




RESEARCH ARTICLE

Developing and testing inter-rater reliability of a data collection tool for patient health records on end-of-life care of neurological patients in an acute hospital ward

Gudrun Jonsdottir^{1,2}  | Erna Haraldsdottir³  | Valgerdur Sigurdardottir² | Asta Thoroddsen¹ | Runar Vilhjalmsón¹ | Gudny Bergthora Tryggvadottir⁴ | Helga Jonsdottir¹ 

¹Faculty of Nursing and Midwifery, School of Health Sciences, University of Iceland, Reykjavik, Iceland

²Landspítali, The National University Hospital of Iceland, Reykjavik, Iceland

³Division of Nursing, Queen Margaret University, Edinburgh, UK

⁴The Social Science Research Institute, University of Iceland, Reykjavík, Iceland

Correspondence

Gudrun Jonsdottir, Faculty of Nursing, University of Iceland, Eirberg, Eiriksgata 34, 101 Reykjavik, Iceland.
Email: gudjonsd@landspitali.is

Funding information

Landspítali, The National University Hospital of Iceland; The Icelandic Nurses Association; The University of Iceland

Abstract

Aim: Develop and test a data collection tool—Neurological End-Of-Life Care Assessment Tool (NEOLCAT)—for extracting data from patient health records (PHRs) on end-of-life care of neurological patients in an acute hospital ward.

Design: Instrument development and inter-rater reliability (IRR) assessment.

Method: NEOLCAT was constructed from patient care items obtained from clinical guidelines and literature on end-of-life care. Expert clinicians reviewed the items. Using percentage agreement and Fleiss' kappa we calculated IRR on 32 nominal items, out of 76 items.

Results: IRR of NEOLCAT showed 89% (range 83%–95%) overall categorical percentage agreement. The Fleiss' kappa categorical coefficient was 0.84 (range 0.71–0.91). There was fair or moderate agreement on six items, and moderate or almost perfect agreement on 26 items.

Conclusion: The NEOLCAT shows promising psychometric properties for studying clinical components of care of neurological patients at the end-of-life on an acute hospital ward but could be further developed in future studies.

KEYWORDS

end-of-life care, face validity, instrument development, inter-rater reliability, neurological patients, patient health records

1 | INTRODUCTION

Studies that have investigated end-of-life care of neurological patients highlight that it is a complex subject (Alonso et al., 2016; Hussain et al., 2018). To shed light on current practice concerning end-of-life care of neurological patients in an acute neurological ward, it was decided to retrospectively retrieve data from patient

health records (PHRs). The aim of this paper is to describe the design and testing inter-rater reliability of a data collection tool to map end-of-life care for neurological patients in the last 7 days of their life. The last 7 days were selected since it could be expected that within that time frame indicators of impending death of neurological patients should have manifested and could be noted (Hui et al., 2014, 2015). Designing the data collection tool is part of a

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Nursing Open* published by John Wiley & Sons Ltd.

larger study of diagnosing dying among neurological patients in an acute hospital ward.

2 | BACKGROUND

The final stage of palliative care or end-of-life care refers to health-care provided in the last weeks and months of life. End-of-life care is given to patients when medical treatment to cure their disease has been brought to an end (Cohen-Mansfield et al., 2018). The ultimate goal is to relieve suffering, for instance by optimizing pain management and providing psychological and social support to assist patients and their next-of-kin with managing physical, emotional, social, and spiritual burden of the imminent death (Huskamp et al., 2012). High quality end-of-life care requires communication between the patient, healthcare professionals involved in the care, as well as relatives in order to create a shared understanding about the person's values and treatment preferences (Fernando & Hughes, 2019; Gonella et al., 2020; Sinuff et al., 2015; Taffurelli et al., 2020). This communication ultimately leads to an individualized care plan that is consistent with the patient's values and needs, considering what treatments, including assessments and interventions, will or will not be used to manage the symptoms of the life-threatening disease (Sinuff et al., 2015). Such individualized care, based on highly tuned clinical judgement, using clinical indicators along with experience and clinical wisdom, is paramount for the right course of action in end-of-life care (Kennedy et al., 2014).

To shed light on current practice in relation to end-of-life care of neurological patients, a retrospective study is attractive. In this situation the main strength of this non-intrusive research design is that at this delicate point in the patients' and the relatives' lives they are not bothered by data collection (Kaji et al., 2014). PHRs are widely used in retrospective studies more generally. The records can be in both electronic and paper form (Gregory & Radovinsky, 2012). PHRs contain already documented information, which is intended for patient care, not research purposes. However, there are multiple advantages of using PHRs to obtain research data. The most prominent one is that they can provide a large amount of data of clinical significance at relatively little cost, without taking up patients' time or disturbing patients in any way (Gregory & Radovinsky, 2012; Kaji et al., 2014). PHRs in electronic form (EHR) are more reliable than the paper form as they are stored permanently, so there is less chance of losing data. The shortcomings of retrospective studies of PHRs are that the data is already collected and unchangeable, so the researcher cannot influence what information could be documented; there may be incomplete or missing data; specific patient information may be lacking; and there may be difficulties in interpreting or verifying the information (Feder, 2017; Fortney & Steward, 2015; Jansen et al., 2005). There is also variability in the quality of how healthcare professionals originally documented patient information in the PHRs (Cassidy et al., 2002; Gianinazzi et al., 2015; Gregory & Radovinsky, 2012; Kaji et al., 2014).

With regard to extracting the data, the data abstractors may read, interpret, code, and transcribe what is written in the patients' records differently from what was initially intended, resulting in low sensitivity and specificity of the data collected (Jansen et al., 2005; Kaji et al., 2014). The use of standardized data collection methods or guidelines to ensure consistent collection of data enhances the quality of the data and minimizes bias (Gregory & Radovinsky, 2012; Jansen et al., 2005; Kaji et al., 2014). Reporting on how inter-rater reliability (IRR) in data collection is achieved shows transparency and may avoid inconsistencies in extraction of the data both within and between the data abstractors (Alonso et al., 2016; Cox et al., 2011; Fortney & Steward, 2015; Kaji et al., 2014). Reports on IRR with percentage agreements and kappa calculations from studies where data has been collected from PHRs have shown the data collection tools to be reliable, with high sensitivity and specificity (Gianinazzi et al., 2015; Ntlholang et al., 2016; Yawn & Wollan, 2005).

Several studies of patients with various neurological diseases have been conducted to improve care and provide evidence for the need of person-centered end-of-life care, with some studies using retrospective PHRs (Alonso et al., 2016; Cheng et al., 2017; Hussain et al., 2018; Liu et al., 2017; Munroe et al., 2007; Ntlholang et al., 2016; Quadri et al., 2018; Wang et al., 2018; Williams et al., 2019). These studies provide limited information on the data collection tools that were used, their psychometric properties, and the training of data abstractors.

3 | DESIGNING A DATA COLLECTION TOOL

To retrospectively study the practice of end-of-life care of neurological patients in an acute hospital ward with the use of PHRs, a data collection tool—Neurological End-Of-Life Care Assessment Tool (NEOLCAT)—was developed. The aim of developing the NEOLCAT is to capture how and if signs that may indicate imminent death of neurological patients are identified and documented. The NEOLCAT has the potential to aid future research of care of neurological patients at end-of-life. Further, the NEOLCAT has the potential to be used in clinical care to support identification of neurological patients who are facing imminent death and thus supporting decision making in relation to shifting the focus of care to end-of-life care, thus improving care for this patients' group.

The NEOLCAT (see File S1) contains key components of end-of-life care of neurological patients on an acute hospital ward as highlighted in the literature, including demographics/background, dates of major decisions about treatment, clinical signs and symptoms, laboratory and other tests undertaken, medically invasive and other treatments, communication with relatives, and healthcare professionals' contribution to end-of-life care.

The development and testing of the NEOLCAT were divided into three phases: (1) *review of the literature*, (2) *expert advice*, and (3) *ensuring and reporting IRR* (Gregory & Radovinsky, 2012; Jansen et al., 2005; Kaji et al., 2014).

3.1 | Review of literature

This *first phase* included identifying and studying relevant literature and guidelines in the relation to the purpose of the NEOLCAT. A medical librarian assisted with a literature search for the keywords: diagnosing dying, end-of-life care, palliative care, terminal care, neurological disease, stroke, ALS, MND, and Parkinson's, which yielded 20 relevant articles and guidelines that informed the NEOLCAT (Alonso et al., 2016; Bruera et al., 2014; Cheng et al., 2017; Cohen-Mansfield et al., 2018; Eriksson et al., 2016; Highet et al., 2014; Hui et al., 2014, 2015; Huskamp et al., 2012; Kennedy et al., 2014; Landspitali, 2017; Liu et al., 2017; Mazzocato et al., 2010; McCusker et al., 2013; Munroe et al., 2007; National Clinical Guideline Centre (NICE), 2015; Ntlholang et al., 2016; Quadri et al., 2018; Registered Nurses' Association of Ontario (RNAO), 2011; Sinuff et al., 2015). Systematically scrutinizing the literature generated a list of possible items, which was listed in an Excel file. Of notice is the emphasis that is placed on communication with relatives, and on spirituality. Although these are important aspects of end-of-life care for neurological patients, the format for documenting them in EHR is yet to be developed which made it unlikely that they would be identified and extracted (Forde-Johnston et al., 2022; Sjöberg et al., 2021). The spectrum of these items is therefore limited in the NEOLCAT. Table 1 shows literature and guidelines that formed the foundation of the NEOLCAT.

3.2 | Expert advice

The *second phase* consisted of an iterative process of developing the item list further to its completion. This included conversations with professional experts in palliative care, neurology, and health informatics to gain consensus on items that would finally make up the NEOLCAT. With several formal and informal meetings with the experts, the item list was refined based on their comments and suggestions and finally approved. The NEOLCAT has both objective and subjective items. The wording of each item was carefully studied. A decision was made to have the NEOLCAT in English, with translations in Icelandic in the coding manual. The first draft consisted of 84 items and the final version has 76 items falling into the following seven categories: Demographics/Background, major decisions in treatments, clinical signs and symptoms, laboratory and other tests, medically invasive and other treatments, communication with relatives, and health professionals contribution. Finally, the NEOLCAT was inserted in Research Electronic Data Capture (RedCap) (Harris et al., 2009).

3.3 | Ensuring and reporting the inter-rater reliability of the NEOLCAT

The *third phase* entailed ensuring a consistent data collection procedure and testing the IRR of the NEOLCAT. A coding manual of

definitions of items in the NEOLCAT and how to locate each item in the PHRs was developed. The coding manual was made as accurate as possible with guiding images presented for how and where the data extractors could find the exact information. An example of a definition in the coding manual is given in Figure 1.

The data abstractor team consisted of two students in medicine and psychology and one registered nurse. They underwent individualized training as data abstractors, which consisted of a detailed presentation of the study and instruction in using the coding manual and in extracting the data. Following this, the abstractors' team met formally twice with the primary researcher (GJ) for co-ordination. Face-to-face contacts between the researcher and each of the abstractors continued with numerous meetings during the data collection period. During this training period inter-rater reliability was calculated three times and the mismatch that was found was explored and corrected. This allowed for discussion around ways of handling conflicting data and clearing up misunderstandings. This was key to facilitating and gaining a shared understanding of each of the items in the NEOLCAT. Examples of issues that needed to be resolved were misunderstandings about the dates of radiological examinations. The study is limited to the last 7 days of life, but the data abstractors, on some occasions, wrongly used information that was documented within a longer range than 7 days. This was corrected after meetings with the researcher. The medium time for extracting data for one patient's EHR was about 60 min.

3.3.1 | Ensuring face validity

Face validity was ensured by thoroughly reviewing the literature and having several rounds of communication with the clinical experts, all with over 20 years of experience. When designing an instrument there should be an exhaustive literature review, rich first-hand knowledge of end-of-life care, consultation with experts, and in-depth conversation with members of the target population, in our case, the healthcare professionals (Polit & Beck, 2017). All available EHRs of patients ($N=170$) who had died in the acute neurological unit over the five previous years were included in the study.

3.3.2 | Testing inter-rater reliability

Inter-rater reliability is defined as the degree of agreement or consensus of raters, in this case, the consistency of extracting data with NEOLCAT by data abstractors (Landis & Koch, 1977; LeBreton & Senter, 2008; McHugh, 2012; Polit & Beck, 2017). It was established by regular communication and meetings between researchers and data abstractors to discuss and handle conflicting data from the patient records and evaluated by calculating percentage agreement and the Fleiss' kappa coefficient. The Fleiss' kappa coefficient was categorized into fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect agreement (0.81–1.00); (Landis & Koch, 1977). There should be a consistency between high percentage

TABLE 1 Key literature and clinical guidelines for end-of-life care foundational to the NEOLCAT.

	Major treatment decisions	Clinical signs and symptoms	Laboratory and other tests	Medically invasive and other treatments	Communications with relatives	Healthcare professionals' contribution to decision-making
Alonso et al. (2016)	✓	✓	✓	✓		✓
Bruera et al. (2014)		✓	✓			
Cheng et al. (2017)	✓	✓	✓	✓		✓
Cohen-Mansfield et al. (2018)		✓				
Eriksson et al. (2016)	✓	✓	✓	✓	✓	✓
Highet et al. (2014)	✓	✓	✓	✓	✓	✓
Huskamp et al. (2012)	✓	✓	✓	✓	✓	✓
Hui et al. (2014), Hui et al. (2015)		✓				
ISCI/McCusker et al. (2013)	✓	✓	✓	✓	✓	✓
Kennedy et al. (2014)	✓	✓	✓	✓	✓	✓
Landspitali (2009/, 2017)	✓	✓	✓	✓	✓	✓
Liu et al. (2017)	✓	✓				✓
Mazzocato et al. (2010)	✓	✓	✓	✓		✓
Munroe et al. (2007)	✓					✓
NICE (2015)	✓	✓	✓	✓	✓	✓
Ntlholang et al. (2016)	✓	✓		✓		✓
RNAO (2011)	✓	✓	✓	✓	✓	✓
Sinuff et al. (2015)	✓	✓			✓	✓
Quadri et al. (2018)	✓	✓	✓	✓	✓	✓

ISCI, Institute for Clinical Systems Improvement; Landspitali, Palliative Care Guidelines; NICE, National Institute for Health and Care Excellence; RNAO, Registered Nurses' Association of Ontario.

<p>Fatigue in the 3-7 days before death</p> <ul style="list-style-type: none"> ○ Yes ○ No ○ Unknown 	<p>The patient's fatigue is rated here. Fatigue is defined as a decrease in physical and psychological activity and a feeling of weakness that can often be attributed to illness or treatment.</p> <p>Traditional remedies such as sleep and rest do not reduce the fatigue, which has a significant effect on quality of life and can be very difficult for the patient.</p> <p>Fatigue is one of the most common symptoms that patients with chronic and progressive illnesses complain of. If a patient is unconscious, then this does not apply.</p> <p>Fatigue is recorded in the nurse's documentation.</p>
--	--

FIGURE 1 Definition of fatigue. An example from the coding manual.

agreement and high Fleiss' kappa coefficient. Therefore, it is important to present both calculations (Feinstein & Cicchetti, 1990; To et al., 2008).

Percentage agreement was calculated without delay in the study period as soon as three abstractors had finished abstracting the same PHRs. All three data abstractors extracted data from eight randomly selected PHRs, subsequently calculating the percentage agreement and the Fleiss' kappa, considering agreement occurring

by chance (Cassidy et al., 2002; Feinstein & Cicchetti, 1990; Landis & Koch, 1977; Polit & Beck, 2017; To et al., 2008; Zaiantz, 2019). A percentage agreement that was under 80% was given special consideration and action was taken to rectify data extraction on that item. Disagreement decreased substantially after issues about data extraction of the first three PHRs had been resolved. The data that was gained in the last round of rectifying the data extraction is used in this paper.

With multiple data abstractors it is possible that IRR can be compromised (Gianinazzi et al., 2015). Gregory and Radovinsky (2012) highlighted that to obtain high IRR it is necessary to conduct continuous monitoring and periodic reviews, and to make detailed reports on limitations that were encountered through the data abstraction processes, and to explain how these limitations were addressed (Gregory & Radovinsky, 2012). The items chosen for IRR calculation were nominal variables having yes/no/unknown coding possibilities. Thirty-two items out of 76 were chosen by importance. Items with the possibility of multiple answers e.g., medical diagnoses, signs of dying, and text answers were not included. In calculating percentage agreement and Fleiss' kappa coefficients, two categories were grouped together, Demographic and Major decisions on treatments, since both contained few items.

4 | RESULTS

We report on average percentage agreement and average Fleiss' kappa coefficient for the 32 items of the NEOLCAT belonging to six categories having one to 16 items each, see Table 2 (see supplement 1 for the total of categories and items). This is followed by calculations of congruence between the NEOLCAT items into fair, moderate, substantial, and almost perfect agreement.

4.1 | Percentage agreement

Percentage agreement on the 32 items of NEOLCAT varied from 71% to 100% with 89% overall percentage agreement, see Table 3. The percentage agreement of the six categories ranged between 83% and 95%. The category Health professionals' contribution (83%) was the lowest and Medically invasive and other treatments the highest (95%). The percentage agreement range was biggest in the category Clinical signs and symptoms (71%–100%).

4.2 | Fleiss' kappa coefficient

The Fleiss' kappa coefficient of the 32 items varied between 0.24 and 1.0 with overall coefficient 0.84, see Table 3. The coefficients of the six categories ranged between 0.71 and 0.91. It was highest in the category Medically invasive and other treatments (range 0.43–1.0) and lowest in the category Clinical signs and symptoms (range 0.24–1.0) which had also the biggest range.

4.3 | Congruence between percentage agreement and Fleiss' kappa coefficient

Congruence between percentage agreement and Fleiss' kappa coefficient is presented on all the 32 items.

TABLE 2 Percentage agreements and Fleiss' kappa coefficients.

Items	Percentage agreement	Fleiss' kappa
Demographics/background		
Number of admittances to the hospital	92	0.89
Major decisions in treatment		
Use of "Care pathway for the dying"	92	0.74
Clinical signs and symptoms		
Assessment of consciousness 48h before death	75	0.50
Assessment of consciousness 3–7 days before death	83	0.75
Assessment of alertness/responsiveness 48h before death	83	0.24
Assessment of alertness / responsiveness in 3–7 days before death	100	1.00
Signs of progression of disease in the 48h before death	92	0.80
Signs of progression of disease in the 3–7h before death	92	0.70
Assessment of mobility 48h	100	1.00
Assessment of mobility 3–7 days	75	0.58
Fatigue in the 48h before death	75	0.43
Fatigue in the 3–7 days before death	100	1.00
Nausea in the 48h before death	83	0.82
Nausea in the 3–7 days before death	83	0.72
Dyspnoea in the 48h before death	71	0.59
Dyspnoea in the 3–7 days before death	71	0.73
Fluid balance in the 48h before death	100	1.00
Fluid balance in the 3–7 days before death	100	1.00
Laboratory and other tests undertaken		
Blood tests in the 48h before death	83	0.60
Urine tests taken in the 48h before death	83	0.75
Urine culture taken in the last 7 days	92	0.88
Medically invasive and other treatments		
Vital signs measured in the 48h before death	100	1.00
Vital signs measured in the 3–7 days before death	92	0.76
Intravenous catheter in place at the time of death	88	0.80
Urinary catheter in place at the time of death	83	0.73
Antibiotics given in the last 48h	100	1.00
Pain medication given in the last seven days	100	1.00
Tube feeding given in the last seven days	100	1.00

(Continues)

TABLE 2 (Continued)

Items	Percentage agreement	Fleiss' kappa
Communication with relatives		
Relatives mentioned in PHRs	100	1.00
Relatives present at the time of death of the patient	92	0.83
Health professionals meeting with relatives prior to death (family meeting, formal or informal)	92	0.83
Health professionals' contribution to decision-making of end-of-life care		
Evidence of health care professionals' assessment of patients' needs for end-of-life care	83	0.66
Overall percentage and Fleiss' kappa	89%	0.84

4.3.1 | Fair congruence

There was a fair agreement for two items: *Alertness/responsiveness* and *fatigue in the last 48h before death*. In *alertness/responsiveness at 48h before death*, the percentage agreement was 83% and Fleiss' kappa coefficient 0.24. For *fatigue in the last 48h*, the percentage agreement was 75% and kappa coefficient 0.43. The low kappa coefficient and high percentage agreement contradict each other.

4.3.2 | Moderate congruence

There was moderate percentage agreement for four items: *Consciousness*, *dyspnea*, *blood tests at 48h*, and *mobility in the 3–7 days before death*. The percentage agreement was ranged from 71% to 83%. The kappa coefficient was reasonably consequent and ranged from 0.50 to 0.60.

4.3.3 | Substantial congruence

There was substantial agreement for 11 items: *Progression of disease* both at 3–7 days and 48h before death, *consciousness* in 3–7 days, *dyspnea* and *nausea* in 3–7 days, *vital signs* in 3–7 days, *urine tests*, *intravenous catheter*, *urinary catheter*, *care pathway for the dying* and *assessment of need for end-of-life care by health professional*. The percentage agreement ranged between 71% and 100%. The kappa coefficient was consequent and ranged from 0.66 to 0.80.

4.3.4 | Almost perfect congruence

There was almost perfect agreement for 15 items: *Number of admissions*, *alertness/responsiveness at 3–7 days before death*, *vital signs* and *mobility at 48h*, *nausea at 48h*, *fatigue at 3–7 days*, *fluid assessment at 48h and 3–7 days*, *urine culture*, *antibiotics*, *pain medication*, *tube*

feeding, *relatives mentioned in PHRs*, *relative's presence at time of death* and *health professionals meeting with relatives prior to death*. The percentage agreement ranged between 83% and 100%. The kappa coefficient was consequent and ranged from 0.82 to 1.00.

5 | DISCUSSION

Findings of this study show that the NEOLCAT data collection tool which was developed to retrospectively collect data from PHRs for research on end-of-life care for neurological patients in an acute hospital ward shows promising psychometric properties. We established face validity in close collaboration with clinical experts and followed guidelines in the literature to enhance accuracy of data extraction. The consistency in the work of the data abstractors was established by calculating inter-rater reliability (IRR). The overall categorical percentage agreement of the NEOLCAT was 89% (range 83%–95%) and overall Fleiss's kappa categorical coefficient was 0.84 (range 0.71 to 0.91). This shows that NEOLCAT has high inter-rater reliability and suggests that it is of acceptable quality to be used in our larger study of end-of-life care of neurological patients in an acute hospital ward.

There was fair congruence between two items and moderate congruence between four. Incongruence could be expected on a few items and can be explained by lack of rigour in interpreting clinical information in free text and difficulties in locating items in the PHRs. The category with the lowest percentage agreement had only one item and the data were to be extracted from free text. Items of the category Laboratory and other tests undertaken showed acceptable but lower percentage agreement and Fleiss' kappa coefficient than could be expected since required interpretation is minimal. This data was, however, to be found in free text rather than having a fixed location in PHR, which may explain the low values that were found. Several of the items in the category, Clinical signs and symptoms (12 items), required interpretation of free text, yet did have acceptable values. It is concluded that overall, the IRR was acceptable. Low percentage agreement and Fleiss' kappa coefficient would to a great extent be explained by lack of rigorousness of the guidelines in the coding manual as to how to interpret clinical information and where to locate data. The coding manual, therefore, needs revision on those items.

In this study, it was not possible to extract data from PHRs of all the patients simultaneously from the data warehouse. For that to be possible, the PHR needs to have a more rigorous structure and built into the coding system. Therefore, we manually extracted data for one patient at a time, which is both time-consuming and increases likelihood of mistakes. An important implication of this study is, therefore, the need for improvements in how the PHR is structured and coded.

The general assumption is that end-of-life care involves teamwork (Fernando & Hughes, 2019; Kennedy et al., 2014). Teamwork was not reflected in the PHR. Nurses, physicians, and other health-care professionals wrote in different sections of the PHR about the

TABLE 3 Inter-rater reliability.

	Number of items	Percent agreement		Kappa	
		Category (%)	Range (%)	Category K (95% CI)	K range across items
Demographics and major treatment decisions	2	88	83–92	0.85 (0.73–0.97)	0.76–0.86
Clinical signs and symptoms	16	86	71–100	0.83 (0.76–0.89)	0.24–1.00
Laboratory and other tests undertaken	3	86	83–92	0.71 (0.48–0.95)	0.40–0.83
Medically invasive & other treatments	7	95	83–100	0.91 (0.90–0.93)	0.43–1.00
Communication with relatives	3	94	92–100	nc	
Health professionals' contribution	1	83	83	nc	
Overall	32	89		0.84 (0.80–0.89)	

nc, Not calculated because of small distribution in the sample.

same issues of the care that was provided. It would be useful to have a structured and coded PHR jointly for all healthcare professionals when documenting end-of-life care. That would not only improve the quality of documented care and identify areas for improvement in clinical practice but would also ease data collection in future research.

5.1 | Strength and limitations

It is a considerable strength of the study to have substantive literature for item selection and expert advice from different healthcare professionals to contribute to the design of the NEOLCAT. There were however only a couple of instruments on which to ground the NEOLCAT. Another strength is the rigorousness of the data extraction process, the training of the data abstractors and the use of a coding manual.

There are inbuilt limitations to data collection tools that are aimed for researching PHR, most importantly that the research data is unchangeable and that they are intended for patient care, not research purposes. Lack of structure and coding possibilities of the PHR is a limitation as well. Items of communication with patient and family, and on spirituality, in the NEOLCAT are incomplete. There should be a strong focus on these issues in future development of the NEOLCAT.

6 | CONCLUSIONS

The NEOLCAT data collection tool which was developed to retrospectively collect data from PHR for research on end-of-life care for neurological patients in an acute hospital ward shows promising psychometric properties and has the potential to be used in clinical research. It contains items of patient demographics, dates of

major treatment decisions, clinical signs and symptoms, laboratory and other tests, medically invasive and other treatments, communication with relatives, and healthcare professionals' involvement in end-of-life care. Items of communication with patient and family and on spirituality are incomplete in the NEOLCAT. Those items should be advanced in future development of the NEOLCAT.

AUTHOR CONTRIBUTIONS

GJ, EH, and HJ were responsible for the study conception and design. RV and GBT supervised the statistical analysis. GJ, EH, VS, AT, RV, GBT, and HJ worked collectively on interpretation of data. GJ was responsible for drafting the manuscript and all the co-authors provided ideas for improvement.

ACKNOWLEDGEMENTS

First and foremost, we thank the data abstractors Andrea Jona Eggertsdottir, Berglind Osk Olafsdottir and Mona Sif Hadaya. We also owe gratitude to Marianne E. Klinker for important insights into designing and developing the DCT. The manuscript has been proof-read by Anchor English.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

FUNDING INFORMATION

This study was supported by the grants received by GJ from the following institutions: The Icelandic Nurses' Association; The University of Iceland and Landspítali, The National University Hospital of Iceland. The funding sources were not involved in the data analysis and interpretation of results.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICAL APPROVAL

The Ethics Committee at Landspítali National University Hospital approved the study with the number #26/2017. No identifiable patient information was used, thus, informed consent was not needed.

ORCID

Gudrun Jonsdottir  <https://orcid.org/0000-0003-2194-3924>

Erna Haraldsdottir  <https://orcid.org/0000-0003-4891-0743>

Helga Jonsdottir  <https://orcid.org/0000-0002-3295-869X>

REFERENCES

- Alonso, A., Ebert, A. D., Dörr, D., Buchheidt, D., Hennerici, M. G., & Szabo, K. (2016). End-of-life decisions in acute stroke patients: An observational cohort study. *BMC Palliative Care*, 15, 38. <https://doi.org/10.1186/s12904-016-0113-8>
- Bruera, S., Chisholm, G., Dos Santos, R., Crovador, C., Bruera, E., & Hui, D. (2014). Variations in vital signs in the last days of life in patients with advanced cancer. *Journal of Pain and Symptom Management*, 48(4), 510–517. <https://doi.org/10.1016/j.jpain.2013.10.019>
- Cassidy, L. D., Marsh, G. M., Holleran, M. K., & Ruhl, L. S. (2002). Methodology to improve data quality from chart review in the managed care setting. *The American Journal of Managed Care*, 8(9), 787–793.
- Cheng, H. W. B., Chan, O. M. I., Chan, C. H. R., Chan, W. H., Fung, K. S., & Wong, K. Y. (2017). End-of-life characteristics and palliative care provision for patients with motor neuron disease. *American Journal of Hospice and Palliative Medicine*, 35(6), 847–851. <https://doi.org/10.1177/1049909117735832>
- Cohen-Mansfield, J., Skornick-Bouchbinder, M., & Brill, S. (2018). Trajectories of end of life: A systematic review. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 73(4), 564–572. <https://doi.org/10.1093/geronb/gbx093>
- Cox, K., Moghaddam, N., Almack, K., Pollock, K., & Seymour, J. (2011). Is it recorded in the notes? Documentation of end-of-life care and preferred place to die discussions in the final weeks of life. *BMC Palliative Care*, 10, 18. <https://doi.org/10.1186/1472-684X-10-18>
- Eriksson, H., Milberg, A., Hjelm, K., & Friedrichsen, M. (2016). End of life care for patients dying of stroke: A comparative registry study of stroke and cancer. *PLoS One*, 11(2), e0147694. <https://doi.org/10.1371/journal.pone.0147694>
- Feder, S. L. (2017). Data quality in electronic health records research: Quality domains and assessment methods. *Western Journal Nursing Research*, 40(5), 753–766.
- Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. the problems of two paradoxes. *Journal of Clinical Epidemiology*, 43(6), 543–549. [https://doi.org/10.1016/0895-4356\(90\)90158-1](https://doi.org/10.1016/0895-4356(90)90158-1)
- Fernando, G., & Hughes, S. (2019). Team approaches in palliative care: A review of the literature. *International Journal of Palliative Nursing*, 25(9), 444–451.
- Forde-Johnston, C., Butcher, D., & Aveyard, H. (2022). An integrative review exploring the impact of electronic health records (EHR) on the quality of nurse-patient interactions and communication. *Journal of Advanced Nursing*, 00, 1–20. <https://doi.org/10.1111/jan.15484>
- Fortney, C. A., & Steward, D. K. (2015). Medical record documentation and symptom management at the end of life in the NICU. *Advances in Neonatal Care: Official Journal of the National Association of Neonatal Nurses*, 15(1), 48–55. <https://doi.org/10.1097/ANC.0000000000000132>
- Gianinazzi, M. E., Rueegg, C. S., Zimmerman, K., Kuehni, C. E., Michel, G., & Swiss Paediatric Oncology Group. (2015). Intra-rater and inter-rater reliability of a medical record abstraction study on transition of care after childhood cancer. *PLoS One*, 10(5), e0124290. <https://doi.org/10.1371/journal.pone.0124290>
- Gonella, S., Basso, I., Clari, M., Dimonte, V., & Di Giulio, P. (2020). A qualitative study of nurses' perspective about the impact of end-of-life communication on the goal of end-of-life care in nursing homes. *Scandinavian Journal of Caring Sciences*, 35, 502–511. <https://doi.org/10.1111/scs.12862>
- Gregory, K. E., & Radovinsky, L. (2012). Research strategies that result in optimal data collection from the patient medical record. *Applied Nursing Research: ANR*, 25(2), 108–116. <https://doi.org/10.1016/j.apnr.2010.02.004>
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42(2), 377–381. <https://doi.org/10.1016/j.jbi.2008.08.010>
- Hight, G., Crawford, D., Murray, S. A., & Boyd, K. (2014). Development and evaluation of the supportive and palliative care indicators tool (SPICT): A mixed-methods study. *BMJ Supportive & Palliative Care*, 4(3), 285–290. <https://doi.org/10.1136/bmjspcare-2013-000488>
- Hui, D., dos Santos, R., Chisholm, G., Bansal, S., Silva, T. B., Kilgore, K., Crovador, C. S., Yu, X., Swartz, M. D., Perez-Cruz, P. E., Leite, R., Nascimento, M. S., Reddy, S., Seriacco, F., Yennu, S., Paiva, C. E., Dev, R., Hall, S., Fajardo, J., & Bruera, E. (2014). Clinical signs of impending death in cancer patients. *The Oncologist*, 19(6), 681–687. <https://doi.org/10.1634/theoncologist.2013-0457>
- Hui, D., Dos Santos, R., Chisholm, G., Bansal, S., Souza Crovador, C., & Bruera, E. (2015). Bedside clinical signs associated with impending death in patients with advanced cancer: Preliminary findings of a prospective, longitudinal cohort study. *Cancer*, 121(6), 960–967. <https://doi.org/10.1002/cncr.29048>
- Huskamp, H. A., Kaufmann, C., & Stevenson, D. G. (2012). The intersection of long-term care and end-of-life care. *Medical Care Research and Review*, 69(1), 3–44. <https://doi.org/10.1177/1077558711418518>
- Hussain, J., Allgar, V., & Oliver, D. (2018). Palliative care triggers in progressive neurodegenerative conditions: An evaluation using a multi-Centre retrospective case record review and principal component analysis. *Palliative Medicine*, 32(4), 716–725. <https://doi.org/10.1177/0269216318755884>
- Jansen, A. C., van Aalst-Cohen, E. S., Hutten, B. A., Büller, H. R., Kastelein, J. J., & Prins, M. H. (2005). Guidelines were developed for data collection from medical records for use in retrospective analyses. *Journal of Clinical Epidemiology*, 58(3), 269–274. <https://doi.org/10.1016/j.jclinepi.2004.07.006>
- Kaji, A. H., Schriger, D., & Green, S. (2014). Looking through the retro-spectroscope: Reducing bias in emergency medicine chart review studies. *Annals of Emergency Medicine*, 64(3), 292–298. <https://doi.org/10.1016/j.annemergmed.2014.03.025>
- Kennedy, C., Brooks-Young, P., Brunton Gray, C., Larkin, P., Connolly, M., Wilde-Larsson, B., Larsson, M., Smith, T., & Chater, S. (2014). Diagnosing dying: An integrative literature review. *BMJ Supportive & Palliative Care*, 4(3), 263–270. <https://doi.org/10.1136/bmjspcare-2013-000621>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Landspítali (Ed.). (2017). *Klínískar leiðbeiningar um líknarmæðferð (e. Palliative care guidelines)* (2nd ed.). Landspítali. <https://www.landspitali.is/library/Sameiginlegar-skrar/Gagnasafn/BRUNNURINN/Klínískar-leiðbeiningar/Líknarmæðferd/Líknarmæðferd.pdf>
- LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational*

- Research Methods*, 11(4), 815–852. <https://doi.org/10.1177/1094428106296642>
- Liu, Y., Kline, D., Aerts, S., Youngwerth, J. M., Kutner, J. S., Sillau, S., & Kluger, B. M. (2017). Inpatient palliative care for neurological disorders: Lessons from a large retrospective series. *Journal of Palliative Medicine*, 20(1), 59–64. <https://doi.org/10.1089/jpm.2016.0240>
- Mazzocato, C., Michel-Nemitz, J., Anwar, D., & Michel, P. (2010). The last days of dying stroke patients referred to a palliative care consult team in an acute hospital. *European Journal of Neurology*, 17(1), 73–77. <https://doi.org/10.1111/j.1468-1331.2009.02744.x>
- McCusker, M., Ceronisky, L., Crone, C., Epstein, H., Greene, B., Halvorson, J., Kephart, K., Mallen, E., Nosan, B., Rohr, M., Rosenberg, E., Ruff, R., Schlecht, K., & Setterlund, L. (2013). *Palliative care for adults*. Institute for Clinical Systems Improvement (ICSI).
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282.
- Munroe, C. A., Sirdofsky, M. D., Kuru, T., & Anderson, E. D. (2007). End-of-life decision making in 42 patients with amyotrophic lateral sclerosis. *Respiratory Care*, 52(8), 996–999.
- National Clinical Guideline Centre (NICE). (2015). *Care of dying adults in the last days of life*. National Institute for Health and Care Excellence (UK). <https://www.nice.org.uk/guidance/ng31>
- Ntlholang, O., Walsh, S., Bradley, D., & Harbison, J. (2016). Identifying palliative care issues in inpatients dying following stroke. *Irish Journal of Medical Science*, 185(3), 741–744. <https://doi.org/10.1007/s11845-015-1290-9>
- Polit, D. F., & Beck, C. T. (2017). *Nursing research: Generating and assessing evidence for nursing practice* (10th ed.). Wolters Kluwer Health.
- Quadri, S. Z., Huynh, T., Cappelen-Smith, C., Wijesuriya, N., Mamun, A., Beran, R. G., McDougall, A. J., & Cordato, D. (2018). Reflection on stroke deaths and end-of-life stroke care. *Internal Medicine Journal*, 48(3), 330–334. <https://doi.org/10.1111/imj.13619>
- Registered Nurses' Association of Ontario (RNAO). (2011). *End-of-life care during the last days and hours*. Registered Nurses' Association of Ontario (RNAO).
- Sinuff, T., Dodek, P., You, J. J., Barwich, D., Tayler, C., Downar, J., Hartwick, M., Frank, C., Stelfox, H. T., & Heyland, D. K. (2015). Improving end-of-life communication and decision making: The development of a conceptual framework and quality indicators. *Journal of Pain and Symptom Management*, 49(6), 1070–1080. <https://doi.org/10.1016/j.jpainsymman.2014.12.007>
- Sjöberg, M., Edberg, A. K., Rasmussen, B. H., & Beck, I. (2021). Documentation of older people's end-of-life care in the context of specialized palliative care: A retrospective review of patient health records. *BMC Palliative Care*, 20(1), 91. <https://doi.org/10.1186/s12904-021-00771-w>
- Taffurelli, C., Barelo, S., Cervantes Camacho, V., Bertuol, M., Savarese, M., & Artioli, G. (2020). Taking care of dying patients through an 'interprofessional ecosystem': A grounded theory study on the experience of an interprofessional team in palliative care. *Scandinavian Journal of Caring Sciences*, 35, 1169–1178. <https://doi.org/10.1111/scs.12934>
- To, T., Estrabillo, E., Wang, C., & Cicutto, L. (2008). Examining intra-rater and inter-rater response agreement: A medical chart abstraction study of a community-based asthma care program. *BMC Medical Research Methodology*, 8, 29. <https://doi.org/10.1186/1471-2288-8-29>
- Wang, V., Hsieh, C. C., Huang, Y. L., Chen, C. P., Hsieh, Y. T., & Chao, T. H. (2018). Different utilization of intensive care services (ICs) for patients dying of hemorrhagic and ischemic stroke, a hospital-based survey. *Medicine*, 97(8), e0017. <https://doi.org/10.1097/MD.00000000000010017>
- Williams, M. T., Zimmerman, E., Barry, M., Trantum, L., Dietrich, M. S., Doersam, J. K., & Karlekar, M. (2019). A retrospective review of patients with acute stroke with and without palliative care consultations. *The American Journal of Hospice & Palliative Care*, 36(1), 60–64. <https://doi.org/10.1177/1049909118787136>
- Yawn, B. P., & Wollan, P. (2005). Interrater reliability: Completing the methods description in medical records review studies. *American Journal of Epidemiology*, 161(10), 974–977. <https://doi.org/10.1093/aje/kwi122>
- Zaiontz, C. (2019). *Fleiss' Kappa. Real Statistics Using Excel*. <http://www.real-statistics.com/reliability/fleiss-kappa/>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Jonsdottir, G., Haraldsdottir, E., Sigurdardottir, V., Thoroddsen, A., Vilhjalmsson, R., Tryggvadottir, G. B., & Jonsdottir, H. (2023). Developing and testing inter-rater reliability of a data collection tool for patient health records on end-of-life care of neurological patients in an acute hospital ward. *Nursing Open*, 00, 1–9. <https://doi.org/10.1002/nop2.1789>