



SOJNR

SOUTHERN ONLINE JOURNAL OF NURSING RESEARCH

Volume 9 – Number 3

www.snrs.org

Data Collection Order: A Primer

T. Deshefy-Longhi, DNSc, RN
Postdoctoral Fellow
Duke University Center for Aging and Human Development
and School of Nursing
Durham, North Carolina

S. Sullivan-Bolyai, DNSc
Associate Professor of Nursing
UMASS School of Nursing
Worcester, MA;

J. K. Dixon, PhD
Professor of Nursing
Yale University School of Nursing
New Haven, CT

Corresponding Author:

Tess Deshefy-Longhi, DNSc, RN
Duke University Center for the Study of Aging and Human Development
Clipp Research Affairs Building , Room 1030
DUMC Box 3322
Durham , NC 27710
email: td40@duke.edu

Partial support for Dr. Deshefy-Longhi is provided by NIH Grant T32 AG000029

Partial support for Dr. Sullivan-Bolyai is provided by NIH-NINR Grant 1 R15 NR008391-01

Data Collection Order: A Primer

Introduction

Assessing the validity of research data and replicating the research study are critical steps towards clinicians' confidence in and application of research findings. An important methods consideration—data collection order—affects both validation and replication of research data. Whenever using two or more measures, or methods (i.e. qualitative and quantitative), for data collection, researchers must carefully consider and document the order in which data are collected.¹ Yet, a search of the literature indicates that few researchers indicate the order in which their data were collected, and even fewer explain their rationale behind the order selected. Neophyte researchers may not be fully aware of the issues involved in data collection order, while experienced researchers may argue that data collection order considerations are intuitive, eliminating the need to document the order used or the rationale for it. Clearly each data collection order choice involves an informed decision regarding its pros and cons, and there is no single correct answer for every research situation.

Therefore the process of how researchers determine their specific data collection order, whether intuitive or not, needs to be addressed. The main purpose of this methodological article is to describe issues to consider in determining data collection order. We also will examine potential threats to validity that may arise relative to the order selected. We intend it to be a primer in data collection order decisions for neophyte researchers, and a review and reminder for more experienced researchers. First we will present issues to consider. This will be followed by a brief discussion of the specific threats to validity that data collection order may engender. We emphasize examples of mixed methods studies, as the more diverse the data collection approaches are, the more the order of collection may matter. Furthermore, this genre of research has been least discussed in relation to data collection order. For the purposes of this article, mixed methods research is defined as the integration of quantitative and qualitative paradigms to explore the research aims.²

Issues to Consider Related to Data Collection Order

Regardless of the type of study a researcher proposes, one must carefully consider the implications of data collection order with respect to 5 issues. These include study purpose, study participants, data content, study logistics, and data collection procedures.

Study purpose

The study purpose should be considered first when determining data collection order in planning a mixed methods study. If the purpose is to develop an instrument that accurately captures the essence of a targeted concept, early participant information is critical; focus group or individual interview data may be collected before the instrument is developed. Conversely, if a researcher wants subject input regarding participation in a survey, the reverse order may be selected—survey administered first followed by focus groups or individual

interviews using a subset of original subjects—to provide context for the quantitative data findings. Habermann and Davis³ administered the Caregiver Assistance Measure (CAM) to elder caregiving spouses of Alzheimer and Parkinson patients. This written survey was followed by interviews with these caregivers. The authors stated that this order was selected to see if resources noted on the CAM by caregivers related to caregiving challenges described in their qualitative interviews. Another example in which the study purpose might determine the data collection order is the description of an elusive issue encountered in clinical practice. For instance, Randall & Barroso⁴ used a grounded theory approach to explore why some HIV-positive adults do not enter into the health care system for routine care and follow up of their condition. Quantitative data were collected first on all subjects in a longitudinal study on HIV; then, using extreme case sampling, the authors selected four subjects who had no designated health care provider or clinic monitor their HIV condition, for further in-depth interviews about this issue.

Study Participants

A second data collection order consideration is the general characteristics of the study sample. Populations with whom data collection may be challenging deserve special attention. These populations include elders and individuals with reduced energy or increased stress due to illness, or vulnerability, as well as children and parents with young children. Data collection sessions should be of reasonable length relative to the sample, and splitting the data collection into two sessions, should sometimes be considered. When determining order, the following questions should be asked about potential participants: 1) Will participants be able to complete all types of data collection methods in one session? 2) Is there potential for some participants to decide not to continue participating in the study? 3) If so, which data collection method best answers the primary aim of the study so as to maximize limited participation? For example, in a study on nonverbal communication in elders with Parkinson disease⁵ observation and filming of the communication task were done first before participants completed several relationship satisfaction questionnaires. This enabled the researcher to do the quantitative component (self-administered questionnaires) at a second visit if elder participants became too fatigued to continue after completing the communication task.

Data Content

Content of data may be sensitive, self-revealing, or culturally specific. Data also may be difficult for participants to recall such as dietary intake. The following are examples of data content considerations that may affect data collection order:

Sensitive or self-revealing data. A researcher may require data that are considered sensitive or self-revealing by study participants. For example, Hughes et al.⁶ described teenage drinking habits in Scotland. Hughes decided to use two

separate samples (non-overlapping subjects) for her two methods. The researchers sampled a large adolescent population using a self-report questionnaire, and also recruited a separate, smaller sample for focus groups. An alternative approach, to capitalize on the self-revealing nature of the data, might have been to use an overlapping sample—same subjects used for both methods—in which the self-report questionnaires were administered first to the participating adolescents. Based on the results, adolescents who self-reported abuse of alcohol could have comprised one focus group, and those who did not report such abuse would comprise the second focus group. This data collection order would have produced more homogeneous focus groups that, in turn, may have reduced potential bias of the group effect on the individual.

Culturally/Regionally specific data. A researcher may wish to gather information from a particular cultural or regional group using qualitative interviews and a well-validated instrument but is uncertain about the instrument's potential biases for that group or region. If the researcher does the qualitative phase of the study first, prior to using the quantitative instrument, the qualitative data may provide insight into the tool's degree of cultural/regional sensitivity. The researcher could then use the qualitative findings to modify the tool accordingly.

Information difficult to recall. The content of the data may also be difficult to recall for study participants. Examples of such content may include elder recall of distant events, or adult recall of daily routine activities such as meals. One needs to consider if administration of a questionnaire first, prior to a qualitative interview, might assist in cueing participants. Alternatively, in some situations, a researcher may use a qualitative method to facilitate recall prior to administering a self-report questionnaire. In a study to explore dietary intake of women with Type II diabetes^{7,8} participants did a 48-hour dietary intake recall first before completing a self-administered survey on their food habits for the previous month. In both situations, one must consider if data collection order potentially biases either the quantitative or qualitative responses and then determine ways to address any potential bias. A simple query asking participants if they felt that the data collection order could have influenced their responses may suffice.

Study Logistics

Another consideration for data collection order is the logistics of implementing the study. The researcher must address the limitations imposed by the study site by determining a practical sequence, flow and timing for the data collection. For instance, one must consider the feasibility of adequate space for completing questionnaires, and for conducting qualitative interviews or discussion groups that allow ongoing privacy without interruptions. If there is no private space for conducting interviews where recruitment is being done (such as a clinic setting), it may be possible to administer the quantitative measures of the study, followed by qualitative interviews arranged at a later time outside the clinic setting. This sequence also could provide an opportunity for a beginning relationship between

the researcher and the study participants prior to a possible home visit interview. However, in using different locations for each data collection method, one must consider the potential for change in setting (or passage of time between sessions) to introduce potential threats to validity, as will be discussed below.

Data Collection Procedures

Once a data collection order is determined, the researcher should make it a standardized part of the data collection process, so that order of data collection is the same for all study participants. There are, however, two potential exceptions to this important principle. The first is systematic variation in order to explore the effects of order. If the researcher is uncertain regarding potential bias due to data collection order, the order may be systematically varied, as by randomly assigning participants to one of two (or more) data collection orders. Floyd⁹ effectively modeled this strategy. She conducted a mixed methods study exploring perceived problems of sleep patterns in older adults. To test for potential biases related to the order of data collection, Floyd randomly assigned each subject to one of two groups. One group did qualitative interviews first followed by self-report questionnaires; the second group provided information in the reverse order. Floyd found that completing the self-report instruments prior to the qualitative interviews influenced the language chosen by the elders in describing their sleep concerns. However, administering the qualitative interviews first did not affect participants' questionnaire responses. Floyd concluded that if her study were replicated, the qualitative data collection methods should precede the quantitative methods. Her findings underscore the effect of one method biasing or influencing participants' responses to a second method.

The second exception involves the need for flexibility in a particular population such as frail elders, severely ill adults, or preschool-aged children. Allowing for flexibility may be paramount for the caregivers of these vulnerable populations. Sullivan-Bolyai and colleagues¹⁰ conducted a mixed methods study comprised of parental interviews, mother-child observations, and a battery of questionnaires with mothers and their young children with Type I diabetes. Consideration was given to the child's mood, medical condition, and potential for hypoglycemia or hyperglycemia during the study visit. It was determined that starting with the interviews first was the best approach to address the study aims, to reduce the influence of the questionnaires on the qualitative responses, and to allow for flexibility of the observations according to the mother's assessment of her child's needs. Similar considerations might be required with other vulnerable populations such as frail or severely ill participants.

Threats to Internal Validity Related to Data Collection Order

Internal validity is the degree to which we can be confident of our interpretation of a study's result—the extent to which we can confidently infer that the dependent variable in a study is actually influenced by the study's independent variable(s)

and not by other, uncontrolled factors or events. Shadish, Cook and Campbell¹¹ cite nine common sources of threats to the internal validity of a study. Three of these stand out as potentially impacted by data collection order, especially when two types of data are collected from the same participants. These threats are testing, history, and maturation. Additionally, these threats may occur in combination.

Testing

Testing is a threat to internal validity concerning the possibility that the effects of taking a pre-test may affect the participants' responses on a post-test. This threat may also apply when two different types of data are collected. Thus it must be carefully considered when planning the data collection order in mixed methods studies. As noted previously, Floyd⁹ reported differences in data due to order of data collection when studying elders and their self-perceived fatigue. More recently, Barroso & Sandelowski¹² explored HIV-related fatigue in adults and chose to begin the study with qualitative interviews prior to using self-report questionnaires. They purposefully chose this order to ensure that participants' discussions of fatigue would not be influenced by items in the scale and inventory they used.

History

History, defined as external events that occur during a study that could affect or influence the study findings,¹¹ could potentially be introduced by data collection order when using mixed methods. There may be a time lag between one method and another being administered, during which an external event might alter the participants' responses. If the order of data collection in the mixed methods study is randomized, then history should not be a threat because it is assumed that there are equal chances of the participants in the different randomized groups being affected. A comparison of these groups should help the researcher identify if a history effect occurred. However, if data collection order is not randomized, outside events, such as publicized changes in treatments, equipment recalls, and educational news, can alter responses from one method to the next. Timing the data collection methods to occur as closely as possible to each other can minimize this potential threat. Another strategy would be to include a question at the end of the study about historical events and how they might have influenced the participants' responses.

Maturation

Maturation refers to the occurrence of natural changes, such as aging. It also may refer to becoming more attuned to or experienced with the study's topic over the course of the study. This threat is often seen in the form of participant boredom or fatigue that is generated by the use of multiple study measures. Last-administered measures, whether qualitative or quantitative, may not receive

participants' full efforts or concentration. Determining the amount of time each measure will require may help the researcher address threats of maturation. For instance, a researcher may include a short snack break between an interview and a battery of questionnaires to renew participant energy level as well as build participant-researcher rapport.

Combination of Internal Validity Threats

Shadish et al.¹¹ point out that threats to internal validity may occur in combination with each other, thereby adding to (or even multiplying) the overall effect. For instance, a participant's exposure to the responses of others may be more influential in the case of a subject who is also feeling fatigue in the context of a lengthy session of data collection. Again, vulnerable populations such as frail elders and children may be especially susceptible to such combined threats. Critically thinking about all of the earlier-mentioned considerations of data collection order should help to minimize these combination effects.

Conclusion

This article presents issues of data collection order that must be considered during the planning and implementation phases of studies. The researcher can achieve consistency in data collection order decisions by considering order in relation to study purpose, study participants, content of the data, study logistics, data collection procedures, and potential threats to internal validity. Through thoughtful consideration and explication of these order decisions, researchers can maximize the validity of their data and allow for easier replication in future studies. Researchers are challenged to account for increasing details in decreasing printed space. Nonetheless, to get the full benefit from the dissemination of studies, data collection order must be carefully addressed and documented.

References

1. Morse, J.M. (2005). Evolving trends in qualitative research: Advances in mixed-method design. *Qualitative Health Research*, **15**: pp. 583-585.
2. Tashakkori, A. and Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Vol. 45. Thousand Oaks, California: SAGE Publications.
3. Habermann, B. and Davis, L.L. (2005). Caring for family with Alzheimer's disease and Parkinson's disease: Needs, challenges, and satisfactions. *Journal of Gerontological Nursing*, **31**(6): pp. 49-54.
4. Randall, M. and Barroso, J. (2002). Delayed pursuit of health care among HIV-positive gay men enrolled in a longitudinal research study. *JANAC: Journal of the Association of Nurses in AIDS Care*, **13**(4): pp. 23-31.

5. Deshefy-Longhi, T. (2008), Nonverbal Communication Accuracy in Elder Couples Living with Moderate to Advanced Parkinson Disease, in *School of Nursing*. Yale: New Haven. pp. 162.
6. Hughes, K., MacKintosh, A.M., Hastings, G., et al. (1997). Young people, alcohol, and designer drinks: quantitative and qualitative study. *BMJ*, **314**(7078): pp. 414-8.
7. Vorderstrausse, A.A., Melkus, G., Chyun, D., et al. (2009). "It's like trying to stop smoking": Psychosocial factors associated with dietary intake in Black women with Type 2 Diabetes, a mixed methods approach *RINAH*, **(in review)**.
8. Amend, A., Melkus, G.D., Chyun, D.A., et al. (2007). Validation of Dietary Intake Data in Black Women with Type 2 Diabetes. *Journal of the American Dietetic Association*, **107**(1): pp. 112-117.
9. Floyd, J.A. (1993). The use of a cross-method triangulation in the study of sleep concerns in healthy older adults. *ANS - Advances in Nursing Science*, **16**(2): pp. 70-80.
10. Sullivan-Bolyai, S., Deatrck, J., Gruppuso, P., et al. (2002). Mothers' experiences raising young children with Type 1 diabetes. *Journal for Specialists in Pediatric Nursing*, **7**: pp. 93-103.
11. Shadish, W.R., Cook, T.D., and Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.
12. Barroso, J. and Sandelowski, M. (2001). In the field with the Beck Depression Inventory. *Qualitative Health Research*, **11**(4): pp. 491-504.