Interobserver agreement in a checklist of care in enteral nutrition therapy

Concordância interobservadores em um *checklist* de cuidados em terapia nutricional enteral

Concordancia interobservadores en una *checklist* de cuidados en terapia nutricional enteral

Ana Paula Almeida Corrêa[1] https://orcid.org/0000-0001-8890-1767
Stella Marys Rigatti Silva[1] https://orcid.org/0000-0002-4124-519X
Camila Camargo Oleques[1] https://orcid.org/0000-0002-7481-1581
Gabriele Peres de Sousa[1] https://orcid.org/0000-0002-9330-0234
Graziela Lenz Viegas[1] https://orcid.org/0000-0001-7093-7470
Franciele Anziliero[1] https://orcid.org/0000-0002-5650-9709
Adriana Catarina de Souza Oliveira[2] https://orcid.org/0000-0001-8600-4413
Mariur Gomes Beghetto[1] https://orcid.org/0000-0002-9437-4999

Como citar:
Corrêa AP, Silva SM, Oleques CM, Sousa GP, Viegas GL, Anziliero F, et al. Interobserver agreement in a checklist of care in enteral nutrition therapy. Acta Paul Enferm. 2021;34:eAPE001525.

DOI
http://dx.doi.org/10.37689/acta-ape/2021A0001525

Keywords
Nutrition therapy; Enteral nutrition; Checklist; Bias

Descritores
Terapia nutricional; Nutrição enteral; Lista de checagem; Viés

Descriptores
Terapia nutricional; Nutrición enteral; Lista de verificación; Sesgo

Submitted
June 20, 2020

Accepted
March 1, 2021

Corresponding author
Ana Paula Almeida Corrêa
Email: anacorrea.enf@gmail.com

Abstract

Objective: To assess interobserver agreement in the application of a checklist of care in enteral nutritional therapy (ENT).

Methods: This is a reliability study that preceded a clinical trial (NCT03497221), carried out at a university hospital in southern Brazil (June and July 2017). A checklist of 25 items related to care in ENT was performed by a nurse (reference standard) and nine research assistants (RA). Assessments were carried out concurrently and independently. Agreement was tested using the Statistical Package for the Social Sciences, version 21.0. Kappa values (k) were considered as poor (0 to 0.19), relative (0.20 to 0.39), moderate (0.40 to 0.59), substantial (0.60 and 0.79), almost perfect (0.80 to 0.99), and perfect (1). The study was approved by an Institutional Review Board (number 16-0534).

Results: Three hundred fifty-one observations were made in duplicate; the lowest number of observations was with RA, 5 (n = 35) and the highest with RA, 8 (n = 45). Items related to ENT were assessed in three blocks: identification of infusion bottles and infusion pump; support materials for administering the therapy; care for patients using ENT. There was almost perfect or perfect agreement in all observation pairs, with lowest Kappa for RA 6 (k = 0.890; 95% CI = 0.86, 0.92) and the highest for RA 3 (k = 0.965; 95% CI = 0.93, 0.99).

Conclusion: Interobserver agreement, when applying a checklist containing 25 items, was excellent, which minimizes the occurrence of measurement bias in subsequent steps.

Resumo

Objetivo: Avaliar a concordância interobservadores na aplicação um *checklist* de cuidados em Terapia Nutricional Enteral (TNE).

Métodos: Estudo de confiabilidade que precedeu um ensaio clínico (NCT03497221), realizado em hospital universitário do sul do Brasil (junho e julho de 2017). C*hecklist* de 25 itens relacionado aos cuidados em TNE foi realizado por uma enfermeira (Padrão de Referência) e por nove Assistentes de Pesquisa (AP). As avaliações foram feitas concomitantemente e de modo independente. A concordância foi testada utilizando-se o *Statistical Package for the Social Sciences* versão 21.0. Valores de Kappa (k) foram considerados como concordância entre: pobre (0 a 0,19); relativa (0,20 a 0,39); moderada (0,40 a 0,59); substancial (0,60 e 0,79); quase perfeita (0,80 a 0,99); e perfeita (1). O estudo foi aprovado pelo Comitê de Ética da Instituição (nº 16-0534).

Resultados: Foram realizadas 351 observações em duplicata, sendo o menor número de observações foi com a AP 5 (n=35) e o maior com a AP 8 (n=45). Foram avaliados itens relacionados a TNE em três blocos: identificação dos frascos de infusões e bomba de infusão; materiais de apoio para administração da terapia;

[1]Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.
[2]Universidad Católica San Antonio de Murcia, Murcia, Spain.
Conflicts to interest: nothing to declare.

e cuidados ao paciente em uso de TNE. Houve concordância quase perfeita ou perfeita em todos os pares de observação, com o menor Kappa para AP 6 (k=0,890; IC95%=0.86, 0.92) e o maior para AP 3 (k=0,965; IC95%=0.93, 0.99).

Conclusão: A concordância interobservadores ao aplicar um *checklist* contendo 25 itens foi excelente, o que minimiza a ocorrência de viés de aferição nas etapas subsequentes.

Resumen

Objetivo: Evaluar la concordancia interobservadores en la aplicación de una *checklist* de cuidados en terapia nutricional enteral (TNE).

Métodos: Estudio de fiabilidad que precedió un ensayo clínico (NCT03497221), realizado en un hospital universitario de la región Sur de Brasil (junio y julio de 2017). Una checklist de 25 ítems relacionados con los cuidados en TNE fue realizada por una enfermera (Estándar de Referencia) y por nueve Asistentes de Investigación (AI). Las evaluaciones fueron llevadas a cabo simultánea e independientemente. La concordancia se comprobó utilizando el *Statistical Package for the Social Sciences* versión 21.0. Los valores de Kappa (k) se consideraron como concordancia entre: pobre (0 a 0,19); relativa (0,20 a 0,39); moderada (0,40 a 0,59); considerable (0,60 a 0,79); casi perfecta (0,80 a 0,99); y perfecta (1). El estudio fue aprobado por el Comité de Ética de la institución (n.° 16-0534).

Resultados: Se realizaron 351 observaciones duplicadas, de las cuales el menor número de observaciones fue de la AI 5 (n=35) y el mayor de la AI 8 (n=45). Se evaluaron ítems relacionados con la TNE en tres grupos: identificación de los frascos de infusiones y bomba de infusión, material de apoyo para la administración de la terapia y cuidados del paciente en uso de TNE. Se observó concordancia casi perfecta o perfecta en todos los pares de observación, con el menor Kappa de la AI 6 (k=0,890; IC95 %=0.86, 0.92) y el mayor de la AI 3 (k=0,965; IC95 %=0.93, 0.99).

Conclusión: La concordancia interobservadores al aplicar una *checklist* de 25 ítems fue excelente, lo que minimiza la ocurrencia de sesgo de medición en las etapas subsiguientes.

Introduction

In clinical studies, the reliability in the application of instruments is influenced by the variability of the subjects. To minimize possible measurement biases due to this variety, agreement assessment is applied, which can be for the same individual or phenomenon, by evaluators, at different times or by different instruments or by a set of these situations so that the results obtained are more equal.[1] To do this, measures must be taken to minimize the occurrence of bias and ensure data reliability, among which, those related to the measurement of the study variables are equipment calibration, standardization of methods and team training and certification.[2]

The interobserver agreement method was applied in some studies that used clinical assessment of subjects[3,4] or analysis by images of diagnostic tests, whether they are ultrasound, magnetic resonance imaging (MRI), computed tomography (CT) and/or radiography (X-ray) for the identification of nodules, masses, tumors and fractures, with a difference in the degree of experience of the evaluators.[5–8] Still considering agreement through images, there are also some studies that use visual aids both with the naked eye and telescopically to assess the degree of pressure injuries or to quantify the evolution of decomposition of corpses.[9,10] Corroborating the importance of applying these studies in different areas, one study states that inter-rater reliability was essential to guarantee the quality of the process, just as the practice must be routine in studies for due transparency in conducting research.[11]

However, it is known that these studies are still more widespread in the medical field, being pioneers in the field of nursing and infrequent with regard more specifically to enteral nutritional therapy (ENT).[12-17] It is of paramount importance that multidisciplinary health teams also develop and disseminate agreement studies, in order to assist researchers in their clinical studies as a tool to ensure the reliability and reproducibility of the data obtained.[18] Thus, agreement assessment is justified as a way to guarantee the "calibration" of those who assess the independent and dependent variables in question.[2]

In addition to the methodological issue, agreement studies can be the basis for implementing care protocols that help health professionals to act more safely in patient care, and it is known that for the application of instruments, the interobserver agreement assessment is a step that allows establishing its reproducibility.[19] At the level of ENT, although it is known that the use of adequate protocols is recommended to promote faster and safer recovery of patients,[20] the development of studies is scarce in this domain, remaining more at the level of assessment of food surveys, anthropometric measures and reliability of tests that confirm the probe positioning.[8,21]

Thus, the objective of this study, as a step that preceded a clinical trial, was to assess interobserver agreement in the application of a checklist of care in ENT of hospitalized adults using a nasoenteral tube (ENT).

Methods

This is a reliability study that took place at a university hospital of high complexity in southern Brazil. It is a preliminary stage of a clinical trial approved by an Institutional Review Board (nº16-0534) and registered in Clinical Trials (NCT03497221).

In June and July 2017, adults (aged 18 years or older) were assessed using ENT from four inpatient units (two clinical and two surgical), except for confused and/or disoriented people who were without a companion, of which it was not possible to obtain the consent form. The identification of patients using a diet by ENT occurred through an Enteral Feeding Center's diet map, which consists of a list of all hospital enteral diet users that is generated from an institution's information system that integrates medical and nutritionist prescriptions.

All assessments were carried out by two evaluators at the same time, which were carried out by a nurse (reference standard), hospital worker and doctoral student, while the assessment carried out by each of the nine undergraduate students in nursing was considered "test assessment", i.e., tested as concordant or not in relation to the assessment carried out by the reference nurse.

Prior to the agreement stage, it was envisaged that research assistants (RA) would be previously trained in order to standardize data collection. These were directly trained and supervised by the nurse responsible for the study for a period of three months before the start of the agreement stage. Training was carried out at the bedside with the same checklist that was applied to a larger study, aiming to standardize: (a) inviting patients and obtaining consent; (b) data collection; (c) assessment and monitoring of the study variables; (d) records on the survey forms. Guidance manuals were prepared, which remained available for consultation by the data collection team throughout the training period and afterwards. These manuals were intended to standardize the collection and filling in of the instrument's variables.

It was chosen to train nursing students from the 5th to the 8th semesters, since all had gone through practical internships at the hospital and because it was an exclusively operational collection. In this first stage of agreement, we considered that what was being tested was the standardization of data collection, in order not to alter the results obtained after the intervention performed in the clinical trial. The selection of students took place based on the disclosure in local universities of the opportunity to participate as a volunteer fellow linked to the larger project, with subsequent analysis of the curriculum and interview.

The observation of care for patients using ENT consisted of the application of a checklist containing 25 items, based on the institution's Standard Operational Protocols (SOPs) for care in ENT, which follow the guidelines of Resolution 63/2000, especially with regard to the inspection script for management activities in ENT, for which there were three possibilities of response: (1) complied, (2) did not comply and (3) did not apply (when it was not possible to observe). The sets of variables in the checklist, used to assess interobserver agreement, were divided into three categories: a) infusions (diets and water) and infusion pump; b) support materials in nutritional therapy administration; c) care for patients at the bedside.[22]

All assessments took place independently, in subsequent moments, and patients and their companions were instructed not to issue comments during the assessments, in order to ensure the blindness of the evaluators to the opposite assessment. Data were collected using a form made up of a checklist at the bedside, using cell phones connected to the internet and an instrument developed using Google Forms®. These data were automatically transferred to a Google Sheets® spreadsheet, and later exported to a Microsoft Excel® spreadsheet, where they were treated and coded by the nurse responsible for the study for later statistical analysis.

Agreement among evaluators was tested by obtaining the Kappa coefficient (k) and its confidence

intervals (95%). The "k" was calculated by comparing assessments performed by the "standard reference" nurse to the others, collected by nine RA. Kappa values (k) were considered as agreement: poor (between 0 and 0.19); relative (0.20 to 0.39); moderate (0.40 to 0.59); substantial (0.60 and 0.79); almost perfect (0.80 to 0.99); perfect (1).[20] For data analysis, the Statistical Package for the Social Sciences - SPSS®, version 21.0 was used. Additionally, for agreement analysis (Kappa and 95% CI), WinPEPI4® and Single Case Research® were used.

Ethical considerations were respected, applying the Informed Consent Form to all subjects, and the Term of Commitment for Data Use was signed by all researchers and RA.[23]

Results

The Kappa coefficient (k) was calculated by comparing assessments made by the nurse who authored the present thesis ("reference standard") to those collected by nine RA. A total of 351 observations were made in duplicate, with the lowest number occurring with RA 5 (n = 35) and the highest with RA 8 (n = 45). The agreement between observers was almost perfect among all pairs, showing better with RA 3 (k = 0.965) and slightly worse with RA 6 (k = 0.890), according to chart 1. Thus, through the analysis of the data, it was contacted that the initial training was effective. However, in order to identify whether there was any discrepancy specifically, all items were assessed individually and when the agreement was very different, the items with less agreement were taken up with each RA specifically.

When assessing agreement among observers in each of the checklist items related to the identification of infusion bottles (diet and water) and to infusion pump use, it was found that it was almost perfect among all pairs of observations. Regarding the observations that assessed patients' identification or the validity of the diet and water bottles, agreement among pairs was perfect or almost perfect. In the item that assessed infusion pump dirtiness, which was more subject to subjectivity, agreement assessment was predominantly substantial (Chart 1).

Agreement among observers in the assessment of items related to the conditions (presence of dirt, identification and validity) of the materials and devices (diet equipment, diet syringes and disposable plastic cups) used for ENT administration and maintenance was shown to be perfect or almost perfect for most assessed items. Except for some isolated verification items, such as agreement between RA 1 in the item protection of the diet equipment with cover while it was not being used, which demonstrates that agreement was low in isolation. This is repeated with one or another evaluator on specific items, suggesting non-systematic errors of interpretation (Chart 2).

When assessing agreement among observers on the items related to direct care for patients (conditions of ENT fixation and headboard position for ENT administration), there was perfect or almost perfect agreement only for the most objective item, which assessed the date of fixing the probe. However, regarding the assessment of more subjective items, such as presence of dirt, oil, traction or detachment of the fixation, with the exception of some RA that had perfect or almost perfect agreement for some of these items, the comparison in most cases was weak, relative, or moderate (Chart 3).

Discussion

Interobserver agreement in a checklist of care in ENT, being performed by a reference nurse and nine RA, in general, showed excellent agreement. It is noteworthy that the objective observations between pairs were perfect or almost perfect. In some isolated items, especially those with greater subjectivity, this agreement was low. A reliability study conducted in Brazil, with the objective of analyzing the use of the Manchester Screening System, measured agreement with the application of clinical cases to nurses. They were previously trained, as occurred in the present study, and nurses who underwent theoretical assessment using 60% or more were considered able to participate. Reliability ranged from moderate to substantial, with Kappa values between 0.55 and 0.72 ($p < 0.001$) and between 0.57 and 0.78 ($p < 0.05$).

Chart 1. Agreement among a nurse (reference standard) and nine RA in the checklist verification items regarding the identification of infusion bottles (diets and water) and the infusion pump used in ENT. Data were expressed by Kappa coefficient (k) values and its 95% CI

Checklist items	RA 1 n=38 k (CI): 0.89 (0.86– 0.92)	RA 2 n=38 k (CI): 0.91 (0.86– 0.92)	RA 3 n=42 k (CI): 0.96 (0.93– 0.99)	RA 4 n=41 k (CI): 0.95 (0.92– 0.97)	RA 5 n=35 k (CI): 0.93 (0.90– 0.96)	RA 6 n=41 k (CI): 0.89 (0.86– 0.92)	RA 7 n=39 k (CI): 0.94 (0.92–0.97)	RA 8 n=45 k (CI): 0.94 (0.92– 0.97)	RA 9 n=38 k (CI): 0.94 (0.92– 0.98)
Identification of the enteral diet bottle being administered									
Same as the patients' bracelet	0.79 (0.69– 0.99)	0.89 (0.77– 1 *)	0.73 (0.66– 0.93)	1.00 (0.85– 1*)	0.94 (0.81– 1*)	1.00 (0.85– 1*)	0.94 (0.81– 1*)	0.95 (0.82– 1*)	0.94 (0.81– 1*)
Expiration (up to 3h)	1.00 (0.85– 1*)	0.91 (0.77– 1*)	0.89 (0.75– 1*)	0.91 (0.78– 1*)	0.94 (0.81– 1*)	1.00 (0.85– 1*)	0.95 (0.81– 1*)	0.96 (0.82– 1*)	0.95 (0.81– 1*)
Identification of the water bottle for cleaning utensils and ENT									
Same as the patients' bracelet	0.81 (0.77– 1*)	1.00 (0.85– 1*)	0.80 (0.60– 0.89)	1.00 (0.85– 1*)	0.95 (0.77– 1*)	1.00 (0.85– 1*)	1.00 (0.85– 1*)	1.00 (0.86– 1*)	1.00 (0.85– 1*)
Expiration (24h)	0.94 (0.81– 1*)	1.00 (0.85– 1*)	1.00 (0.85– 1*)	1.00 (0.85– 1*)	0.91 (0.77– 1*)	0.95 (0.81– 1*)	1.00 (0.85– 1*)	1.00 (0.86– 1*)	1.00 (0.85– 1*)
Identification of the water bottle for hydration									
Same as the patients' bracelet	1.00 (0.85–1*)	1.00 (0.85– 1*)	0.25 (0.7- 0.46)	1.00 (0.85– 1*)	0.90 (0.77– 1*)	0.94 (0.81– 1*)	1.00 (0.85– 1*)	1.00 (0.86– 1*)	1.00 (0.85– 1*)
Expiration (24h)	1.00 (0.85– 1*)	1.00 (0.85– 1*)	0.94 (0.82– 1*)	1.00 (0.85– 1*)	0.89 (0.77– 1*)	0.85 (0.74– 1*)	1.00 (0.85– 1*)	1.00 (0.86– 1*)	1.00 (0.85– 1*)
Identification of the infusion pump used for ENT									
No dirt	0.93 (0.81– 1*)	0.82 (0.77– 1*)	0.92 (0.75– 1*)	0.86 (0.78– 1*)	0.87 (0.77– 1*)	0.74 (0.78– 1*)	0.92 (0.81– 1*)	0.87 (0.79 - 1*)	0.91 (0.81– 1*)

RA - Research Assistant; k - Kappa coefficient - from 0 to 0.19 = poor agreement; 0.20-0.39 = relative agreement; 0.40-0.59 = moderate agreement; 0.60-0.79 = substantial agreement; 0.80-0.99 = almost perfect agreement; 1, 00 = perfect agreement; CI - Confidence Interval (*Mathematically calculated confidence interval values> 1, considered = 1)

Chart 2. Agreement among a nurse (reference standard) and nine RA in the checklist verification items referring to support materials in ENT administration. Data were expressed by Kappa coefficient (k) values and its 95% Confidence Interval (CI)

Checklist items	RA 1 n=38 k (CI): 0.89 (0.86– 0.92)	RA 2 n=38 k (CI): 0.91 (0.86– 0.92)	RA 3 n=42 k (CI): 0.96 (0.93– 0.99)	RA 4 n=41 k (CI): 0.95 (0.92– 0.97)	RA 5 n=35 k (CI): 0.93 (0.90– 0.96)	RA 6 n=41 k (CI): 0.89 (0.86– 0.92)	RA 7 n=39 k (CI): 0.94 (0.92–0.97)	RA 8 n=45 k (CI): 0.94 (0.92– 0.97)	RA 9 n=38 k (CI): 0.94 (0.92– 0.98)
Identification of the equipment used for ENT administration (blue)									
At the bedside	1.00 (0.85– 1*)	1.00 (0.85– 1*)	0.73 (0.66– 0.93)	1.00 (0.85– 1*)	1.00 (0.85– 1*)	0.84 (0.70– 0.99)	1.00 (0.85– 1*)	1.00 (0.86– 1*)	1.00 (0.85– 1*)
Expiration (24h)	0.88 (0.77–1*)	0.90 (0.77– 1*)	0.93 (0.78– 1*)	0.89 (0.78– 1*)	1.00 (0.85– 1*)	0.94 (0.81. 1*)	0.93 (0.81– 1*)	0.94 (0.82– 1*)	1.00 (0.85– 1*)
No diet dirt	0.95 (0.81–1 *)	0.80 (0.69– 0.99)	0.94 (0.67– 0.96)	0.96 (0.81– 1*)	0.95 (0.77– 1*)	0.65 (0.52– 0.81)	0.67 (0.54– 0.84)	0.96 (0.82– 1*)	0.90 (0.77– 1*)
Identification of the tip of the blue equipment for ENT protected with cover									
When not infusing the diet	0.57 (0.45-0.75)	0.76 (0.69– 0.99)	1.00 (0.67– 0.96)	0.87 (0.78– 1 *)	0.94 (0.81– 1 *)	0.93 (0.90– 0.96)	1.00 (0.85– 1 *)	0.94 (0.82– 1 *)	1.00 (0.85– 1 *)
Identification of the syringe used for ENT (Oralpak®)									
At the bedside	1.00 (0.85–1 *)	1.00 (0.85– 1 *)	1.00 (0.85– 1 *)	1.00 (0.85– 1 *)	1.00 (0.85– 1 *)	1.00 (0.85– 1 *)	0.94 (0.81– 1 *)	1.00 (0.86– 1 *)	1.00 (0.85– 1 *)
Expiration (6h)	1.00 (0.85–1 *)	0.86 (0.77– 1 *)	0.90 (0.78– 1 *)	0.94 (0.81– 1 *)	0.89 (0.77– 1 *)	0.94 (0.81– 1 *)	0.87 (0.73– 1 *)	1.00 (0.86– 1 *)	0.93 (0.90– 0.96)
Clean and without residues	0.90 (0.77–1 *)	0.90 (0.77– 1 *)	1.00 (0.75– 1*)	0.79 (0.67– 0.96)	0.91 (0.77– 1 *)	0.66 (0.56– 0.85)	0.96 (0.81– 1 *)	0.96 (0.82– 1 *)	1.00 (0.85– 1 *)
Labeled (name and medical record number)	0.95 (0.81–1 *)	1.00 (0.85– 1 *)	0.95 (0.78– 1 *)	0.96 (0.81– 1 *)	0.85 (0.70– 0.99)	0.92 (0.78– 1 *)	0.96 (0.81– 1 *)	1.00 (0.86– 1 *)	1.00 (0.85– 1 *)
Labeled (bed, date and shift)	0.87 (0.77–1 *)	0.94 (0.81– 1 *)	1.00 (0.75– 1 *)	0.94 (0.81– 1 *)	0.78 (0.70– 0.99)	0.95 (0.81– 1 *)	0.95 (0.81– 1 *)	0.95 (0.82– 1 *)	1.00 (0.85– 1 *)
Identification of the plastic cup used to sanitize devices used in ENT									
Expiration (6h)	0.94 (0.81–1 *)	0.89 (0.77– 1 *)	1.00 (0.82– 1 *)	1.00 (0.85– 1 *)	1.00 (0.85– 1 *)	0.95 (0.74– 1 *)	1.00 (0.85– 1 *)	1.00 (0.86– 1 *)	0.94 (0.81– 1 *)
Clean and without residues	0.81 (0.73–1 *)	0.91 (0.77– 1 *)	0.94 (0.77– 1 *)	0.95 (0.81– 1 *)	1.00 (0.85– 1 *)	0.95 (0.81– 1 *)	0.90 (0.77– 1 *)	0.95 (0.82– 1 *)	0.95 (0.81– 1 *)
Dry	0.82 (0.69–0.99)	0.90 (0.77– 1 *)	1.00 (0.78– 1 *)	0.89 (0.74– 1 *)	0.95 (0.81– 1 *)	0.88 (0.74– 1 *)	1.00 (0.85– 1 *)	0.96 (0.82– 1 *)	0.90 (0.77– 1 *)

RA - Research Assistant; k - Kappa coefficient - from 0 to 0.19 = poor agreement; from 0.20 to 0.39 = relative agreement; from 0.40 to 0.59 = moderate agreement; 0.60 to 0.79 = substantial agreement; 0.80-0.99 = almost perfect agreement; 1.00 = perfect agreement; CI - Confidence Interval (*Mathematically calculated confidence interval values> 1, considered = 1

Chart 3. Agreement among a nurse (reference standard) and nine RA in the checklist verification items related to care at the bedside with ENT. Data were expressed by Kappa coefficient (k) values and its 95% Confidence Interval (CI)

Checklist items	RA 1 n=38 k (IC): 0.89 (0.86– 0.92)	RA 2 n=38 k (IC): 0.91 (0.86– 0.92)	RA 3 n=42 k (IC): 0.96 (0.93– 0.99)	RA 4 n=41 k (IC): 0.95 (0.92– 0.97)	RA 5 n=35 k (IC): 0.93 (0.90– 0.96)	RA 6 n=41 k (IC): 0.89 (0.86– 0.92)	RA 7 n=39 k (IC): 0.94 (0.92–0.97)	RA 8 n=45 k (IC): 0.94 (0.92– 0.97)	RA 9 n=38 k (IC): 0.94 (0.92– 0.98)
Observation and identification of the fixation of ENT related to conservation and validity									
Clean	0.76 (0.69– 0.99)	0.64 (0.77– 1*)	0.80 (0.60– 0.89)	0.87 (0.81– 1*)	1.00 (0.85– 1*)	0.45 (0.70– 0.99)	0.89 (0.81– 1*)	0.73 (0.66– 0.93)	0.28 (0.69– 0.99)
No oil	0.69 (0.65– 0.95)	0.63 (0.57– 0.87)	0.81 (0.75– 1*)	0.57 (0.63– 0.92)	0.53 (0.47– 0.77)	0.20 (0.44– 0.72)	0.63 (0.58– 0.87)	0.61 (0.56– 0.83)	0.68 (0.61– 0.91)
Not detached	0.27 (0.57– 0.87)	0.47 (0.61– 0.91)	1.00 (0.82– 1*)	0.84 (0.78– 1*)	0.87 (0.81– 1*)	0.41 (0.63– 0.92)	0.87 (0.81– 1*)	0.85 (0.79– 1*)	0.47 (0.61– 0.91)
Not pulled	0.72 (0.77– 1*)	0.16 (0.61– 0.91)	1.00 (0.82– 1 *)	0.65 (0.81– 1 *)	0.70 (0.74– 1 *)	0.74 (0.78– 1 *)	0.65 (0.81– 1 *)	0.41 (0.62– 0.90)	0.44 (0.69– 0.99)
Dated (up to 24h previous)	1.00 (0.85–1*)	1.00 (0.85– 1 *)	1.00 (0.82– 1 *)	0.94 (0.81– 1 *)	1.00 (0.85– 1 *)	0.78 (0.70– 0.99)	0.94 (0.81– 1 *)	1.00 (0.86– 1 *)	1.00 (0.85– 1 *)
Observation and identification of the elevated head of patients' beds = > 30º									
In diet/water administration by ENT	0.42 (0.37– 0.67)	0.94 (0.81– 1 *)	0.89 (0.64– 0.92)	0.90 (0.78– 1 *)	0.83 (0.70– 0.99)	0.76 (0.67– 0.96)	0.85 (0.73– 1 *)	0.86 (0.78– 1 *)	0.94 (0.81– 1 *)

RA - Research Assistant; k - Kappa coefficient - from 0 to 0.19 = poor agreement; from 0.20 to 0.39 = relative agreement; from 0.40 to 0.59 = moderate agreement; 0.60-0.79 = substantial agreement; 0.80-0.99 = almost perfect agreement; 1.00 = perfect agreement; CI - Confidence Interval (*Mathematically calculated confidence interval values > 1, considered = 1)

[24] It is emphasized that our study had higher levels of agreement; however, there is a discrepancy in this comparison despite the theme. It is necessary further studies, of methodological value, aimed at nurses, as well as ENT, in order to improve future research in nursing and better clinical practices.

The previous training of RA, of different levels of professional training, was important for obtaining the excellent agreement found, especially considering the clinical or care-related inexperience established in the SOP of the institution in which the study occurred. Confirming this, a study, with the objective of determining the level of interobserver agreement in the classification of breast nodules by ultrasound, found that the level of agreement was higher among the most experienced radiologists. Agreement was made by three radiologists with different degrees of experience in breast imaging (15 years, eight years and two years), and agreement was regular to excellent (ICC=0.9503).[5] Another study corroborates this finding about professional experience, since there is greater agreement among nurses with more clinical experience (even after training), being significant (<0.001) (Kappa=0.51 for less than one year of experience versus Kappa=0.58 for more than 10 years of experience).[24]

Another study, with the objective of assessing the training demand and number of repetitions required to perform some imaging exams, to obtain greater interobserver agreement (Kappa>0.80), was conducted with 22 medical trainees and a standard reference evaluator. The Kappa index achieved in the second stage of training was k=0.80 for carotid, k=0.39 vertebral and k=0.54 ultrasound scans for transcranial Doppler, which allows us to conclude that a fixed training does not guarantee high interobserver agreement.[25] Although all RA in the present study received the same training and supervision before the assessment in duplicate, a slight difference in performance were observed. For some, the agreement compared to the reference standard was perfect, while others presented more disagreements in some items. This suggests that other conditions, in addition to training, such as the subjectivity involved in some checklist items, may affect the performance of observers.

There are few studies assessing interobserver agreement from the perspective of patient safety using enteral nutrition. As the present study that addresses this theme, a study assessed the between the use of auscultation test and X-ray in the detection of enteral probe positioning. Two nurses performed auscultation and gave opinions about the anatomical positioning of the probe. There was weak agreement among nurses (PABAK=0.054; p=0.103). Moreover, the agreement between the methods (auscultation and X-ray) was also very weak for both nurses (PABAK=0.188; p=0.111 and PABAK=0.128; p=0.107).[8] In fact, in

ENT, it is necessary to conduct studies that allow understanding the points of agreement and divergence involving the assessment and practices of nurses and nursing technicians in patient care.

It is observed that there was greater interobserver agreement in this study when the observed data had lower subjectivity, such as the expiration date of some item, since when the data required a certain interpretation, it was observed less agreement, as when the dirtiness of some device was assessed. As in our study, which used bedside observation of patients, a study assessed interobserver agreement in the clinical observation of focal and generalized epilepsy of 512 patients and found Kappa 0.91 (p <0.0001), which identifies high agreement. Unlike our study, clinical observation was considered reliable interobservers, using only epilepsy assessment according to the image provided on video. This demonstrates that when the concepts are well defined for an observer about the item assessed, such as what is considered generalized and focal epilepsy, interobserver agreement may be higher, even though it is an item that requires a certain interpretation.[3]

Also in the safety issue, we highlight that the observation and assessment of the fixation of the probe, even in isolated items, had cases of low or very low agreement. Literature demonstrates the importance of this assessment ascertained by health professionals who provide care to patients who use ENT, since an adequate fixation can avoid an accidental traction of the probe.[19] It is considered an important placement in this study, for some assessments of what was considered a good state of fixation, was very subjective for the different evaluators.

Conclusion

It is suggested that the high interobserver agreement found in this study was related to previous training as a methodology applied to minimize possible bias, succeeding greater methodological rigor and greater reliability of the results obtained. In this regard, there is a recommendation that the agreement among the different observers involved in data collection, especially in clinical studies, be assessed in order to minimize biases. It is encouraged that this is a step that precedes data collection from more studies of this type, especially by nursing, with a view to seeking greater reliability of the data obtained and greater credibility of studies in this area. Despite the fact that RA presented themselves at different stages of training, standardization was able to guarantee the quality of the data collected in clinical research. Special attention is recommended in obtaining agreement among observers in the presence of variables subject to interpretation and personal values (subjectivity).

Acknowledgments

We would like to thank the Coordination for the Improvement of Higher Education Personnel (CAPES - *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*), with a scholarship for the Sandwich Doctorate Program Abroad (PDSE – *Programa de Doutorado Sanduíche no Exterior*), Brasília, FD - Brazil; the Research and Graduate Group (GPPG) of *Hospital de Clínicas de Porto Alegre* (HCPA), Porto Alegre, RS - Brazil; the HCPA Research Incentive Fund (FIPE), Porto Alegre, RS - Brazil; the members of the research group Nursing Care Study Group (NUCAS), Porto Alegre, RS - Brazil.

Collaborations

Corrêa AP, Silva SM, Oleques CM, Sousa GP, Viegas GL, Anziliero F, Oliveira AC, Beghetto MG contributed to the project design, data analysis and interpretation, writing of the article, critical review of relevant intellectual content and final approval of the version to be published.

References

1. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.

2. Souza AC, Alexandre NM, Guirardello EB. Psychometric properties in instruments evaluation of reliability and validity. Epidemiol Serv Saude. 2017;26(3):649–59.

3. Kumar S, Singh MB, Shukla G, Vishnubhatla S, Srivastava MV, Goyal V, et al. Effective clinical classification of chronic epilepsy into focal and generalized: A cross sectional study. Seizure. 2017;53(5): 81–8.
4. Nielsen LG, Folkestad L, Brodersen JB, Brabrand M. Inter-observer agreement in measuring respiratory rate. PLoS One. 2015;10(6):e0129493.
5. Fleury EF, Marcomini K. Breast elastography: diagnostic performance of computer-aided diagnosis software and interobserver agreement. Radiol Bras. 2020;53(1):27-33.
6. Greer MD, Brown AM, Shih JH, Summers RM, Marko J, Law YM, et al. Accuracy and agreement of PIRADSv2 for prostate cancer mpMRI: a multireader study. J Magn Reson Imaging. 2017;45(2):579–85.
7. Stirma GA, Secundino AR, Gonzalez GF, Sola WC, de Souza GA, Dau L. Inter/intra-observer evaluation between radiographs and tomographies for proximal humerus fracture. Acta Ortop Bras. 2020;28(1):36-9.
8. Beghetto MG, Anziliero F, Leães DM, de Mello ED. Feeding tube placement: auscultatory method and x-ray agreement. Rev Gaúcha Enferm. 2015;36(4):98–103.
9. Cortés OL, Alvarado PA, Rojas YA, Salazar LD, Acuña X, Esparza M. Digital Photography: a Tool for Nursing on the Assessment of Pressure Lesions. Invest Educ Enferm. 2018;36(2):e07.
10. Dabbs GR, Connor M, Bytheway JA. Interobserver Reliability of the Total Body Score System for Quantifying Human Decomposition. J Forensic Sci. 2016;61(2):445–51.
11. Girianelli VR, Tomazelli JG, Nogueira MC, Corrêa CS, Souza EO, Gabrielli L, et al. Interobserver reliability in the classification of pairs of records formed by probabilistic linkage of SISMAMA databases. Rev Bras Epidemiol. 2019;22:e190045.
12. Stuchi DP, Rossato J, Bocaccio FJ, Procianoy F. Intra- and interobserver reliability of a modified distraction test based on digital images to assess lower eyelid horizontal tension. Arq Bras Oftalmol. 2020;83(2):127-31.
13. Tammaa A, Fritzer N, Lozano P, Krell A, Salzer H, Salama M, et al. Interobserver agreement and accuracy of non-invasive diagnosis of endometriosis by transvaginal sonography. Ultrasound Obstet Gynecol. 2015;46(6):737–40.
14. Casas M, Vergara M, Brullet E, Junquera F, Bauer EM, Miquel M, et al. Concordancia inter e intraobservador en el diagnóstico de la gastropatía por hipertensión portal. Rev Esp Enferm Dig. 2018;110(3):166–71.
15. Almeida VA, Fernandes CH, dos Santos JB, Schwarz-Fernandes FA, Faloppa F, Albertoni WM. Evaluation of interobserver agreement in Albertoni's classification for mallet finger. Rev Bras Ortop. 2018;53(1):2–9.
16. Lima MA, Pagliuca LM, Nascimento JC, Caetano JÁ. Comparing Interrater reliability between eye examination and eye self-examination 1. Rev Lat Am Enfermagem. 2017;25:e2966.
17. Alves GR, Malaquias SG, Bachion MM. Pressure Ulcer Scale for Healing (PUSH): confiabilidade interobservadores na avaliação de úlceras venosas. Rev Eletr Enf. 2018;20(20):a33.
18. Silva AF, Velo MM, Pereira AC. Importância da reprodutibilidade dos métodos para diagnóstico em odontologia. RFO UFP. 2016;21(1):115–20.
19. Instituto de estudos de saúde suplementar (IESS). Erros acontecem: a força da transparência para o enfrentamento dos eventos adversos assistenciais em pacientes hospitalizados. Belo Horizonte: IESS; 2016 [citado 2021 Fev 16]. Disponível em: Acessado em: http://documents.scribd.com.s3.amazonaws.com/docs/5x5i1j985c5jwcsp.pdf
20. Brasil. Ministério da Saúde. Agência Nacional de Vigilância Sanitária. Resolução - RDC Nº 63, de 6 de julho de 2000. Regulamento Técnico para a Terapia de Nutrição Enteral. Brasília (DF): Ministério da Saúde; 2020 [citado 2021 Fev 16] Disponível em: http://bvsms.saude.gov.br/bvs/saudelegis/anvisa/2000/rdc0063_06_07_2000.html
21. Braga Azambuja F, Beghetto MG, de Assis MC, de Mello ED. Food intake reported versus nursing records: is there agreement in surgical patients? Nutr Hosp. 2015;31(6):2735–9.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.
23. Brasil. Ministério da Saúde. Conselho Nacional de Saúde. Resolução Nº 466, de 12 de dezembro de 2012. Brasília (DF): Ministério da Saúde; 2012 [citado 2021 Fev 16]. Disponível em: https://bvsms.saude.gov.br/bvs/saudelegis/cns/2013/res0466_12_12_2012.html
24. Souza CC, Chianca TC, Júnior WC, Rausch MC, Nascimento GF. Análise de confiabilidade do sistema de triagem de Manchester: acordo interobservador e intraobservador. Rev Lat Am Enfermagem. 2018;26:e3005.
25. Egido JA, Garcia AM, Del Prado-Gonzalez N, Fuentes-Ferrer M, Lopez-Herranz M, Simal-Hernández P, et al. Impact of clinical training on supra-aortic duplex and transcranial doppler examination concordance. J Clin Ultrasound. 2016;44(9):571–9.