Designing and Using Surveys in Nursing Research: A Contemporary Discussion

Siobhan O'Connor, BSc, CIMA CBA, BSc, RGN, PhD[1] (iD)

Abstract

This commentary summarizes the contemporary design and use of surveys or questionnaires in nursing science, particularly in light of recent reporting guidelines to standardize and improve the quality of survey studies in healthcare research. The benefits, risks, and limitations of these types of data collection tools are also briefly discussed.

Keywords

healthcare, nursing, questionnaire, survey, research

The use of research questionnaires or surveys in nursing is a long standing tradition, dating back to the 1960s (Logan, 1966) and 1970s (Oberst, 1978), when the scientific discipline emerged. This type of tool enables nursing researchers to gather primary data from a specific population, whether it is patients, carers, nurses, or other stakeholders to address gaps in the existing evidence base of a particular clinical, pedagogical, or policy area. However, the recent creation of a checklist for reporting survey studies called CROSS: Consensus-Based Checklist for Reporting of Survey Studies, hints at problems in their design, development, administration, and reporting (Sharma et al., 2021). This commentary discussion focuses on the process of developing, validating, and administering surveys in nursing research and some ways to strengthen this methodological approach.

Ideally, surveys should be constructed to gather the minimum amount of information from respondents to provide good quality data about a problem or phenomenon. Gathering large amounts of unnecessary data may complicate a survey, leading to a low response rate and weak findings. Therefore, time and expertise is needed when designing research surveys (Robb & Shellenbarger, 2020). Firstly, existing evidence should be reviewed to identify if an existing survey could be utilized or refined. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN), and associated database and critical appraisal checklist, could be employed to examine the psychometric properties of an established tool and its methodological quality before use (Mokkink et al., 2016). For instance, Charette et al. (2020) followed the COSMIN approach when conducting a systematic review of the psychometric properties of scales that assessed new nurses clinical competence.

If a new instrument needs to be developed, then reviewing relevant literature could help inform what should be measured for example, what people know, think, feel, or do, along with guiding the content of specific survey questions (Polit & Beck, 2020). Other techniques can be employed to create questions including a Delphi study to gather expert opinion (Bender et al., 2018), or focus groups with patients, clinicians, educators, students, or policy makers (Tajik et al., 2010). Decisions about the style of survey questions also needs consideration as each brings advantages and disadvantages. An open question gives a respondent free reign with their answer which could uncover fresh insights on a topic. However, it may contribute to respondent fatigue if too many are asked, and the data can be time consuming to analyze (O'Cathain & Thomas, 2004).

How each question is worded is also important to avoid leading, composite, or presumptive questions, ones that are vague, overly lengthy and complex, or include double negatives, jargon, or terminology unfamiliar to the reader, so that what is being asked and answered is clear and consistent. The sequence of questions should also be logical, opening with more general non-threatening questions, followed by more specific ones that can be grouped or filtered accordingly, and closing with socio-demographic variables and a thank you (Boynton & Greenhalgh, 2004). Closed fixed choice questions can be formulated in a number of ways including checklists, frequency or Likert-type scales,

[1]National University of Ireland Galway, Ireland

Corresponding Author:
Siobhan O'Connor, School of Nursing and Midwifery, National University of Ireland Galway, Aras Moyola, Upper Newcastle, Galway, Ireland.
Email: s.oconnor@nuigalway.ie

Guttman or cumulative scales, Thurstone scales, and rankings, which vary in their content, structure, and layout, and require either a dichotomous or multiple-choice response. The sensitivity of any measurement scale is important to ensure it accurately represents the respondents answer and reduces the risk of bias (Polit, 2014). Hence, piloting a draft survey with a small sample of intended respondents can help identify problems with ambiguous content, the format of questions, or confusing instructions or layouts.

The validity and reliability of a survey instrument should also be established to demonstrate the questions are worded appropriately and illicit accurate answers. Validity is about accuracy, in terms of how well a survey measures what it is supposed to. It can be assessed in three ways: (1) face or content validity, (2) construct validity, and (3) criterion validity (Rattray & Jones, 2007). Content validity looks at comprehensiveness, and whether questions adequately measure all aspects of a subject matter. For example, Devriendt et al. (2012) examined the content validity of the Safety Attitudes Questionnaire through expert clinician review and using the content validity index and a modified kappa index. Construct validity focuses on whether the concept(s) that underpin the questions in a survey correspond with contemporary theory and scientific knowledge. For instance, McSherry et al. (2002) employed factor analysis to determine construct validity for a Spirituality and Spiritual Care Rating Scale. Numerous research studies are often required to evaluate and refine the construct validity of a survey instrument to ensure it is robust. Some go further and investigate both convergent and discriminate validity, the two sub-types of construct validity (Hallett et al., 2018; Zhao et al., 2020)

Criterion validity refers to how much the scores in a survey measure agree with a gold-standard that is considered an ideal measure of the constructs or variables. This approach is not always feasible, if there are no reliable measures for independent comparison (Polit & Beck, 2020). It can be done in two ways, the first by calculating a correlation coefficient which tests the strength of the association (not agreement) between the results from a survey and an external independent measure. Secondly, sensitivity and specificity can be calculated, although there is usually a trade-off between the two (Groves, 2009). Sensitivity measures the ability of a survey to detect all instances of its subject matter (true positives), while specificity measures the ability of a survey to discriminate all instances of its subject matter from those which are not related (true negatives). Both false negatives and false positive errors may occur, so the nature of the research and survey instrument should guide which type of error should be minimized as much as possible (Dillman, 2014). The two types of criterion validity, concurrent and predictive validity, can also be measured. Concurrent validity compares survey questions or items to a related validated measure, both of which are assessed at the same time, whereas predictive validity compares survey items against some criterion measure at a later time (Kim & Abraham,

2016). While validity testing can be time consuming, expensive, and require a significant amount of statistical expertise, it is a robust way to develop and improve the psychometric properties of surveys.

The other major concept used to evaluate the quality of surveys is reliability which focuses on the consistency of a survey and its items, to ensure it would give the same results if repeated under the same conditions (Rattray & Jones, 2007). The three kinds of reliability testing are: (1) test retest, (2) inter-rater, and (3) internal consistency. Test retest looks at consistency of a measure across time and whether survey results from the same person were the same on at least two or more occasions. This can be measured using a number of statistical techniques such as the intraclass correlation coefficient and Wilcoxon signed rank test (Lovén Wickman et al., 2019). Inter-rater reliability examines the consistency of a measure across raters or observers to determine if a person scores items in a survey in the same way multiple times. Cohen's kappa (Dancey et al., 2012) and the intra-class correlation coefficient (Ryu et al., 2013) are common statistical measures for this. Finally, internal consistency is how consistently respondents' answer items in a survey, if pairs of questions measuring the same concept are asked in different ways which can be calculated using Chronbach's alpha (Paans et al., 2010). Although reliability may be established and the survey results reproducible, this does not mean they are valid and may be incorrect unless the instruments' validity is also determined.

Once a survey is designed, it then needs to be administered to the appropriate population, once ethical approval is granted. Self-completing surveys are the most common as they can quickly and easily be given to a large population using online, electronic or paper methods which are affordable options. An interview administered survey is an alternative, where questions are answered in the presence of a researcher, if sensitive topics need to be discussed, if vulnerable populations need to be reached, or if a survey is long and complex. A Cochrane review of interventions for administering postal and electronic questionnaires reported several strategies such as utilizing stamped addressed envelopes, financial incentives, and personalized communications, were effective in increasing the response rate (Edwards et al., 2009), as low response rates can negatively impact the results of survey studies. After data is gathered, verified, cleaned, and anonymized, it needs to be coded using suitable analyses. Epi Info™ (https://www.cdc.gov/epiinfo/index.html) is a popular tool for entering and storing survey data before exporting it to a statistical analysis package (Kebede et al., 2017).

Finally, surveys are frequently published in scientific nursing journals. However, Sharma et al. (2021) highlighted the substantial variability and inconsistency in how research surveys are reported which can weaken the evidence base on a topic. They emphasized that despite two guidelines for reporting non-web and web-based surveys, SURGE

(Grimshaw, 2014) and CHERRIES (Eysenbach, 2004), improvements in the reporting of survey research have not materialized and these tools have limitations. Hence, a new comprehensive survey reporting guideline called CROSS was developed to enhance the transparency and replicability of survey based health research (Sharma et al., 2021). This new guideline should be used in nursing to enable survey studies to be better designed, conducted, and reported. By undertaking rigorous, high-quality surveys, researchers can advance nursing science, strengthen the evidence base on which nurses practice, and help make a positive impact on patient care and health service delivery.

Author Contributions

The sole author drafted and wrote the manuscript.

Declaration of Conflicting Interests

Funding

ORCID iD

Siobhan O'Connor (iD) https://orcid.org/0000-0001-8579-1718

References

Bender, M., Avolio, A., Baker, P., Harris, J. L., Hilton, N., Hites, L., Roussel, L., Shirley, B., Thomas, P. L., & Williams, M. (2018). Developing the clinical nurse leader survey instrument: A modified Delphi study. *Journal of Nursing Care Quality*, *33*(4), 300–308.

Boynton, P. M., & Greenhalgh, T. (2004). Selecting, designing, and developing your questionnaire. *BMJ*, *328*(7451), 1312–1315. https://doi.org/10.1136/bmj.328.7451.1312

Charette, M., McKenna, L. G., Maheu-Cadotte, M.-A., Deschênes, M.-F., Ha, L., & Merisier, S. (2020). Measurement properties of scales assessing new graduate nurses' clinical competence: A systematic review of psychometric properties. *International Journal of Nursing Studies*, *110*, 103734–103734. https://doi.org/10.1016/j.ijnurstu.2020.103734

Dancey, C. P., Reidy, J., & Rowe, R. (2012). *Statistics for the health sciences: A non-mathematical introduction* (1st ed.). SAGE Publications Ltd.

Devriendt, E., Van Den Heede, K., Coussement, J., Dejaeger, E., Surmont, K., Heylen, D., Schwendimann, R., Sexton, B., Wellens, N. I., Boonen, S., & Milisen, K. (2012). Content validity and internal consistency of the Dutch translation of the Safety Attitudes Questionnaire: An observational study. *International Journal of Nursing Studies*, *49*(3), 327–337. https://doi.org/10.1016/j.ijnurstu.2011.10.002

Dillman, D. A. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Wiley.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., Kwan, I., & Cooper, R. (2009). Methods to increase response rates to postal questionnaires. *Cochrane Database of Systematic Reviews*, *3*, MR000008. https://doi.org/10.1002/14651858.MR000008.pub4

Eysenbach, G. (2004). Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *Journal of Medical Internet Research*, *6*(3), e34–e34. https://doi.org/10.2196/jmir.6.3.e34

Grimshaw, J. (2014). SURGE (The SUrvey Reporting GuidelinE). In D. Moher, D. Altman, G. Schulz, K. F. Simera & I. E. Wager. (Eds.), *Guidelines for reporting health research: A user's manual* (pp. 206–213). Wiley Blackwell.

Groves, R. M. (2009). *Survey methodology* (2nd ed.). Wiley.

Hallett, N., Huber, J., Sixsmith, J., & Dickens, G. L. (2018). Measuring the violence prevention climate: Development and evaluation of the VPC-14. *International Journal of Nursing Studies*, *88*, 97–103. https://doi.org/10.1016/j.ijnurstu.2018.09.002

Kebede, M., Endris, Y., & Zegeye, D. T. (2017). Nursing care documentation practice: The unfinished task of nursing care in the University of Gondar Hospital. *Informatics for Health and Social Care*, *42*(3), 290–302. https://doi.org/10.1080/17538157.2016.1252766

Kim, H.-J., & Abraham, I. (2016). Psychometric comparison of single-item, short, and comprehensive depression screening measures in Korean young adults. *International Journal of Nursing Studies*, *56*, 71–80. https://doi.org/10.1016/j.ijnurstu.2015.12.003

Logan, W. M. (1966). The shortage of nurse educators–report of a research project carried out in Scotland. *International Journal of Nursing Studies*, *2*(4), 283–292. https://doi.org/10.1016/0020-7489(66)90003-4

Lovén Wickman, U., Yngman-Uhlin, P., Hjortswang, H., Wenemark, M., Stjernman, H., Riegel, B., & Hollman Frisman, G. (2019). Development of a self-care questionnaire for clinical assessment of self-care in patients with inflammatory bowel disease: A psychometric evaluation. *International Journal of Nursing Studies*, *89*, 1–7. https://doi.org/10.1016/j.ijnurstu.2018.08.016

McSherry, W., Draper, P., & Kendrick, D. (2002). The construct validity of a rating scale designed to assess spirituality and spiritual care. *International Journal of Nursing Studies*, *39*(7), 723–734. https://doi.org/10.1016/S0020-7489(02)00014-7

Mokkink, L. B., Prinsen, C. A. C., Bouter, L. M., Vet, H. C. W. D., & Terwee, C. B. (2016). The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian Journal of Physical Therapy*, *20*(2), 105–113. https://doi.org/10.1590/bjpt-rbf.2014.0143

O'Cathain, A., & Thomas, K. J. (2004). "Any other comments?" Open questions on questionnaires - a bane or a bonus to research? *BMC Medical Research Methodology*, *4*(1), 25–25. https://doi.org/10.1186/1471-2288-4-25

Oberst, M. T. (1978). Priorities in cancer nursing research. *Cancer Nursing*, *1*(4), 281–290. https://doi.org/10.1097/00002820-197808000-00002

Paans, W., Sermeus, W., Nieweg, R. M., & van der Schans, C. P. (2010). D-Catch instrument: Development and psychometric testing of a measurement instrument for nursing documentation in hospitals. *Journal of Advanced Nursing*, *66*(6), 1388–1400. https://doi.org/10.1111/j.1365-2648.2010.05302.x

Polit, D. F. (2014). *Essentials of nursing research: Appraising evidence for nursing practice* (8th, international ed.). Wolters Kluwer Health/Lippincott Williams & Wilkins.

Polit, D. F., & Beck, C. T. (2020). *Essentials of nursing research: Appraising evidence for nursing practice* (10th ed.). Lippincott Williams & Wilkins.

Rattray, J., & Jones, M. C. (2007). Essential elements of questionnaire design and development. *Journal of Clinical Nursing*, *16*(2), 234–243. https://doi.org/10.1111/j.1365-2702.2006.01573.x

Robb, M., & Shellenbarger, T. (2020). Mastering Survey Design and Questionnaire Development. *Journal of Continuing Education in Nursing*, *51*(6), 248–249. https://doi.org/10.3928/00220124-20200514-02

Ryu, E., Kim, K., Choi, S. Y., Cleland, C. M., & Fu, M. R. (2013). The Korean version of the Symptom Experience Index: A psychometric study. *International Journal of Nursing Studies*, *50*(8), 1098–1107. https://doi.org/10.1016/j.ijnurstu.2012.12.008

Sharma, A., Minh Duc, N. T., Luu Lam Thang, T., Nam, N. H., Ng, S. J., Abbas, K. S., Huy, N. T., Marušić, A., Paul, C. L., Kwok, J., Karbwang, J., de Waure, C., Drummond, F. J., Kizawa, Y., Taal, E., Vermeulen, J., Lee, G. H. M., Gyedu, A., To, K. G., . . . Karamouzian, M. (2021). A Consensus-Based Checklist for Reporting of Survey Studies (CROSS). *Journal of General Internal Medicine*, *36*, 3179–3187. https://doi.org/10.1007/s11606-021-06737-1

Tajik, M., Galvão, H. M., & Eduardo Siqueira, C. (2010). Health survey instrument development through a community-based participatory research approach: Health Promoting Lifestyle Profile (HPLP-II) and Brazilian immigrants in Greater Boston. *Journal of Immigrant and Minority Health*, *12*(3), 390–397. https://doi.org/10.1007/s10903-008-9209-4

Zhao, M., Mou, H., Zhu, S., Li, M., & Wang, K. (2020). Cross-cultural adaptation and validation of the FRAIL-NH scale for Chinese nursing home residents: A methodological and cross-sectional study. *International Journal of Nursing Studies*, *105*, 103556–103556. https://doi.org/10.1016/j.ijnurstu.2020.103556

Author Biography

Siobhan O'Connor, BSc, CIMA CBA, BSc, RGN, PhD, is a Lecturer at the School of Nursing and Midwifery, National University of Ireland Galway, Ireland. She has a multidisciplinary background in both nursing and information systems. Hence, her research focuses on the design, implementation, and use of a range of technologies in healthcare.