privacy and time to reflect on their decision?	
Randomization and Blinding	Is the randomization procedure effective?	• Test various methods to allocate to group (e.g., opaque envelope vs. telephone line vs. computer-generated or website)
	Are the methods used to blind participants, assessors, and/or staff feasible?	• Number ineligible persons randomized • Reasons and circumstances for which ineligible persons were randomized
Outcome Measures: Selection	What outcome constructs should be assessed?	• Assess two similar constructs and determine which is most appropriate (e.g., a general measure of self-care vs. one that is specific to a particular illness)
	What measures will best capture the constructs of interest? (i.e., validity)	• Administer two measures of same construct and document responses of participants to each (i.e., ease of responding) • Comments and questions of participants during data collection

Table continues

Element/Procedure	*Questions*	*Indicators*
Outcome Measurement: Sensitivity to Change/Timing	Is the outcome measure sensitive to change?	• Assess test-retest reliability without any intervention, and it should be high
	When will change occur (possible delayed effects)?	• Assess outcomes at different point of time to identify when change occurs
	How long will change persist (sustained effects)?	• Seek evidence that intervention has an effect on outcome (eyeballing vs. hypothesis testing, qualitative interviews with participants)
	Are the effects in the expected direction?	Document and record any unanticipated effects
	Are there any negative effects or harms?	• % measures completed as indicated in protocol

	time to reflect on their decision?	
Randomization and Blinding	Is the randomization procedure effective? Are the methods used to blind participants, assessors, and/or staff feasible?	• Test various methods to allocate to group (e.g., opaque envelope vs. telephone line vs. computer-generated or website) • Number ineligible persons randomized • Reasons and circumstances for which ineligible persons were randomized
Outcome Measures: Selection	What outcome constructs should be assessed? What measures will best capture the constructs of interest? (i.e., validity)	• Assess two similar constructs and determine which is most appropriate (e.g., a general measure of self-care vs. one that is specific to a particular illness) • Administer two measures of same construct and document responses of participants to each (i.e., ease of responding) • Comments and questions of participants during data collection

Table continues

Element/Procedure	*Questions*	*Indicators*
Outcome Measurement: Sensitivity to Change/Timing	Is the outcome measure sensitive to change? When will change occur (possible delayed effects)? How long will change persist (sustained effects)? Are the effects in the expected direction? Are there any negative effects or harms?	• Assess test-retest reliability without any intervention, and it should be high • Assess outcomes at different point of time to identify when change occurs • Seek evidence that intervention has an effect on outcome (eyeballing vs. hypothesis testing, qualitative interviews with participants) Document and record any unanticipated effects
Data Collection Procedures	Is it possible to collect data as planned? How long does it take participants to complete the measures? Can participants complete study measures in the time available? Is the sequencing of measures administration appropriate? Are measures being completed by the appropriate participant? (e.g., determine whether peers or family members respond as proxies for participants) Can the outcome assessors be trained to reliably assess the outcome?	• % measures completed as indicated in protocol • Reasons for deviations from plan • Length of time to complete • Number of missing responses on measure items • Other issues (e.g.. missing data in the last questionnaire; sensitive questions at the beginning or the end of the questionnaire) • Inter-rater agreement/reliability • Issues with training of the assessors
Retention	What are the rate, timing and reasons for attrition? Is there a relationship between attrition and participant characteristics or site? How effective are the methods of follow-up? What is the best window of time to reach participants for follow-up?	• Reasons for withdrawing • Reasons those who are retained remain • % outcome data obtained • Reasons for not obtained outcome data as indicated • Assess strategy to increase retention (e.g., birthday card, regular telephone call)
Contamination	Is contamination possible? How?	• % of control group participants who are exposed to experimental content
Cross-over	Does cross-over of participants from one group to another occur? How?	• % of participants who cross-over from one group to another

Note: The overall design should first be considered in light of the scientific question under study. Indicators should be documented and examined.

Table 13.2 Design elements and procedures: assessment for acceptability in a pilot RCT

Element/Procedure	*Questions*	*Indicators*
Recruitment	Will eligible potential participants agree to participate?	• % of persons screened who enroll • Reasons for refusal and how they might be addressed
Consent	Do eligible potential participants obtain all the information that they desire to make an informed decision?	• Difficulties with consent (e.g., language, literacy, or cognitive ability)
Randomization and Blinding	Will participants agree to be randomized? Will participants agree to be blinded? Will clinicians find the allocation methods acceptable?	• Responses to the allocation methods • Reasons for not agreeing to random assignment • Reasons for refusing to be blind to the intervention • Clinicians responses to the allocation methods
Outcome Measures: Selection	Are the measures acceptable to participants? Is one measure more acceptable to participants than another? Can participants comprehend the study measures?	• Assess similar constructs and determine which is most appropriate. (e.g., general versus illness-specific measure of self-care) • Administer two measures of same construct and document responses of participants to each (i.e., ease of responding) • Comments and questions of participants during data collection • Participant responses to systematically queries about the acceptability of measures
Outcome Measurement: Sensitivity to Change/ Timing	Is the timing of the outcome assessment acceptable to participants?	• Participants responses to systematic queries about acceptability of the timing of measures
Data Collection Procedures	Will participants agree to and be comfortable providing the data desired? (e.g., blood draws, observations, questions of a sensitive nature) Is the time to complete measures acceptable to participants? Is the sequencing of measures administration acceptable?	• % measures completed as indicated in protocol • Reasons for deviations from plan • Length of time to complete • Missing responses to items • Other issues (e.g., missing data in the last questionnaire; sensitive questions at the beginning or the end of the questionnaire) • Responses to systematic queries about acceptability of procedures
Retention	Will participants find the study procedures acceptable and complete these as planned? What strategies might enhance the retention of participants?	• % outcome data obtained • Responses to interview or questionnaire to elicit reasons for withdrawing, and/or reasons those who are retained remain • Reasons for not obtaining outcome data as indicated • Responses about what strategies enhanced retention
Contamination/Cross-over	Do participants seek other or alternative treatments during the study?	• % of participants who cross-over from one group to another

Note. The overall design should first be considered in light of the scientific question under study. Indicators should be documented and examined.

important for providing evider
2008), and also provide inform
reflect the acceptability of the
to estimate how much time is
(Lancaster, Dodd, & Williamso

The characteristics of the pa
mine whether the inclusion and
2009). Often researchers discov
study require modification beca
or are too restrictive and need t
consent process should be explo
of these data are used to determ
recruitment procedures or sites,

I

Randomization is an essential fo
zation is not suitable to achieve
can be utilized such as block, str
population will agree to be ran
a variety of reasons that the stu
group. For example, if the expe
a trial. People with life-threate
placebo group if they could ha

Various methods can be use
bers table, or a computer gener
concealing the allocation also n
cannot acquire knowledge of
based on this knowledge (e.g.,
et al., 2010). Feasibility of proc
of the allocation should both b

Another important methods
ticipants, their health care provi
Blinding means concealing inf
uals so that they will not be ir
CONSORT statement for RC
must report the blinding status
(Boutron et al., 2008b).

Blinding participants in nurs
In many nursing interventions
cognizant of which interventio
sification scheme to describe c
health care providers in RCTs
scheme can be useful for inve
of blinding, such as blinding p
study hypothesis has been use
the intervention. Participants c
course, there may be ethical co

		• Reasons for deviations from plan
		• Length of time to complete
	Is the time to complete measures acceptable to participants?	• Missing responses to items
	Is the sequencing of measures administration acceptable?	• Other issues (e.g., missing data in the last questionnaire; sensitive questions at the beginning or the end of the questionnaire)
Retention	Will participants find the study procedures acceptable and complete these as planned?	• Responses to systematic queries about acceptability of procedures • % outcome data obtained
	What strategies might enhance the retention of participants?	• Responses to interview or questionnaire to elicit reasons for withdrawing, and/or reasons those who are retained remain • Reasons for not obtaining outcome data as indicated • Responses about what strategies enhanced retention
Contamination/Cross-over	Do participants seek other or alternative treatments during the study?	• % of participants who cross-over from one group to another

Note. The overall design should first be considered in light of the scientific question under study. Indicators should be documented and examined.

important for providing evidence of the viability of the RCT in grant applications (Hertzog, 2008), and also provide information about the eventual generalizability of the trial as it may reflect the acceptability of the intervention (Rothwell, 2006). These data should also be used to estimate how much time is required to recruit the desired sample for the full-scale study (Lancaster, Dodd, & Williamson, 2004).

The characteristics of the participants who agree to participate should be examined to determine whether the inclusion and exclusion criteria are effective or require revision (Arnold et al., 2009). Often researchers discover that the inclusion and exclusion criteria as stated for the pilot study require modification because the criteria fail to exclude persons who should be excluded, or are too restrictive and need to be changed to enlarge the pool of eligible persons. As well, the consent process should be explored if the researcher anticipates that this might be challenging. All of these data are used to determine if revisions are needed to the inclusion and exclusion criteria, recruitment procedures or sites, or consent process for the full-scale study.

Randomization and blinding

Randomization is an essential feature of an RCT. If the pilot demonstrates that simple randomization is not suitable to achieve equivalence between groups, other approaches to randomization can be utilized such as block, stratified, or minimization. It is important to ascertain that the target population will agree to be randomly assigned to group (Lackey & Wingate, 1998). There are a variety of reasons that the study population might not be willing to be randomly assigned to group. For example, if the experimental intervention is available, it would be difficult to conduct a trial. People with life-threatening illnesses may not be willing to be randomly assigned to a placebo group if they could have access to cutting-edge experimental treatment.

Various methods can be used to generate the random allocation sequence (e.g., random numbers table, or a computer generated list of random numbers (Moher et al., 2010). A method of concealing the allocation also needs to be utilized, so that staff and others who enroll participants cannot acquire knowledge of group assignment and include or exclude possible participants based on this knowledge (e.g., centralized randomization telephone service or a website (Moher et al., 2010). Feasibility of procedures for generation of the allocation sequence and concealment of the allocation should both be tested at the pilot stage.

Another important methodological feature of RCTs utilized to reduce bias is blinding of participants, their health care providers, outcome assessors, and data analysts (Boutron et al., 2008a). Blinding means concealing information about the group assignment from all of these individuals so that they will not be influenced by this information (Moher et al., 2010). The current CONSORT statement for RCTs of non-pharmacological interventions states that researchers must report the blinding status of participants, their health care providers, and outcome assessors (Boutron et al., 2008b).

Blinding participants in nursing intervention studies is often not possible or difficult to achieve. In many nursing interventions participants often take part in the intervention and are thus fully cognizant of which intervention they receive. Boutron and colleagues (2007) developed a classification scheme to describe creative methods that have been used to blind participants or their health care providers in RCTs assessing non-pharmacological interventions. This classification scheme can be useful for investigators interested in developing and testing different methods of blinding, such as blinding participants to the study hypothesis. Blinding participants to the study hypothesis has been used when participants or their care providers cannot be blinded to the intervention. Participants can be given only partial information in the consent process. Of course, there may be ethical concerns about such methods depending on the context. A modified

Zelen design involving a two-step consent is another approach to blind participants to the study hypothesis (Boutron et al., 2007). As a first step, participants are asked to provide consent to take part in a cohort study (Quilty, Tucker, Campbell, & Dieppe, 2003). After completion of the cohort phases, blinded randomization occurs and only participants randomized to the experimental group are informed that they can receive an experimental intervention. They provide a second consent for this second phase if they wish to do so. The participants allocated to the control group are not informed about the intervention tested in the experimental group and continue to participate in the cohort.

The current CONSORT statement indicates that information concerning how the effectiveness of blinding was assessed is no longer required (Schulz et al., 2010). This is because when researchers assess effectiveness of blinding, most find that it is not successful, and the methods used to assess blinding may not be valid (Sackett, 2007). For example, Hrobjartsson et al.'s (2007) survey of RCTs indexed in the Cochrane Central Register of Controlled RCTs in 2001 found that it was rare for authors to report that they assessed the effectiveness of blinding. Less than half found that blinding was effective, and the most common method used to assess effectiveness was to ask participants to guess their group assignment. Most authors now concur that reporting needs to be improved by describing precisely what was done, to whom, and how (Boutron et al., 2008a; Hrobjartsson, Forfang, Haahr, Als-Nielsen, & Brorson, 2007; Sackett, 2007). Thus, the procedures to assess blinding should be put in place during the pilot study, and their feasibility examined and revised if needed for RCT. However, the assessment of the effectiveness of blinding is currently controversial until more valid methods are developed to do so.

Selection of outcome measures

The pilot study can be used to determine what construct is the best outcome to assess and with what measure (Conn, 2010; Lancaster et al., 2004). For example, the researcher can explore whether a specific or a generic measure of anxiety is more appropriate and sensitive to change. The clarity and acceptability of the measures for participants should also be explored to ensure that the measures chosen capture what the investigators intend them to capture, and that the participants are able to and comfortable providing the data. Test-retest reliability and internal consistency of responses can be assessed. Hertzog (2008) provides guidelines to determine the sample size required to evaluate the psychometric properties of a measure in a pilot study. Information about responsiveness (sensitivity to change) is needed when an outcome will be studied over time (Terwee, Dekker, Wiersinga, Prummel, & Bossuyt, 2003; de Vet & Beurskens, this volume).

Timing of outcome measurement

In nursing intervention studies, the researcher is interested in assessing whether the intervention can bring about the desired change. The anticipated full-scale RCT study design may involve repeated measures, longitudinal follow up, or trajectories as outcomes (Henly, Wyman, & Gaugler, 2011). When making design decisions, the researcher requires an understanding of change processes (Gottlieb & Feeley, 1999). They must understand: When will change occur? How long will it persist? Will there be any delayed effects? In the pilot study, questions concerning the timing of outcome measurements (i.e., duration, frequency), possible mediating variables (Conn et al., 2010), and whether intermediate intervention effects lead to longer-term outcomes, or if short-term outcomes persist (Craig et al., 2008) should be explored if applicable. These data can be particularly useful for justifying every data point in the full-scale study, and their timing. The

issue of whether the effect of the intervention on the outcomes should be examined in a pilot RCT is discussed later in this chapter in the section on current controversies.

Data collection

Many researchers are familiar with the notion that a pilot study should test and refine data collection methods and procedures. A multitude of procedures can be scrutinized including: the time required for data collection, the optimal sequence of measure administration, the feasibility of the data collection (i.e., can the data be collected as planned?), and the acceptability for the participants and clinicians if applicable (i.e., will the participants agree to the data collection procedures?). Assessment of these design aspects is vital to optimize data collection and minimize missing data, thus maintaining power in the full-scale RCT.

The current CONSORT statement for reporting RCTs of non-pharmacological interventions indicates that researchers need to describe any methods used to enhance the quality of their measurements, such as training of outcome assessors. Thus, the pilot study is also an opportunity to address issues, such as inter-rater reliability if applicable.

Participant retention

The timing, rate and reasons for attrition, as well as the relationship of attrition to participant characteristics and site should be studied (Conn et al., 2010). Inclusion and exclusion criteria can then be revised as needed to minimize attrition going forward. Questionnaires can be administered or interviews conducted to determine why pilot study participants drop out, and why they remain.

How effective are the methods for following up study participants? The pilot study also presents an opportunity to implement and assess the effectiveness of evidence-based strategies to retain study participants, in particular when the study design involves repeated measurement of outcomes over time.

Contamination and cross-over

Contamination occurs when participants in the comparison group receive some or all of the components of the experimental intervention (Sidani & Braden, 2011). In nursing intervention studies, participants may be in close proximity to one another, or able to interact, such as in waiting rooms or hospital. They may observe the intervention being provided to others, or they may have opportunities to share information about the intervention. In these ways, participants in the control group may receive the experimental intervention. Cross-over refers to situations where a participant assigned to the control group receives the experimental intervention; or a participant assigned to an intervention group receives instead the control condition or the intervention of another comparison group when more than one intervention is being evaluated (Friedman, Furberg, & DeMets, 2010). The extent of cross-over, reasons for it, and methods for minimizing it should be assessed in pilot studies.

What next? Decision-making following a pilot RCT

How does the researcher decide whether the pilot study findings indicate that methods are sufficiently tested to warrant the full-scale RCT? Unfortunately, there has been little discussion concerning how to arrive at this decision. We propose that prior to the pilot study the researcher

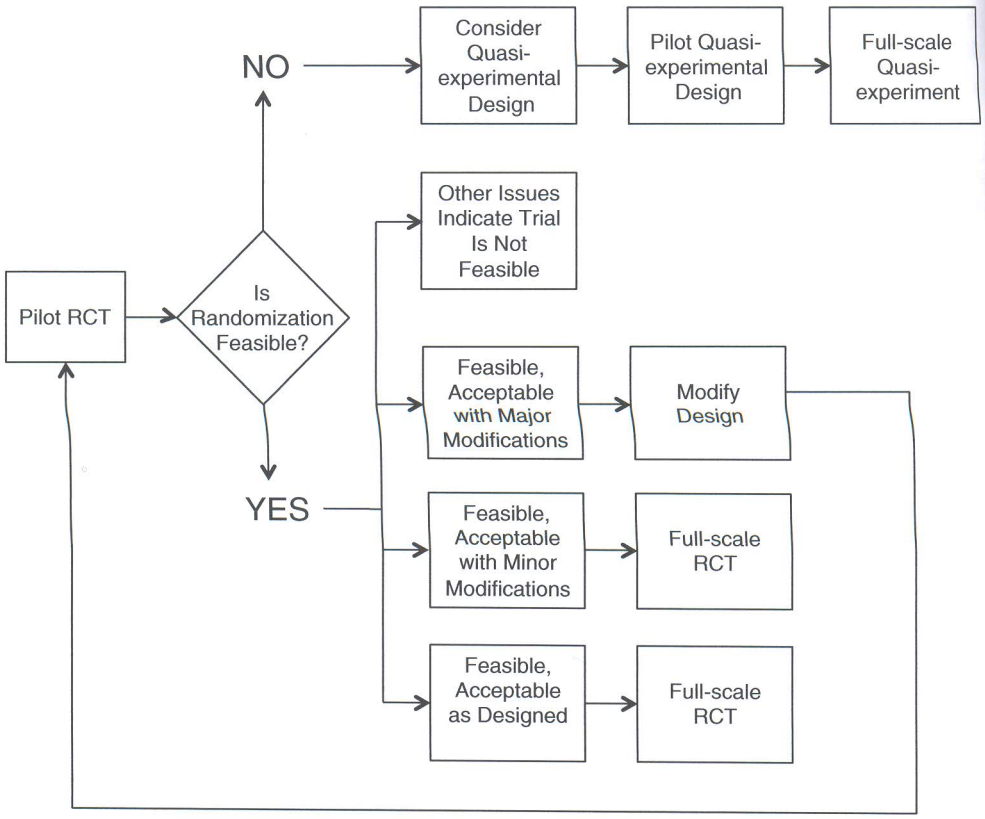
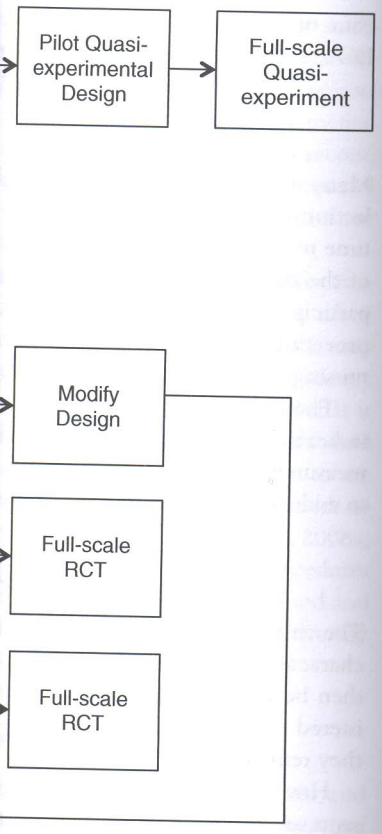


Figure 13.1 Decision-making process from pilot to full-scale RCT.

should identify specific indicators that will be utilized to answer each pilot study question. In Table 13.1 we outline some of the possible indicators for various pilot study questions. For example, if one of the questions is: Will participants find the data collection measures and schedule acceptable? Examples of indicators to answer this question could include: the percentage of measures completed as planned, responses to specific questions about the acceptability of the data collection procedures, and a record of any difficulties encountered. The analyses of the pilot data should then focus on these pre-specified indicators. An overall assessment of these data will guide the ultimate decision about whether to proceed to a full-scale study.

Figure 13.1 describes the various outcomes that might arise following a pilot RCT. If during the course of the pilot it becomes apparent that randomization of participants to group is not feasible or acceptable (i.e., eligible participants do not agree to random assignment or the setting is not suitable for simple randomization), then the researcher needs to consider other designs. These designs may include a cluster randomized trial where a group of patients or different settings (e.g., hospital unit) are randomized instead of individuals (Friedman et al., 2010). A preference trial design can permit participants not willing to be randomized to be included in an RCT. Participants with treatment preferences are given their desired group assignment; while those who do not have strong views are randomly assigned a group as they would be in an RCT (Torgerson & Sibbald, 1998). A historical control design could be employed when the benefit of a new intervention is almost demonstrated, but the effect size has yet to be determined. For



T.

nswer each pilot study question. In various pilot study questions. For data collection measures and sched- on could include: the percentage of s about the acceptability of the data tered. The analyses of the pilot data l assessment of these data will guide e study.

se following a pilot RCT. If during ion of participants to group is not o random assignment or the setting er needs to consider other designs. e a group of patients or different viduals (Friedman et al., 2010). A be randomized to be included in eir desired group assignment; while group as they would be in an RCT uld be employed when the benefit size has yet to be determined. For

instance, if a new intervention is almost ready to be transferred into clinical practice, a historical control design will permit the researcher to determine the effect of the intervention based on historical data (Friedman et al., 2010).

If randomization is possible, there are four possible outcomes. The first possibility is that the pilot findings indicate that a full-scale RCT is not feasible because of other serious feasibility or acceptability challenges that come to light. Another possibility is that major modifications are required because of problems such as contamination. Another pilot study should be conducted to assess the feasibility and acceptability of the revised study design. The new pilot does not necessarily have to be an RCT if the researcher wishes to assess questions other than random assignment. In many cases, the outcome will be to proceed with the full-scale RCT but with minor modifications to the study design or processes. The last possibility, which is less likely, is that the RCT proceeds but without any modifications as fine-tuning is inevitable.

Current controversies

Sample size for a pilot study

The question of how many participants are needed for a pilot study has received little attention in the literature. As researchers have had little guidance as to how to handle this issue, many pilot studies provide no justification (Shanyinde et al., 2011). Currently, given that the main objective is considered to be an assessment of feasibility and acceptability, a general guideline is that the pilot sample should be large enough to detect flaws in methodology, although factors such as time and budget may also come into play (Thabane et al., 2010). More specifically, several authors propose that the sample size required to answer one of the pilot study's main feasibility questions should be computed using well-known methods to construct confidence intervals around estimates (Arnold et al., 2009; Hertzog, 2008; Thabane et al., 2010). Using this approach, Hertzog (2008) illustrated how to determine sample size based on a pilot study's purpose, and provided guidelines for the sample size required for pilot studies with different purposes.

Estimating effect size in a pilot study

In the past, pilot studies were often used to provide an estimate of effect size prior to proceeding with the full-scale RCT. Conn et al. (2010) proposed that pilot studies should report effect sizes for the outcomes, and these effect sizes should be compared to those in the literature, and discussed in the context of the clinical significance. Because the number of participants in a pilot study is small, estimates of effect size based on these data are imprecise (Hertzog, 2008; Loscalzo, 2009). Thus, some now advocate that estimates obtained from pilot studies should not be used to compute power for the full-scale RCT (Sidani & Braden, 2011; Shanyinde et al., 2011), while others assert that these can be used but very cautiously (Arain, Campbell, Cooper, & Lancaster, 2010; Arnold et al., 2009; Hertzog, 2008; Loscalzo, 2009; Thabane et al., 2010). Hertzog (2008) recommended that researchers estimate confidence limits around a pilot study effect size that has been corrected for bias, and provides information on how to do so.

Most importantly, it is not advisable to draw conclusions about the efficacy of an intervention based on pilot data because sampling variability of the estimates is large. Premature conclusions based on pilot study data could lead to the unwarranted rejection of an intervention that may be efficacious (Loscalzo, 2009; Sidani & Braden, 2011). The opposite is also possible, a significant effect may be observed in the pilot study and not in the full-scale RCT.

Ethical issues

Ethical issues should be considered. CONSORT guidelines specify that any adverse or unintended effects be described in RCT reports (Schulz, Altman, Moher, & CONSORT Group, 2010). The pilot study should explore whether any potential unexpected harmful outcomes arise from the study procedures, such as participant distress. Investigators should also be alert to any indications of collateral harm, such as an increase or decrease in use of needed usual care services due to participation in an intervention study.

In the consent process, participants should be informed that the purpose of the study is to assess feasibility and acceptability questions, and not efficacy (Thabane et al., 2010). The researcher may need to develop an argument to justify the conduct of the pilot study to the Institutional Review Board, indicating that the study purpose is to assess feasibility and acceptability.

Registration and publication

Researchers have proposed that pilot RCTs should be registered in the same way as full-scale clinical RCTs (Arnold et al., 2009; Loscalzo, 2009). Pilot studies are being registered in the database of Current Controlled Trials (De Angelis et al., 2004). Consistent with what is reported in reviews cited earlier, the primary outcome included in these studies labeled as a "pilot study" ranged from feasibility and acceptability issues to a clinical endpoint with statistical testing of the effect of the intervention.

The subject of the publication of pilot studies is highly controversial. Much of the debate on this matter is clouded by the larger question of what is a pilot study, and what are the objectives of such studies. Conn et al. (2010) described the contributions that pilot study reports make to nursing knowledge, including providing information to alert other researchers to problems in methods and procedures and prevent them from unnecessarily conducting a similar study. Specific guidelines for publishing pilot studies based on the CONSORT guidelines have recently been proposed (Thabane et al., 2010). Given the purpose of the pilot study, publications should focus primarily on reporting these findings about feasibility and acceptability (Conn et al., 2010; Thabane et al., 2010).

Conclusion

Pilot studies are an important tool to inform the design and ensure the success of full-scale RCTs of nursing interventions. Thus, it is important that researchers learn to utilize these studies appropriately and effectively to ensure the development of the most rigorous RCTs to contribute to the knowledge needed to guide nursing practice.

Acknowledgment

Nancy Feeley was supported by a Research Scholar Award from the Fonds de recherche du Québec – Santé (FRQ-S).

References

Arain, M., Campbell, M. J., Cooper, C. L., & Lancaster, G. A. (2010). What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Medical Research Methodology, 10*(67). doi:10.1186/1471–2288–10–67

Arnold, D. M., Burns, K. E., Adhikari, N. K., Kho, M. E., Meade, M. O., Cook, D. J. for the McMaster Critical Care Interest Group. (2009). The design and interpretation of pilot RCTs in clinical research in critical care. *Critical Care Medicine, 37*(1 Suppl.), S69–S74.

Ballman, K. (2008, February). *Pilot study design issues.* Presentation at the Center for Health Trajectory Research Seminar Series, School of Nursing, University of Minnesota, Minneapolis. Retrieved from www.nursing.umn.edu/research/research-seminars/SeminarArchive/index.htm

Boutron, I., Guittet, L., Estellat, C., Moher, D., Hrobjartsson, A., & Ravaud, P. (2007). Reporting methods of blinding in randomized RCTs assessing nonpharmacological treatments. *PLoS Medicine, 4*(2), 0370–0380.

Boutron, I., Moher, D., Altman, D. G., Schulz, K. F., Ravaud, P., & CONSORT Group. (2008a). Extending the CONSORT statement to randomized RCTs of nonpharmacologic treatment: Explanation and elaboration. *Annals of Internal Medicine, 148*, 295–309.

Boutron, I., Moher, D., Altman, D. G., Schulz, K. F., Ravaud, P., & CONSORT Group. (2008b). Methods and processes of the CONSORT Group: Example of an extension for RCTs assessing nonpharmacologic treatments. *Annals of Internal Medicine, 148*, W60–W66.

Brown, S. J. (2002). Nursing intervention studies: A descriptive analysis of issues important to clinicians. *Research in Nursing & Health, 25*, 317–327.

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. *BMJ, 321*, 694–696.

Conn, V. S. (2010). Rehearsing for the show: The role of pilot study reports for developing nursing science. *Western Journal of Nursing Research, 32*, 991–993.

Conn, V. S., Algase, D. L., Rawl, S. M., Zerwic, J. J., & Wyman, J. F. (2010). Publishing pilot intervention work. *Western Journal of Nursing Research, 32*, 994–1010.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ, 337*, a1655.

De Angelis, C., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., . . . Van Der Weyden, M. B. (2004). Clinical trials registration: A statement from the International Committee of Medical Journal Editors [Editorial]. Retrieved from http://www.icmjeorg/clin_trial.pdf

Feeley, N., Cossette, S., Côté, J., Héon, M., Stremler, R., Martorella, G., & Purden, M. (2009). The importance of piloting an RCT intervention. *Canadian Journal of Nursing Research, 41*, 84–99.

Feeley, N., Zelkowitz, P., Charbonneau, L., Cormier, C., Lacroix, A., Ste. Marie, C., & Papageorgiou, A. (2008). Assessing the feasibility and acceptability of an intervention to reduce anxiety and enhance sensitivity among mothers of very low birth-weight infants. *Advances in Neonatal Care, 8*, 276–284.

Feeley, N., Zelkowitz, P., Shrier, I., Stremler, R., Westreich, R., Dunkley, D., . . . Papageorgiou, A. (2012). Follow up of the Cues and Care randomized controlled trial: Mother and infant outcomes at 6 months. *Journal of Early Intervention, 34*(2), 65–81.

Friedman, L. M., Furberg, C. D., & DeMets, D. L. (2010). *Fundamentals of clinical trials* (4th ed.). New York, NY: Springer.

Gottlieb, L. N., & Feeley, N. (1999). Nursing intervention studies: Issues related to change and timing in children and families. *Canadian Journal of Nursing Research, 30*, 193–212.

Henly, S. J., Wyman, J. F., & Gaugler, J. E. (2011). Health trajectory research: A call to action for nursing science. *Nursing Research, 60*(3 Suppl.), S79–S82.

Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. *Research in Nursing & Health, 31*, 180–191.

Hrobjartsson, A., Forfang, E., Haahr, M. T., Als-Nielsen, B., & Brorson, S. (2007). Blinded RCTs taken to the test: An analysis of randomized clinical RCTs that report tests for the success of blinding. *International Journal of Epidemiology, 36*, 654–663.

Lackey, N. R., & Wingate, A. L. (1998). The pilot study: One key to research success. In P. J. Brink & M. J. Wood (Eds.), *Advanced design in nursing research* (2nd ed.,) (pp. 375–386). Thousand Oaks, CA: Sage.

Lancaster, G. A., Dodd, S., & Williamson, P. R. (2004). Design and analysis of pilot studies: Recommendations for good practice. *Journal of Evaluation in Clinical Practice, 10*, 307–312.

Loscalzo, J. (2009). Pilot RCTs in clinical research: Of what value are they? *Circulation, 119*, 1694–1696.

Maxwell, S. E., & Delaney, H. D. (2003). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Florence, KY: Taylor & Francis.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gotzsche, P. C., Devereaux, P. J., . . . Consolidated Standards of Reporting Trials Groups. (2010). CONSORT 2010 Explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology, 63*(8), e1–37.

Polit, D. F., & Beck, C. T. (2012). *Nursing research: Generating and assessing evidence for nursing practice* (9th ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Polit, D. F., & Gillespie, B. M. (2010). Intention-to-treat in randomized controlled trials: Recommendations for a total trial strategy. *Research in Nursing and Health, 33*, 355–368.

Quilty, B., Tucker, M., Campbell, R., & Dieppe, P. (2003). Physiotherapy, including quadriceps exercises and patellar taping, for knee osteoarthritis with predominant patello-femoral joint involvement: Randomized controlled trial. *Journal of Rheumatology, 30*, 1311–1317.

Rothwell, P. M. (2006). Factors that can affect the external validity of randomized controlled trials. *PLoS Clinical Trials, 1*(1), e9.

Sackett, D. L. (2007). Commentary: Measuring the success of blinding in RCTs: Don't, must, can't or needn't? *International Journal of Epidemiology, 36*, 664–665.

Schulz, K. F., Altman, D. G., Moher, D., & CONSORT Group. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ, 340*, c332.

Shanyinde, M., Pickering, R. M., & Weatherall, M. (2011). Questions asked and answered in pilot and feasibility randomized controlled RCTs. *BMC Medical Research Methodology, 11*(117). doi:10.1186/1471–2288–11–117

Sidani, S., & Braden, C. J. (2011). *Design, evaluation, and translation of nursing interventions*. Chichester, England: Wiley-Blackwell.

Terwee, C. B., Dekker, F. W., Wiersinga, W. M., Prummel, M. F., & Bossuyt, P. M. M. (2003). On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. *Quality of Life Research, 12*, 349–362.

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., . . . Goldsmith, C. H. (2010). A tutorial on pilot studies: The what, why and how. *BMC Medical Research Methodology, 10*(1). doi:10.1186/1471–2288–10–1

Torgerson, D., & Sibbald, B. (1998). Understanding controlled trials: What is a patient preference trial? *British Medical Journal, 316*, 360.

Whittemore, R., & Grey, M. (2002). The systematic development of nursing interventions. *Journal of Nursing Scholarship, 34*, 115–120.

CAUSAL
AND OBS

Dona

Inference for causal effects is
The field of statistics is critica
experiments and observationa
standard for inferring causal ef
studies, requires great care. Her
causal effects in both random
inferences in both settings. Th
can dramatically differ in prac
However, we use well-accepte
design and analysis of observa
clinical studies.

Randomized experiments, ofte
sciences, are commonly used
placebo control. The simplest
sciences context) has a known
RCTs, investigators are often b
often also blinded to the assign
the opportunity to observe ou
blood pressure outcome one ye
protocols, with the design of t
journals in the health sciences
trial begins in order for results
the study. Many of these journa
analysis plan, be submitted wit
Here we denote the observed
strata indicators for male versu

A critical feature of random
group are balanced on measu