Canadian Medical Education Journal
Revue canadienne de l'éducation médicale

Eight ways to get a grip on intercoder reliability using qualitative-based measures
Huit façons de maîtriser la fidélité intercodeur en utilisant des mesures qualitatives

Nicholas Cofie, Heather Braund et Nancy Dalgarno

Volume 13, numéro 2, 2022

URI : https://id.erudit.org/iderudit/1090341ar
DOI : https://doi.org/10.36834/cmej.72504

Aller au sommaire du numéro

Éditeur(s)

Canadian Medical Education Journal

ISSN

1923-1202 (numérique)

Découvrir la revue

Citer ce document

Cofie, N., Braund, H. & Dalgarno, N. (2022). Eight ways to get a grip on intercoder reliability using qualitative-based measures. *Canadian Medical Education Journal / Revue canadienne de l'éducation médicale*, *13*(2), 73–76. https://doi.org/10.36834/cmej.72504

Résumé de l'article

L'utilisation de mesures quantitatives de la fidélité intercodeur dans l'analyse de données de recherche qualitative a souvent suscité des débats acrimonieux parmi les chercheurs qui considèrent qu'en raison de leurs traditions ontologiques et épistémologiques différentes, les méthodologies de recherche quantitative et qualitative sont incompatibles. Bien que ces mesures soient précieuses dans de nombreux contextes, les critiques soulignent que leur utilisation dans l'analyse qualitative constitue une tentative d'importer des normes dérivées de la recherche positiviste. Nous nous appuyons sur les recherches existantes et sur notre expérience en recherche qualitative pour soutenir qu'il est possible de développer une mesure qualitative de la fidélité intercodeur qui soit compatible avec le paradigme épistémologique interprétativiste de la recherche qualitative. Nous proposons huit recommandations, fondées sur des lignes directrices en recherche qualitative pour évaluer et rapporter la fidélité intercodeur en recherche qualitative. Nous espérons qu'elles seront particulièrement utiles pour guider les chercheurs débutants dans les processus de codage et d'analyse des données qualitatives.

Canadian Medical Education Journal

Black Ice

Eight ways to get a grip on intercoder reliability using qualitative-based measures

Huit façons de maîtriser la fidélité intercodeur en utilisant des mesures qualitatives

Nicholas Cofie,[1] Heather Braund,[1] Nancy Dalgarno[1]

[1]Faculty of Health Sciences, Queen's University, Ontario, Canada

Correspondence to: Nicholas Cofie; email: nicholas.cofie@queensu.ca

Published ahead of issue: March 29, 2022, published: May 3, 2022. CMEJ 2022, 13(2) Available at https://doi.org/10.36834/cmej.72504

Abstract

The use of quantitative intercoder reliability measures in the analysis of qualitative research data has often generated acrimonious debates among researchers who view quantitative and qualitative research methodologies as incompatible due to their unique ontological and epistemological traditions. While these measures are invaluable in many contexts, critics point out that the use of such measures in qualitative analysis represents an attempt to import standards derived for positivist research. Guided by extant research and our experience in qualitative research, we argue that it is possible to develop a qualitative-based measure of intercoder reliability that is compatible with the interpretivist epistemological paradigm of qualitative research. We present eight qualitative research process-based guidelines for evaluating and reporting intercoder reliability in qualitative research and anticipate that these recommendations will particularly guide beginning researchers in the coding and analysis processes of qualitative data analysis.

Résumé

L'utilisation de mesures quantitatives de la fidélité intercodeur dans l'analyse de données de recherche qualitative a souvent suscité des débats acrimonieux parmi les chercheurs qui considèrent qu'en raison de leurs traditions ontologiques et épistémologiques différentes, les méthodologies de recherche quantitative et qualitative sont incompatibles. Bien que ces mesures soient précieuses dans de nombreux contextes, les critiques soulignent que leur utilisation dans l'analyse qualitative constitue une tentative d'importer des normes dérivées de la recherche positiviste. Nous nous appuyons sur les recherches existantes et sur notre expérience en recherche qualitative pour soutenir qu'il est possible de développer une mesure qualitative de la fidélité intercodeur qui soit compatible avec le paradigme épistémologique interprétativiste de la recherche qualitative. Nous proposons huit recommandations, fondées sur des lignes directrices en recherche qualitative pour évaluer et rapporter la fidélité intercodeur en recherche qualitative. Nous espérons qu'elles seront particulièrement utiles pour guider les chercheurs débutants dans les processus de codage et d'analyse des données qualitatives.

Introduction

The debate

The use of quantitative intercoder reliability (ICR) measures, such as the kappa statistic, weighted kappa statistic, and binomial intraclass correlation coefficients (ICC), in the analysis of qualitative research data has often generated acrimonious debates among researchers who view quantitative and qualitative research methodologies as incompatible due to their unique ontological and epistemological traditions.[1-3] Braun and Clarke,[1] for example, assert that reliability is not an appropriate criterion for judging qualitative work and that quantitative measures of ICR are epistemologically problematic. ICR has been defined as a numerical measure of the agreement between different coders regarding how the same data should be coded.[3] ICR can help provide confidence that systematic efforts were made to ensure the final qualitative data analytic framework is a credible and accurate representation of the data.[3]

ICR measures are used to assess the rigor and transparency of the coding frame and its application to the data.[4-7] A high ICR may be used to assure the research team and audience

that the coding frame is sufficiently well specified to allow for its communicability across persons.[5,8,9] Performing an ICR assessment also ensures that multiple researchers can understand and contribute to the analytic process and provides confidence that the analysis transcends the imagination of a single individual. ICR assessment further ensures that the patterns in the latent content is fairly robust to the degree that if readers were to code the same qualitative text, they would make the same judgments or produce the same results.[10] ICR fosters reflexivity and can serve as a badge of trustworthiness[3] to the extent that some journal editors and reviewers now request or require a measure of ICR before agreeing to publish qualitative studies.[11] Taken together, ICR might improve the systematicity, communicability, and transparency of the coding process and promote reflexivity and dialogue within research teams. [3] Other critics note, however, that the use of ICR in qualitative analysis represents an attempt to import standards derived for positivist research[12,13] and that its use could mask the fact that a rigorous, in-depth qualitative analysis was not undertaken.

A major pitfall surrounding the use of quantitative ICR measures in qualitative research is that such use may create the incorrect assumption that somehow quantitative ICR measures do not essentially contradict the interpretative agenda of qualitative research[1,14-16] which requires the researcher to see the research field as composed of multiple perspectival realities that are intrinsically constituted by an individual's social context and personal history.[17] As O'Conner and Joffe[3] note, the role of the qualitative researcher is not to reveal universal objective facts but to apply their theoretical expertise to interpret and communicate the diversity of perspectives on a given topic. Despite this inherent pitfall, some qualitative researchers often resort to quantitative based ICR measures or use their own methods that may not be well grounded in the literature. Also, in the absence of clear or adequate guidelines, some authors hesitate to engage in ICR assessments. We present eight process-based guidelines on ways to get a grip on intercoder reliability using qualitative-based measures. This paper is intended for use by researchers across the continuum and is particularly valuable for beginning researchers.

An alternative measure of ICR

We argue that it is possible to develop a robust measure of ICR that is unique and compatible with the interpretivist epistemological paradigm of qualitative research. This paradigm is premised on relativist ontology and subjectivist epistemology and assumes that reality as we know it is constructed intersubjectively through the meanings and understandings developed socially and experientially and that we cannot separate ourselves from what we know.[18] This measure need not be statistical or quantitative. It can be descriptive and must be able to qualitatively characterize the extent to which independent coders agree or disagree on codes produced from interview, focus group, visual, and textual data. This approach must emphasize the need to achieve consistency between coders rather than mere quantification of the extent of agreement between coders and encourages reflexivity and authenticity throughout the qualitive analysis process. This alternative view of ensuring consistency is echoed by many qualitative researchers who argue that coding and identification of themes by independent researchers could be followed by a group discussion of overlaps and divergences[19] without necessarily quantifying the degree of consensus achieved between the coders.[3] In the rest of the commentary, we present and discuss a set of guidelines for evaluating and reporting ICR in qualitative data analysis based on prior research and the authors' own experiences in the application of qualitative and quantitative research methods.[3,20-22] These guidelines are intended to be used in conjunction with other guidelines including those described elsewhere in the lietrature.[23] We have several years of diverse experience in mixed research methodology including coding and analyzing interviews, focus groups, and textual data, as well as narrative responses from survey data.

Ways to get a grip on evaluating and reporting ICR

Guided by extant research and our experience in qualitative research, we recommend eight ways to get a grip on evaluating and reporting ICR in qualitative research with the goal of achieving consistency in the coding process. These are summarized in Table 1.

1. We suggest that at least two researchers must code the data, except in situations where the goal of the coding is to assess the extent of intracoder (within a single coder) reliability, wherein emphasis is placed on the extent of consistency with respect to how the same person codes data at multiple time points.[20-21] As Conner and Joffe[3] describe, if the same person returns to the data at another time, it is possible to assess the

extent of consistency in the coding process, thereby promoting researcher reflexivity.[5]

2. To ensure transparency and minimize bias, we recommend that at least one of the coders in the research team must be external to or removed from the data collection process in such a way that this external coder may view and code the data from a fresh perspective.

3. We recommend that at least one of the coders have expertise and previous experience with coding qualitative data to ensure that the coding and development of themes are done in a rigorous and robust manner, thereby increasing the consensus among coders.

4. Steps must be taken to ensure that use of novice coders (together with experienced coders) does not produce unreasonable discrepancies in coding and development of themes.

5. We also suggest that if a project includes multiple participant groups, a minimum of two researchers should code transcripts from each participant group.[3,20]

6. We highly recommend that the coders use the same framework for analysis to ensure that basic concepts or themes developed within the analysis are consistent with the theoretical framework guiding the research.

7. Accordingly, we suggest that coders should focus on shared meaning of codes through a dialogue and consensus processes. However, where discrepancies in codes and themes emerge, we recommend that another coder with expertise in qualitative methods is consulted to resolve such observed discrepancies.

8. We recommend that the resulting codebook (based on consensus reached from selected transcripts) should be used to code the remaining transcripts. In inductive and abductive analyses, coding can be an iterative process; therefore, we suggest that new codes may be added to the codebook until a reasonable code saturation is reached.[24] The researchers could therefore schedule regular team meetings to discuss and achieve consensus on the newly added codes. We recommend that researchers should try to use as many criteria as often as possible to increase the rigor, trustworthiness, authenticity, and meaningfulness of qualitative research. However, if the researcher is unable to use all criteria, they should reflect and justify why they were unable to apply all the criteria.

Table 1. Ways to get a grip on Intercoder Reliability

Aspects of Intercoder Reliability	Present		Justification (If 'no' selected)
There was a minimum of two coders.	Yes	No	
At least one coder was more removed from data collection (to address bias).	Yes	No	
At least one coder had expertise and previous experience with coding qualitative data.	Yes	No	
If there were multiple participant groups, a minimum of two researchers (coders) coded transcripts from each participant group.	Yes	No	
The coders used the same framework for analysis (e.g., inductive, deductive, abductive).	Yes	No	
Coders focused on shared meaning of *codes through dialogue and consensus.	Yes	No	
Another coder with expertise in qualitative methods was consulted to resolve outstanding conflicts.	Yes	No	
Coder consensus resulted in a codebook** that was applied when coding the remaining transcripts.	Yes	No	

** The code names do not have to be identical, but the meaning of the codes must be the same.*
***In inductive and abductive analyses, coding can be an iterative process; therefore, new codes may be added to the codebook until code saturation is reached.*

Conclusion

We note that while there are valid reasons for incorporating quantitative-based measures of ICR into qualitative research, it is possible to develop a qualitative-based measure of ICR that is unique and compatible with the interpretivist epistemological paradigm of qualitative research. Drawing on prior research and research experience, we note further that this alternative measure does not need to be statistical in nature, however it must be able to characterize the extent to which independent researchers agree or disagree on codes produced from qualitative data and encourage reflexivity and authenticity throughout the qualitive analysis process. We anticipate that the recommendations presented here will guide researchers across the continuum, particularly beginning researchers in assessing the degree to which quality of process in ICR was met for qualitative data analysis.

Conflicts of Interest: The Authors declare no conflicts of interest.
Acknowledgement: The authors acknowledge Amber Hastings-Truelove, PhD, Britney Lester, M.Ed, Shannon Hill, MA, Eleftherios Soleas, PhD for their review and feedback.

References

1. Braun V, Clarke V. Successful qualitative research. Thousand Oaks, CA: Sage Publications Ltd; 2013.
2. Davey JW, Gugiu PC, Coryn CLS. Quantitative methods for estimating the reliability of qualitative data. *J MultiDiscip Eval.* 2010; 6(13): 140-162.
3. O'Conner C, Joffe H. Intercoder reliability in qualitative research: Debates and practical guidelines. *Int J Qual Methods.* 2020; 19: 1–13. https://doi.org/10.1177/1609406919899220
4. Hruschka DJ, Schwartz D, St John DC, Picone-Decaro E, Jenkins RA, Carey JW. Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field Methods*. 2004;16: 307–331. https://doi.org/10.1177/1525822X04266540
5. Joffe H, Yardley L. Content and thematic analysis. In: Marks DF, Yardley L, eds. *Research methods for clinical and health psychology*. Thousand Oaks, CA: Sage Publications Ltd; 2003. p. 56–68.
6. MacPhail C, Khoza N, Abler L, Ranganathan M. Process guidelines for establishing intercoder reliability in qualitative studies. *Qual Res*. 2016;16(2):198-212. https://doi.org/10.1177/1468794115577012
7. Mays N, Pope C. Qualitative research: Rigour and qualitative research. *Brit MedJ*. 1995; 311, 109–112. https://doi.org/10.1136/bmj.311.6997.109
8. Joffe H. Risk and 'the other'. Cambridge University Press; 1999. https://doi.org/10.1017/CBO9780511489846
9. Joffe H, Rossetto T, Solberg C, O'Connor C. Social representations of earthquakes: A study of people living in three highly seismic areas. *Earthq Spectra*. 2013; 29: 367–397. https://doi.org/10.1193/1.4000138
10. Potter WJ, Levine-Donnerstein D. Rethinking validity and reliability in content analysis. *J Appl Commun*. 1999; 27: 258–284. https://doi.org/10.1080/00909889909365539
11. Wu S, Wyant DC, Fraser MW. Author guidelines for manuscripts reporting on qualitative research. *J Soc Soc Work Res*. 2016; 7: 405–425. https://doi.org/10.1086/685816
12. Guba EG, Lincoln YS. Competing paradigms in qualitative research. In: Denzin NK, Lincoln YS, eds. Handbook of qualitative research. Thousand Oaks, CA: Sage Publications Ltd; 1994. p. 105–117.
13. Madill A, Jordan A, Shirley C. Objectivity and reliability in qualitative analysis: Realist, contextualist and radical constructionist epistemologies. *Br J Psychol.* 2000; 91: 1–20. https://doi.org/10.1348/000712600161646
14. Hollway W, Jefferson T. Doing qualitative research differently: A psychosocial approach. 2nd ed. Sage; 2013. https://doi.org/10.4135/9781526402233
15. Vidich J, Lyman SM. Qualitative methods: Their history in sociology and anthropology. In: Denzin NK, Lincol YS, eds. Handbook of qualitative research. Sage; 1994. p. 23-59.
16. Yardley L. Dilemmas in qualitative health research. *Psychol & Health*. 2000;15: 215–228. https://doi.org/10.1080/08870440008400302
17. Bauer M, Gaskell G, Allum N. Quality, quantity and knowledge interests: Avoiding confusions. In: Bauer MW, Gaskell G, eds. Qualitative Researching with Text, Image and Sound. London: Sage Publications Ltd; 2000. p. 4-17. https://doi.org/10.4135/9781849209731
18. Cohen D, Crabtree B. Qualitative research guidelines project. 2006. Online. Available from http://www.qualres.org/HomeInte-3516.html
19. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med Res Methodol.* 2008; 8(45). https://doi.org/10.1186/1471-2288-8-45.
20. Campbell JL, Quincy C, Osserman J, Pedersen OK. Coding In-depth Semistructured Interviews: Problems of Unitization and Intercoder Reliability and Agreement. *Sociol Methods & Res*. 2013;42(3):294-320. https://doi.org/10.1177/0049124113500475
21. Creswell JW, Creswell JD. Research design: Qualitative, quantitative, and mixed methods approaches. 5th ed. Sage; 2018.
22. Braun V, Clarke V. *Thematic analysis: A practical guide*. Sage; 2022.
23. Patton MQ. Qualitative research & evaluation methods: Integrating theory and practice. 4th ed. Thousand Oaks, CA: Sage; 2015.
24. Braun V, Clarke V. To saturate or not saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qual Res Sport Exerc Health.* 2021; 13(2): 201-216. https://doi.org/10.1080/2159676X.2019.1704846