

Development of an instrument to evaluate diagnosis accuracy*

CONSTRUÇÃO DE INSTRUMENTO PARA AVALIAR A ACURÁCIA DIAGNÓSTICA

CONSTRUCCIÓN DE INSTRUMENTO PARA EVALUAR LA EXACTITUD DIAGNÓSTICA

Fabiana Gonçalves de Oliveira Azevedo Matos¹, Diná de Almeida Lopes Monteiro da Cruz²

ABSTRACT

Accurate diagnoses are important for choosing adequate interventions. This study reports on a methodological research that developed an instrument to estimate nursing diagnosis accuracy, based on written data from patients' assessment. A definition of accuracy was elaborated and the items that compose the instrument were constructed and submitted to content validation and pilot test. The instrument was named Nursing Diagnosis Accuracy Scale - NDAS and comprised four items: Presence of cues; Relevance of cues; Specificity and Coherence of cues. The answers of 12 expert nurses regarding the application of NDAS to the diagnoses of five written cases permitted the identification of values for each item and estimation of validity and reliability of the NDAS.

KEY WORDS

Nursing diagnosis.
Validation studies.
Reproducibility of results.

RESUMO

Diagnósticos acurados são importantes para a escolha adequada de intervenções. Neste artigo relata-se pesquisa metodológica de desenvolvimento de instrumento para estimar a acurácia de diagnósticos de enfermagem a partir de dados escritos da avaliação de paciente. Elaborou-se a definição de acurácia e a construção dos itens que compõem o instrumento, submetendo-os a validação de conteúdo e teste piloto. O instrumento foi denominado Escala de Acurácia de Diagnóstico de Enfermagem - EADE e foi composto por 4 itens: Presença de pistas; Relevância da pista; Especificidade da pista e Coerência da pista. As respostas de 12 enfermeiros especialistas à aplicação da EADE aos diagnósticos de 5 casos escritos permitiram identificar valores para cada item e estimar a validade e confiabilidade da EADE.

DESCRIPTORIOS

Diagnóstico de enfermagem.
Estudos de validação.
Reprodutibilidade dos testes.

RESUMEN

Diagnósticos precisos son importantes para la elección adecuada de las intervenciones. Este artículo se refiere al desarrollo de herramienta metodológica de investigación para estimar la precisión de los diagnósticos de enfermería a partir de datos escritos de la evaluación de paciente. Se elaboró la definición de la precisión y la construcción de los elementos que componen el instrumento, sometidos a prueba del contenido y el test piloto. El instrumento fue llamado Exactitud Diagnóstica de Enfermería - EADE y estaba compuesto por 4 elementos: Presencia de pistas; Pertinencia de la pista; especificidad de la pista y la coherencia de la pista. Las respuestas de 12 enfermeros especialistas a la aplicación de EADE a los diagnósticos de 5 casos por escrito permitieron identificar los valores para cada tema y estimar la validez y la fiabilidad de EADE.

DESCRIPTORES

Diagnóstico de enfermería.
Estudios de validación.
Reproducibilidad de resultados.

* Extracted from dissertation "Construção de instrumento para a avaliação da acurácia diagnóstica", Nursing School of the University of São Paulo, 2006.
¹ Nurse. Doctoral Student of Graduate Program in Adult Health of Nursing School of the University of São Paulo. Assistant Professor of Oeste do Paraná State University (UNIOESTE). Cascavel, PR, Brazil. fabianamatos@hotmail.com ² Nurse. Full Professor of Medical-Surgical Nursing Department of Nursing School of the University of São Paulo. São Paulo, SP, Brazil. dinamacruz@usp.br

INTRODUCTION

The interest in developing this study emerged because few studies have been developed on diagnosis accuracy, and especially because diagnoses are still little used and valued in clinical practice. Nurses are continually interpreting individuals' responses, however, in general, they do not consider the accuracy of these interpretations⁽¹⁾.

There is, in the literature, a method to estimate the accuracy of diagnoses established by nurses, which includes a scale of degree to which a diagnosis corresponds to data presented by the patient (Lunney's scale)⁽²⁾. It was developed for situations in which the nurse who applies the scale has the opportunity to evaluate the patient for whom the diagnosis was established⁽²⁾. Although the validity of Lunney's scale was evidenced in two studies carried out by the author herself⁽³⁻⁴⁾, it was not evidenced in a Brazilian study in which the scale was applied to written data from patients' evaluations⁽⁵⁾. Nonetheless, this Brazilian study elicited important ideas on the measurement of diagnosis accuracy. Thus, the development of an instrument to evaluate the accuracy of nursing diagnoses based on written data was based on Lunney's scale⁽²⁾.

OBJECTIVE

This article aims to report on the development of the Nursing Diagnosis Accuracy Scale - NDAS, an instrument to evaluate the accuracy of nursing diagnoses based on data written on patients' evaluations, and present the NDAS for further test and refinement.

LITERATURE REVIEW

Nursing diagnosis is the expression of a clinical situation that can be modified by nurses, using standard language⁽⁶⁻⁷⁾. Nursing diagnoses aim to establish a link between patients' answers and the objectives expected by nursing care and interventions needed to achieve them. Diagnoses that originate from data compose a basis for the selection of nursing interventions⁽⁸⁾.

Research has showed that there are variations in nurses' interpretations, and also that patients' responses to health problems are misunderstood⁽¹⁾. An accurate diagnosis is one that reflects patients' real condition.

One of the problems with the concept of accuracy is that, in common sense, accuracy is defined as a dichotomous property. But this is not what happens with nursing diagnoses. The accuracy of a nursing diagnosis is a continuous property, that is, a diagnosis can vary between more and less accuracy, it is not a matter of *all or nothing*. The idea of accuracy as a dichotomous variable was kept until

1992. The analysis of studies from the 1960's to 1980's evidenced that accuracy of nursing diagnoses is very complex and is characterized by a continuum⁽¹⁻²⁾.

Accuracy in interpreting human responses is important because these interpretations base the selection of interventions, which, in turn, contribute to the achievement of desirable results. Little accurate interpretations can lead to omission of care and, consequently, harm the patient^(1-4,9).

The topic under study is an important aspect of knowledge on nursing diagnoses. Including the concept of accuracy in nursing teaching can allow future professionals to establish useful references for their role in diagnosing. The availability of reliable and valid methods to evaluate the accuracy of nursing diagnoses can advance knowledge on the diagnosis process and increase the reliability of clinical studies on nursing diagnoses.

METHOD

This is a methodological development study. The stages of this study were organized in theoretical, empirical (experimental) and analytical (statistical) procedures⁽¹⁰⁾. Analytical procedures depend on the results of theoretical and empirical procedures. Thus, the results of these two stages are presented in the method section.

Theoretical Procedures:

The accuracy of a nursing diagnosis was defined as *the judgment of an evaluator regarding the level of relevance, specificity and consistency of cues existent for the diagnosis*.

This definition guided the development of the instrument to be applied to written data, as well as the remaining procedures. The definitions of cues, number of cues, their relevance, specificity and consistency that guided the construction of the items were:

1. Presence of cues: presence of patients' manifestations that represent indicators, vestiges, signs or defining characteristics of a nursing diagnosis.
2. Number of cues: quantity of cues for a nursing diagnosis.
3. Relevance of cues: cue property of being important as indicator of a nursing diagnosis
4. Specificity of cues: cue property of being proper and distinctive of a cue for a nursing diagnosis
5. Consistency of cues: cue property of being coherent with the available data set.

These assumed definitions generated the items that compose the instrument to be applied to the stated nursing diagnoses, jointly with patients' written data (interview, physical exam, and patients' file data). The developed items

The accuracy of a nursing diagnosis was defined as the judgment of an evaluator regarding the level of relevance, specificity and consistency of cues existent for the diagnosis.

represent judgments of accuracy of a diagnosis, that is, they refer to a set of variables that aim to represent the whole range of accuracy indicators a diagnosis can present.

The items developed for the instrument, as well as the three categories that indicate degrees [many, few, not at all] created for the measurement of each item, were submitted to a face validity test. A group of three acknowledged Brazilian researchers on nursing diagnoses, with experience in advising master and doctoral students on this subject, were invited to judge the validity of the developed items. They were asked to analyze the instrument items regarding their pertinence, clarity and applicability and, in addition, to appoint suggestions of items and changes they found relevant.

Initially, the raters worked individually and their answers were registered on a specific form. Later, they discussed them with the researchers in person. The decisions they reached regarding the items after the content validation procedures were the following:

- 1) To change the item *Number of cues* to *Presence of cues*, assuming *yes/no* answer categories. Justification: the number of cues is of little importance as accuracy indication of a diagnosis because a diagnosis can be stated in the presence of only one cue; the categories *many* or *few* cues demanded multiple combinations in the subsequent items, because there can be cues with different degrees of relevance, specificity and consistency for the same diagnosis.
- 2) To keep the item *Relevance of cues* and alter categories (degrees) to *high/moderate* and *low/null*. Justification: the existence of non-relevant cues superposes the lack of cues, characteristic met by changing the first item (presence of cues).
- 3) To keep the item *specificity of cue* and alter categories (degrees) to *high/moderate* and *low/null*. Justification: the current development of diagnosis classification does not support the judgment of three degrees of specificity and two degrees would preserve correspondence with the categories of the previous item (relevance of cues).
- 4) To keep the item *Consistency of cue* and alter categories (degrees) to *high/moderate* and *low/null*. Justification: to keep the pattern with the answer categories for the relevance and specificity items.

After the content validation procedures, the set of items was submitted to two pilot tests. The items were organized, according to the results of the content validation, in a form with orientations for application. The members of a research group on nursing diagnoses were invited to apply the pilot instrument, in two sessions, to diagnoses and written data from interviews and physical exams of four patients. During the group meeting, the participants were asked to individually apply the instrument to the provided data. After individual work, the group discussed impressions and difficulties and made suggestions to improve the instrument.

As a result of the discussion on the first pilot test, the item *Consistency* was changed to *Coherency* because it was believed the term consistency would not be well understood. Another change was the inclusion of a final item in which the evaluator should answer whether (s)he would state the evaluated diagnosis. Participants reached the consensus that some diagnoses, even in the presence of cues with some level of relevance, specificity and coherence, would not be highly accurate and, therefore, should not be indicated. The main reason that led to this decision was the possible existence of redundancy between diagnoses, which would require that the evaluator choose the one that would better explain the presented data. This argument based the inclusion of the question *would you enunciate this diagnosis?* with the response categories *yes/no*.

This inclusion was also suggested as a resource for analytical procedures. Because of the quantity and quality of suggestions from the first pilot test, the instrument was readjusted and submitted to a second pilot test with the same group, following the same procedure of the first pilot test. The group positively evaluated the incorporated changes and suggestions proposed for this stage in terms of format. The created instrument was then submitted to empirical procedures.

Empirical Procedures

The target population, which the diagnosis accuracy instrument was developed for, comprised nurses with experience in: 1) patients' clinical area; 2) use of classification of diagnoses; and 3) concept of accuracy of nursing diagnosis which the instrument is based on. The following material was prepared for the empirical procedures: NDAS with orientations regarding its use and five written evaluations of patients with their respective nursing diagnoses, called *written data*.

Written data were elaborated by the authors, who concomitantly evaluated five patients hospitalized in medical and surgical clinics of a teaching hospital in São Paulo. This evaluation was performed through interviews, physical exams and consultation of patients' charts. Data were registered on a form, adapted for this study, according to the NANDA-I⁽⁷⁾ diagnosis classification domains. The two researchers independently formulated the diagnoses of evaluated patients and then discussed the most accurate ones, defining by consensus. A list of diagnoses was obtained, corresponding to the records of interviews and physical exams of each patient. This list contained diagnoses defined as the most accurate and also some diagnoses considered, by the authors, as little accurate or not accurate at all. For each case, a *template* of accuracy (high/low) of the listed diagnoses was elaborated. The *gold standard* was used to base the estimated instrument validity. There were 43 listed diagnoses for the five listed cases.

Analytical Procedures

The 12 nurse experts in nursing diagnoses were asked to apply the NDAS to the written data of the five patients evaluated by the authors, so as to test the psychometric proprieties of the elaborated instrument. The experts' choice criteria were guided by their practical and theoretical experience in the area of nursing diagnoses, which corresponds to the target population the NDAS was developed for.

The validity of a new instrument can be evaluated by comparing the answers it provides to answers that result from an acknowledged valid instrument and considered gold standard for comparisons⁽¹¹⁾. Since there is no other instrument for comparison, the gold standard adopted in this study was the templates of the five patients' diagnoses.

To estimate the validity of the NDAS, the hypothesis that the instrument items (presence of cues, relevance, specificity and coherence) predict accurate diagnoses (correct diagnoses according to the adopted gold standard) was tested. Generalized Estimating Equations (GEEs) for binomial distribution⁽¹²⁾ were applied. The dependent variable was defined as the *correct* diagnosis (the expert's answer matches the gold standard) and the instrument items (presence of cues, relevance of cues, and cues specificity and coherence) as independent variables.

In the GEEs, response dependence for the evaluated diagnoses due to repetition of patients (five) and experts⁽¹²⁾ was taken into account. It was assumed that the presence of cues would influence the choice of diagnosis in the same way for any of the evaluated diagnoses. In summary, it was considered there was intra-patient and intra-expert correlation in the model but that the diagnoses were independent. That is, when an expert examined patient one, she evaluated several diagnoses; we considered that there was correlation between these evaluations because it was the same expert and the same patient. Between equal diagnoses in different cases (diagnosis-patient), however, we considered independence. All estimations were calculated using the Statistical Analysis System (SAS) Version 8.02.

The effects of the items relevance, specificity and coherence of cues in the accuracy of diagnosis were evaluated for the patients-diagnoses with presence of cues only (correct answer according to the gold standard). Thus, associations of answers of relevance, specificity and coherence of cues with the correct diagnoses-patients were tested.

The evaluation was carried out by odds ratios (OR), that is, the odds ratio of a correct diagnosis with a certain answer for each item of the NDAS (i.e. high/moderate relevance, low/null specificity and high/moderate coherence) with relation to the odds of the correct answer under other characteristic (i.e. high/moderate relevance, high/moderate specificity and low/null coherence). With this in mind, the effects of each of the four items in the NDAS were evaluated.

The reliability of NDAS was tested in terms of inter-rater agreement, estimated by the Kappa coefficient, considering the diagnosis-patient as the information unit. The coefficient was estimated for the four items of the instrument (cues, relevance, specificity and coherence). A 5% significance level was adopted in all tests.

Complementary analyses were also carried out to attribute final scores to each item. These scores were attributed based on approximate Odds Ratio values obtained in the GEEs for combined effect of relevance, specificity and coherence. This procedure was carried out for the NDAS to provide ordinal results in accuracy degrees. To define accuracy degrees, the possible total scores of the NDAS were interpreted according to its item definitions.

The research followed the guidelines determined by resolution 196/96 - National Health Council for research involving human beings.

RESULTS

On the average, the 12 experts who applied the NDAS were 43.8 (± 9.7) years old and time since graduation was 21.8 (± 9.5) years; 83.3% had doctoral degrees; 91.7% predominantly worked in the adult health area and 75.0% were involved in teaching and research; 91.6% used classifications in their practice and 83.2% evaluated their own ability in formulating diagnoses as above 70% of accuracy.

As described in the method section, GEEs were applied to estimate the NDAS validity. Thus, the hypothesis that the items of the instrument (cues, relevance, specificity and coherence) predict accurate diagnoses (diagnoses that match the gold standard) was tested.

The dependent variable was the correct diagnosis. Independent variables were the instrument items regarding cues.

The results showed great variation in the effect of presence of cues for each expert. It means that there was expert-cue interaction. For one expert, for example, cues increased the odds of correct diagnosis five-fold. For another expert, there was no evidence of cue effects because the odds ratio was lower than one and not significant. For this expert, the presence or lack of cues did not interfere in her correct diagnosis. The general effect of presence of cues was 2.16 (CI 95% [1.39-3.35], $p = 0.001$) on the average (for all experts included in this model).

The results regarding the effects of relevance, specificity and coherence of cues on the correct diagnoses were estimated by the applied GEEs only for the diagnoses-patients for which the experts indicated the presence of cues. From the total of 516 evaluations (12 experts X 43 diagnoses), only 418 evaluations remained for this part of the analysis because the experts stated there were no cues for diagnoses-patients in 98 evaluations.

Table 1 - Results of separate GEEs for the effects of relevance, specificity and coherence of NDAS - São Paulo - 2006

Item		OR	SE(OR)	Confidence Interval 95%		p - value
				Lower Lim	Upper Lim	
Relevance*	High or Moderate/Low or Null	16.20	5.25	8.58	30.58	<.0001
Specificity**	High or Moderate/Low or Null	12.94	3.66	7.44	22.52	<.0001
Coherence***	High or Moderate/Low or Null	19.48	6.04	10.61	35.76	<.0001

* Model with relevance effects ($p < 0.001$) and expert control ($p = 0.006$)

** Model considering expert ($p = 0.004$), specificity ($p < 0.001$) and expert *specificity ($p = 0.007$)

*** Model with coherence effects ($p < 0.001$) and expert control ($p < 0.001$)

There was effect of relevance, specificity and coherence on the correct diagnosis (Table 1). The expert variable was considered, in the three models, for the calculation of effects of items on correct diagnoses because of inter-rater variations for the considered comparisons (high or moderate versus low or null). There was interaction between expert and specificity only for the item specificity, that is, the OR of experts on the effect of their judgments regarding specificity of cue being high/moderate or low/null significantly differed among them.

The correct diagnosis, for those who considered cues of high or moderate relevance, was 16.2 times the odds of correct diagnosis in relation to those who considered cues of low or null relevance, with statistical significance ($p < 0.001$). The model for the effect of cue relevance on the correct diagnosis was controlled by the *expert* variable ($p < 0.001$), which indicates that experts judged the relevance of cues of diagnoses-patients differently. The fact that there was no interaction of experts with relevance indicates that the effect of judgment of relevance on the correct diagnosis did not differ significantly among nurses.

The odds of correct diagnosis of those who identified cues of high or moderate specificity was 12.9 times the odds

of correct diagnosis of those who identified cues of low or null specificity, which was statistically significant ($p < 0.001$). There was interaction between expert and specificity in the model of effect of cue specificity on the correct diagnosis ($p = 0.007$), which indicates significant inter-rater differences in OR on the effect of judgment of cue specificity. That is, there were differences among experts regarding the effect of judgment of specificity on the correct diagnosis.

The odds of correct diagnosis of experts who identified high or moderate coherence was 19.4 times the odds of those who identified low or null coherence, which was also significant ($p < 0.001$). The model of effect of cue coherence on the correct diagnosis was controlled by the *expert* variable ($p < 0.001$), which indicates that the experts judged cue coherence in a statistically different way. The fact that there was no interaction between expert and coherence shows that the effect of judgment of cue coherence on the correct diagnosis was not statistically different among the experts.

When analyzed separately (Table 1), all items were significant for the correct diagnosis. However, results in Table 2 show that the model considering the three combined items presented only one significant factor, coherence.

Table 2 - Results of GEE for the combined effect of relevance, specificity and coherence - São Paulo - 2006

Item	OR	SE(OR)	CI 95%	p - value
Relevance: High/Low	1.01	0.75	[0.23-4.35]	0.994
Specificity: High/Low	3.41	2.57	[0.78-14.92]	0.103
Coherence: High/Low	8.05	5.53	[2.10-30.93]	0.002

GEE: nurse ($p = 0.222$), relevance ($p = 0.994$), specificity ($p = 0.103$) and coherence ($p = 0.002$)

The results of Table 3 show agreement among experts.

Table 3 - Kappa results for the NDAS items evaluated by 12 experts - São Paulo - 2006

Items	Kappa (general)		
	Coefficient	p - value	CI 95%
Presence of cue	0.329	< 0.001	[0.292-0.366]
Relevance of cue	0.257	< 0.001	[0.229-0.284]
Specificity of cue	0.242	< 0.001	[0.215-0.269]
Coherence of cue	0.276	< 0.001	[0.249-0.303]

Kappa general coefficients varied from 0.242 (specificity) to 0.329 (presence of cue). Kappa can vary from 0 to 1, where 0 is total absence of agreement and 1 is total agreement among raters.

Approximate values of OR obtained in the GEEs for combined effect were used (Table 2) to attribute ordinal values to the NDAS. No score was attribute to the item *presence of cues* because, when the evaluator gives a negative answer to this item, (s)he is oriented not to answer the other items. To define degrees of accuracy, the possible total scores obtained with the OR were interpreted according to definitions of items of NDAS and then grouped, aiming to reduce the number of categories of the total score:

1. Degree **0** (there are no cues that indicate the diagnosis under study or the existent cues are of low relevance, low specificity and low coherence) - Accuracy category: **null** (there is no diagnosis or the diagnosis presents null accuracy)

2. Degrees **1 / 3.5 / 4.5** (there are only highly relevant cues for the diagnosis under study; or there are only highly specific cues for the diagnosis; or there are highly relevant and specific cues for the diagnosis) - Accuracy category: **moderate** (little accurate ED or moderately accurate ED).

3. Degrees **8 / 9 / 11.5 / 12.5** (There are only cues highly coherent with the context of the diagnosis under study; or cues are highly relevant to the context of the diagnosis; or there are cues highly specific to the context of diagnosis; or there are cues highly relevant, specific and coherent with the context of diagnosis) - Accuracy category: **high** (accurate ED or very accurate ED).

It is observed that the three generated accuracy categories are theoretically relevant. The *null* accuracy applies to the judgments in which there are no cues for the diagnosis or the existent cues are of low relevance, low specificity or low coherence. The *moderate* accuracy category applies to judgments in which there are only highly relevant and/or highly specific cues. In this category, cues are also considered as not coherent with the set of existent data, despite their relevance, specificity or both. The *high accuracy* category applies to judgments in which the existent cues are highly coherent, by themselves or jointly with any other combination regarding specificity or relevance of cues.

The NDAS developed in this study should be applied to a set of written data (data from patients' evaluations and list of nursing diagnoses) and is composed of: 1 - Orientation for applying the instrument; 2 - Table to list judgments; 3 - Answer score system A and B. The appendix presents an example of NDAS application that contains these elements, a set of written data and a template.

DISCUSSION

Results confirm that even experts reach different diagnostic conclusions based on the same set of data. This situ-

ation is not exclusive to nursing diagnoses but is found in other clinical areas. In all cases, it was observed that experts judge each item differently, as confirmed by the Kappa coefficients (Table 3).

All Odds Ratio estimates obtained by the GEEs in the univariate analysis were statistically significant based on the established alpha (≤ 0.05) (Table 1), which permits rejecting the null hypothesis that the items of the NDAS do not predict diagnosis accuracy, and confirm the instrument's validity. When combining the answers for relevance, specificity and coherence in the GEEs, only coherence continued significance (Table 2). We argued, based on the literature^(2,13), that the item *relevance* would be the most important in predicting accuracy of enunciated diagnoses because it is not sufficient to have cues. Instead, these cues need to be important (relevant) so that one can enunciate this diagnosis. We also expected that the item *specificity* would obtain the second higher score in the scale. Sharing cues in nursing diagnoses is common^(2,5) and a specific cue would be essential for an accurate diagnosis. A lower score was predicted for the item *coherence* because this item does not analyze the cue based on the diagnosis per se, but on a set of data this diagnosis is inserted in. Results contradicted expectations because the item *coherence* presented the highest predictive power for a *correct* diagnosis (Table 2).

Data obtained through multivariate analysis were used to attribute *weight* to the NDAS items, generating ordinal scores that express the continuous nature of variables. Eight scores were obtained. The interpretation of these scores, based on the theoretical and validated definitions of NDAS items, guided the reduction of the eight scores for three categories of accuracy: high accuracy, moderate accuracy and null accuracy.

Regarding NDAS reliability, the general Kappa coefficients varied from 0.242 (specificity) to 0.329 (presence of cues). Results showed that reliability was fair⁽¹⁴⁾. However, the estimated low reliability should be analyzed in view of the NDAS characteristics and the context it is applied in.

The application of NDAS demands that each rater re-thinks part of the diagnostic rationale necessary to establish each diagnosis. In this perspective, it is reasonable to compare the results obtained here with those of studies on accuracy of clinical diagnoses. Studies of this nature are carried out with diagnoses in other contexts, with the difference that these studies are carried out with only one diagnosis. In the present study, each evaluator rated 43 diagnoses.

In a study on inter-rater reliability of the classification system for pressure ulcers for the European Pressure Ulcer Advisory Panel, in which 1,452 nurses from five European countries classified twenty validated photographs of skin lesions, the estimated Kappa was 0.33⁽¹⁵⁾. In a literature review on the accuracy of a diagnosis regarding bone quality, quantity and density before and during dental implants, authors identified Kappa values between 0.33 and 0.67⁽¹⁶⁾.

This kind of results suggest that the results of the present study are compatible with what has been observed in terms of inter-rater agreement in clinical diagnoses.

CONCLUSION

This study permitted the creation of a four-item scale (presence, relevance, specificity and coherence of cues) to evaluate accuracy of nursing diagnoses based on written data. The NDAS was developed for application by raters with experience in the use of nursing diagnosis classifications and knowledge in the clinical area of the patient whose diagnosis will be evaluated. The evaluation of nursing diagnoses is based on cues (defining characteristics) they present.

With written data from patients' evaluation and list of enunciated diagnoses, the rater judges whether there are cues for each formulated diagnosis. In the presence of cues, the rater judges them in terms of relevance, specificity based on the evaluated diagnosis and coherence, based on the set of available data. Answers to each NDAS item, except for *presence of cues*, correspond to one score (High relevance = 1/ High specificity = 3.5/ High coherence = 8), and the sum of all answers results in a final score that indicates the degree of accuracy of the obtained diagnosis (0 / 1 / 3.5 / 4.5 / 8 / 9 / 11.5 / 12.5). Finally, based on the

obtained degree of accuracy, the category of accuracy the diagnosis fits in can be identified (high, moderate or null).

The obtained degrees of accuracy and the categories the diagnoses were classified in when using the NDAS should be reported. The instrument presented reasonable validity and reliability estimates. It was tested for use with written data and any other form of application requires previous validity and reliability estimates.

This study presents great advancements in comparison to its precursor⁽⁵⁾, in which the translated version of the Lunney Scoring Method for Rating Accuracy of Nursing Diagnoses - LSM⁽²⁾ did not present acceptable reliability estimates when using written data.

Regarding the study limitations, the fragility of the established *gold standard* and the low level of inter-rater agreement that resulted from the NDAS are highlighted. What motivated the definition of *gold standard* used in the study was the lack of a better alternative. The low level of agreement obtained seems to be consistent with what is found in clinical diagnoses, however, the search for strategies to improve agreement in the use of NDAS is the main challenge for its refinement. It is recommended that evaluators be previously trained with case studies and have adequate understanding of the item definitions for using NDAS.

REFERENCES

1. Lunney M. Pensamento crítico e diagnósticos de enfermagem: estudos de caso e análises. Trad. de Rômulo Marques. Porto Alegre: Artmed; 2004. A precisão no diagnóstico das respostas humanas: a necessidade do pensamento crítico; p. 39-51.
2. Lunney M. Accuracy of nursing diagnosis: concept development. *Nurs Diagn*. 1990;1(1):12-7.
3. Lunney M. Divergent productive thinking factors and accuracy of nursing diagnoses. *Res Nurs Health*. 1992;15(4):303-11.
4. Lunney M, Karlik B, Kiss M, Murphy P. Accuracy of nurses' diagnoses of psychosocial responses. *Nurs Diagn*. 1997;8(4):157-66.
5. Cruz DALM, Fontes CMB, Braga CG, Volpato MP, Azevedo SL. Adaptação para a língua portuguesa e validação do Lunney Scoring Method for Rating Accuracy of Nursing Diagnoses. *Rev Esc Enferm USP*. 2007;41(1):127-34.
6. Souza MSA. Diagnóstico de enfermagem em clientes vivenciando estressores pré-cirúrgicos [dissertação]. Salvador: Escola de Enfermagem de Salvador, Universidade Federal da Bahia; 1997.
7. North American Nursing Diagnosis Association (NANDA). Diagnósticos de Enfermagem da NANDA: definições e classificações – 2005/2006. Porto Alegre: Artes Médicas; 2006.
8. Cruz DALM, Perez RCFG. Accuracy of nursing diagnosis: interrater agreement. In: *Proceedings of the Conference NNN; 2002; Philadelphia. Philadelphia: NNN; 2002. p. 47.*
9. Lunney M, Paradiso C. Accuracy of interpreting human responses. *Nurs Manage*. 1995; 26(10):48H-48K.
10. Pasquali L. Princípios de elaboração de escalas psicológicas. *Rev Psiq Clin* [periódico na Internet]. 1998 [citado 2005 jan. 5];25(5): [cerca de 8 p.]. Disponível em: <http://www.hcnet.usp.br/ipq/revista/r255/conc255ahtm>
11. Cummings SR, Hulley SB, Stewart AL. Elaboração de questionário e instrumento de coleta de dados. In: Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman T.B. *Delineando a pesquisa clínica: uma abordagem epidemiológica*. Trad. de Michael Schmidt Duncan e Ana Rita Peres. 2ª ed. Porto Alegre: Artmed; 2003. p. 265-81.
12. Hair JF, Anderson RE, Tatham RL, Black WC. *Multivariate data analysis*. 5th ed. New Jersey: Prentice Hall; 1998.
13. Cianfrani KL. The influence of amounts and relevance of data on identifying health problems. In: *Proceedings of the Fifth National Conference Classification of Nursing Diagnoses; 1984; Toronto. Toronto: Mosby; 1984. p.625.*
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33(1): 159-74.

15. Beeckman D, Schoonhoven L, Fletcher J, Furtado K, Gunningberg L, Heyman H, et al. EPUAP classification system for pressure ulcers: European reliability study. *J Adv Nurs* 2007;60(6):682-91.
16. Ribeiro-Rotta RF, Lindh C, Rohlin M. Efficacy of clinical methods to assess jawbone tissue prior to and during endosseous dental implant placement: a systematic literature review. *Int J Oral Maxillofac Implants*. 2007; 22(2):289-300.

Project funded by FAPESP (Processo n. 06/01731-5)

APPENDIX

Escala de Acurácia de Diagnóstico de Enfermagem – Dados Escritos (EADE-DE)

A EADE-DE foi desenvolvida para estimar o grau com que uma afirmação diagnóstica tem sustentação num conjunto de informações clínicas escritas do paciente. Para o uso adequado da escala o avaliador deve estar suficientemente esclarecido sobre conceitos e termos envolvidos na escala. A aplicação da EADE requer que você analise cada diagnóstico formulado para um paciente, com base nos dados escritos de avaliação clínica..

1. Orientações:

A EADE tem 5 itens com respostas dicotômicas. Os itens 1 e 5 indicam se o diagnóstico deve ser pontuado. Os escores dos itens 2 a 4 permitem uma interpretação da acurácia.

- 1 Leia cuidadosamente os dados escritos da avaliação do paciente (entrevista, exame físico e de outras fontes);
- 2 Responda cada item da EADE para cada diagnóstico estabelecido para o paciente;
- 3 Leia cada item e siga as orientações específicas;
- 4 Sempre que necessário, releia os dados de avaliação;
- 5 Consulte a classificação da NANDA-I para comparar os dados da avaliação com os diagnósticos;
- 6 Use o Quadro de Respostas da EADE para documentar seu julgamento.

Item 1 - Há pista(s) para o diagnóstico? Sim Não

Orientação: Considere a definição de pistas como manifestações dos pacientes que representam indícios, vestígios, sinais, indicações ou características de um diagnóstico de enfermagem. Se houver pelo menos uma pista para o diagnóstico, independente de sua relevância, especificidade e coerência, marque a resposta sim. Se a resposta for NÃO, os outros itens não se aplicam. Interrompa aqui a aplicação do EADE para esse diagnóstico.

Item 2 - A relevância da(s) pista(s) existente(s) é: Alta/Moderada Baixa/Nula

Orientação: Considere a definição de relevância da pista como a propriedade de uma pista de ser importante como indicador de um diagnóstico de enfermagem, e indique o grau de relevância da(s) pista(s) existente(s). Se você julgar que há pista(s) nos dois graus de relevância, indique apenas o mais elevado (Alta/Moderada).

Item 3 - A especificidade da(s) pista(s) existente(s) é: Alta/Moderada Baixa/Nula

Orientação: Considere a definição de especificidade da pista como a propriedade de uma pista de ser própria e distintiva de um diagnóstico de enfermagem, e indique o grau de especificidade da(s) pista(s) existente(s). Se você julgar que há pista(s) nos dois graus de especificidade, indique apenas o mais elevado (Alta/Moderada).

Item 4 - A coerência da(s) pista(s) existente(s) é: Alta/Moderada Baixa/Nula

Orientação: Considere a definição de coerência da pista como a propriedade de uma pista de ser consistente com o conjunto das informações disponíveis, e indique o grau de coerência da(s) pista(s) existente(s). Se você julgar que há pista(s) nos dois graus de coerência, indique apenas o mais elevado (Alta/ Moderada).

2. Quadro de Respostas da EADE

Diagnósticos de enfermagem enunciados*	Item 1		Item 2		Item 3		Item 4		Total (grau de acurácia)	Categoria de acurácia
	Há pistas? **		Relevância (pista x diagnóstico)		Especificidade (pista x diagnóstico)		Coerência (Pista x conjunto dos dados)			
	Sim	Não	A/M (1)	B/N (0)	A/M (3,5)	B/N (0)	A/M (8)	B/N(0)		

*Inserir número de linhas igual ao número de diagnósticos a serem avaliados pela EADE. **Quando a resposta for não, não preencher as demais células para o diagnóstico específico (n/a). A: Alta; M: Moderada; B: Baixa; N: Nula.

3. Sistema de pontuação das respostas

Escores para os itens da EADE				
Itens		Definições	Categorias	Escores de Acurácia
1	Presença de pista	Presença de manifestações dos pacientes que representam indícios vestígios, sinais, ou características de um diagnóstico de enfermagem	Sim Não	Nenhum
2	Relevância da pista	Propriedade de uma pista de ser importante como indicador de um diagnóstico de enfermagem	Alta/Moderada Baixa/Nula	1 0
3	Especificidade da pista	Propriedade de uma pista de ser própria e distintiva de outro diagnóstico de enfermagem	Alta/Moderada Baixa/Nula	3,5 0
4	Coerência da pista	Propriedade de uma pista de ser consistente com o conjunto das informações disponíveis	Alta/Moderada Baixa/Nula	8 0

Aplique os escores às respostas de cada item e some todos os escores de cada diagnóstico. O escore total é interpretado da seguinte forma:

Categorias de acurácia da EADE		
Escores de acurácia	Interpretação	Categoria de acurácia
0	Não há pistas que indiquem o diagnóstico em questão OU As pistas existentes têm baixa relevância, baixa especificidade e baixa coerência	NULA
1 3,5 4,5	A(s) pista(s) presente(s) nos dados de avaliação tem (têm) coerência baixa/nula com os dados da avaliação, mas há pista(s) altamente relevante(s) e/ou altamente específica(s) para o diagnóstico em questão	MODERADA
8,0 9,0 11,5 12,5	A(s) pista(s) presente(s) nos dados de avaliação é (são) altamente coerente(s) com os dados da avaliação E/OU altamente relevante, E/OU altamente específicas para o diagnóstico em questão	ALTA