

Construção de Instrumento para avaliar a acurácia diagnóstica*

DEVELOPMENT OF AN INSTRUMENT TO EVALUATE DIAGNOSIS ACCURACY

CONSTRUCCIÓN DE INSTRUMENTO PARA EVALUAR LA EXACTITUD DIAGNÓSTICA

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

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.

DESCRIPTORES

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

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.

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, sometiendo 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.

* Extraído da dissertação "Construção de instrumento para a avaliação da acurácia diagnóstica", Escola de Enfermagem, Universidade de São Paulo, 2006.

¹ Enfermeira. Doutoranda do Programa de Enfermagem na Saúde do Adulto, Escola de Enfermagem, Universidade de São Paulo (EEUSP). Professora Assistente da Universidade Estadual do Oeste do Paraná (UNIOESTE). Cascavel, PR, Brasil. fabianamatos@hotmail.com ² Enfermeira. Professora Titular do Departamento de Enfermagem Médico-Cirúrgica da Escola de Enfermagem, Universidade de São Paulo (EEUSP). São Paulo, SP, Brasil. mtmllf@usp.br

INTRODUÇÃO

O interesse em desenvolver esta pesquisa surgiu pela escassez de estudos sobre a acurácia diagnóstica e, principalmente, por ela ainda ser pouco utilizada e valorizada na prática clínica. Os enfermeiros estão continuamente interpretando as respostas dos indivíduos, mas em geral não consideram a acurácia de tais interpretações⁽¹⁾.

Há na literatura um método para estimar a acurácia dos diagnósticos estabelecidos pelas enfermeiras, que inclui uma escala do grau com que um diagnóstico corresponde aos dados apresentados pelo paciente (Escala de Lunney)⁽²⁾. Foi desenvolvido para situações em que a enfermeira que aplica a escala tem a oportunidade de avaliar o paciente para o qual o diagnóstico foi estabelecido⁽²⁾. Apesar de a validade da Escala de Lunney ter sido demonstrada em dois estudos realizados pela própria autora⁽³⁻⁴⁾, o mesmo não ocorreu em estudo brasileiro em que a escala foi aplicada a dados escritos da avaliação do paciente⁽⁵⁾. Esse estudo brasileiro, no entanto, forneceu idéias importantes sobre a mensuração da acurácia diagnóstica. Diante disso, partimos da Escala de Lunney⁽²⁾ para desenvolver um instrumento para avaliar a acurácia dos diagnósticos de enfermagem com base em dados escritos.

OBJETIVO

Este artigo tem a finalidade de relatar o desenvolvimento da Escala de Acurácia de Diagnóstico de Enfermagem (EADE), um instrumento para avaliar a acurácia de diagnósticos de enfermagem a partir de dados escritos da avaliação do paciente, e apresentar a EADE para testes e refinamento.

REVISÃO DA LITERATURA

O diagnóstico de enfermagem é a expressão de uma situação clínica que pode ser modificada pelo enfermeiro, utilizando uma linguagem padronizada⁽⁶⁻⁷⁾. A finalidade dos diagnósticos de enfermagem é estabelecer um elo entre as respostas humanas, os objetivos desejados para a assistência de enfermagem e as intervenções necessárias para alcançá-los. Os diagnósticos provenientes do levantamento dos dados consistem em um ponto de apoio para a seleção das intervenções de enfermagem⁽⁸⁾.

Estudos têm mostrado que existem variações nas interpretações dos enfermeiros, e que as respostas dos pacientes aos problemas de saúde são facilmente mal interpretadas⁽¹⁾. Um diagnóstico é acurado quando reflete o real estado do paciente.

Um dos problemas com o conceito de acurácia é que, no senso comum, a acurácia é veiculada como uma pro-

priedade dicotômica. Mas isso não é o que ocorre com os diagnósticos de enfermagem. A acurácia de um diagnóstico de enfermagem é uma propriedade contínua, isto é, um diagnóstico pode ser mais ou menos acurado, não se tratando de *tudo ou nada*. A idéia de acurácia como uma variável dicotômica foi mantida até 1992. A análise de estudos das décadas de 1960 a 1980 mostrou que a acurácia dos diagnósticos de enfermagem é bastante complexa, caracterizando-se por um *continuum*⁽¹⁻²⁾.

A acurácia das interpretações das respostas humanas é importante porque essas interpretações fundamentam a seleção das intervenções, que, por sua vez, contribuem para a obtenção dos resultados desejáveis. As interpretações de baixa acurácia podem levar à omissão de cuidados acarretando prejuízos ao paciente^(1-4, 9).

O tema em estudo é um aspecto importante do conhecimento sobre diagnóstico de enfermagem. Incluir no ensino de enfermagem o conceito de acurácia permitirá aos futuros profissionais estabelecerem referências úteis para

o papel de diagnosticadores. Dispor de métodos confiáveis e válidos para avaliar a acurácia dos diagnósticos de enfermagem permitirá avançar o conhecimento sobre o processo diagnóstico e aumentar a confiabilidade dos estudos clínicos sobre diagnósticos de enfermagem.

MÉTODO

Tratou-se de estudo de desenvolvimento metodológico. As fases deste estudo foram organizadas em procedimentos teóricos, empíricos (experimentais) e analíticos (estatísticos)⁽¹⁰⁾. Os procedimentos analíticos

dependem dos resultados dos procedimentos teóricos e empíricos. Por essa razão, os resultados dessas duas etapas estão apresentados no método.

Procedimentos teóricos:

A acurácia de um diagnóstico de enfermagem foi definida como *o julgamento de um avaliador quanto ao grau de relevância, especificidade e consistência das pistas existentes para o diagnóstico*. Essa definição norteou o desenvolvimento do instrumento para aplicação em dados escritos, guiando os procedimentos subsequentes. As definições de pista, número de pista, relevância, especificidade e consistência das pistas que direcionaram a construção dos itens foram:

1. Pistas: manifestações dos pacientes que representam indícios, vestígios, sinais, indicações ou características definidoras de um diagnóstico de enfermagem.
2. Número de pistas: quantidade de pistas para um diagnóstico de enfermagem.

3. Relevância da pista: propriedade de uma pista de ser importante como indicador de um diagnóstico de enfermagem.

4. Especificidade da pista: propriedade de uma pista de ser própria e distintiva de um diagnóstico de enfermagem.

5. Consistência da pista: propriedade de uma pista de ser coerente com o conjunto das informações disponíveis.

As definições assumidas geraram os itens para compor o instrumento a ser aplicado aos diagnósticos de enfermagem enunciados, junto com os dados escritos do paciente (entrevista, exame físico, dados de prontuário). Os itens desenvolvidos tratavam de expressões do julgamento quanto a acurácia de um diagnóstico, isto é, referiam-se a um conjunto de variáveis que pretendia representar todo o espectro de indicadores de acurácia que um diagnóstico pode ter.

Os itens desenvolvidos para o instrumento, assim como as três categorias indicativas de graus [muito(as), pouco(as), nada] criadas para a mensuração de cada item, foram submetidos a teste de validade aparente. Um grupo de três pesquisadores sobre diagnóstico de enfermagem, nacionalmente reconhecidos, com experiência na orientação de pesquisas de mestrado e doutorado sobre o assunto foi composto para julgar a validade dos itens desenvolvidos. Foi solicitado aos avaliadores que analisassem os itens do instrumento quanto à pertinência, clareza e aplicabilidade. Foi também solicitado aos juízes que apontassem sugestões de itens e de modificações que considerassem pertinentes.

Os trabalhos dos avaliadores foram primeiramente individuais, solicitando-se respostas registradas em formulário próprio, que, depois foram discutidas presencialmente entre avaliadores e autoras. As decisões sobre os itens decorrentes dos procedimentos de validação de conteúdo foram os seguintes:

1) Modificação do item *Número de pistas para Presença de pistas*, assumindo as categorias de resposta *sim/não*. Justificativa: o número de pistas é pouco importante como indicativo da acurácia de um diagnóstico, visto que um diagnóstico pode ser enunciado apresentando apenas uma pista; as categorias de respostas *muitas* ou *poucas* pistas exigiriam múltiplas combinações nos itens subsequentes, uma vez que pode haver pistas com diferentes graus de relevância, especificidade e consistência para um mesmo diagnóstico.

2) Manutenção do item *Relevância das pistas* e alteração das categorias (graus) para *alta/moderada* e *baixa/nula*. Justificativa: a existência de pistas nada relevantes superpõe-se à ausência de pistas, característica atendida com a modificação do primeiro item (presença de pistas).

3) Manutenção do item *Especificidade da pista* e alteração das categorias (graus) para *alta/moderada* e *baixa/nula*. Justificativa: o desenvolvimento atual da classificação

dos diagnósticos ainda não dá suporte para o julgamento dos três graus de especificidade e os dois graus manteriam correspondência com as categorias do item anterior (relevância das pistas).

4) Manutenção do item *Consistência da pista* e alteração das categorias (graus) para *alta/moderada* e *baixa/nula*. Justificativa: compõe padrão com as categorias de respostas dos itens de relevância e de especificidade.

Ao final dos procedimentos de validação de conteúdo, o conjunto de itens foi submetido a dois testes pilotos. Os itens, conforme os resultados da validação de conteúdo, foram organizados num formulário contendo orientações sobre como aplicá-los. Os participantes de um grupo de pesquisa sobre diagnósticos de enfermagem foram convidados a aplicar o instrumento piloto a diagnósticos e dados escritos de entrevista e exame físico de quatro pacientes, em duas sessões. Foi solicitado às participantes que aplicassem individualmente o instrumento aos dados fornecidos durante uma reunião do grupo. Após o trabalho individual, o grupo discutiu as impressões, as dificuldades e fez sugestões para melhorar o instrumento.

As discussões do primeiro teste piloto conduziram à modificação do item *Consistência* para *Coerência* por acreditar que o termo consistência poderia não ser bem entendido. Outra alteração foi a inclusão de um item final em que o avaliador deveria responder se enunciaria ou não o diagnóstico avaliado. Os participantes chegaram ao consenso de que, para alguns diagnósticos, mesmo havendo pistas com algum grau de relevância, especificidade e coerência, eles não seriam altamente acurados e, por isso, não deveriam ser indicados. O principal motivo que levaria a essa decisão seria a existência de alguma redundância entre diagnósticos, o que exigiria do diagnosticador a escolha daquele que melhor explanasse os dados presentes. Esse argumento fundamentou a inclusão da pergunta *você enunciaria esse diagnóstico?*, com as categorias de respostas *sim/não*.

Essa inclusão também foi sugerida como recurso para os procedimentos analíticos. Frente ao montante e qualidade das sugestões do primeiro teste piloto, o instrumento foi reajustado e submetido ao segundo teste piloto no mesmo grupo, seguindo os mesmos procedimentos do primeiro teste piloto. O grupo avaliou positivamente as mudanças incorporadas e as sugestões dessa etapa foram de ordem de formatação do material. A EADE assim criada foi submetida aos procedimentos empíricos.

Procedimentos empíricos

A população alvo para a qual o instrumento de avaliação da acurácia diagnóstica foi desenvolvido é de enfermeiras com experiência: 1) na área clínica do paciente; 2) no uso de classificação de diagnósticos; e 3) no conceito de acurácia de diagnóstico de enfermagem em que o instrumento se fundamenta. Para os procedimentos empíricos foram preparados os seguintes materiais: a EADE com suas orientações de uso e cinco avaliações de

pacientes com seus respectivos diagnósticos de enfermagem, designados a partir daqui de *dados escritos*.

Os dados escritos foram elaborados pelos autores que, concomitantemente, avaliaram cinco pacientes internados em clínica médica e cirúrgica de um hospital de ensino da cidade de São Paulo. A avaliação dos pacientes foi realizada por entrevista, exame físico e consulta ao prontuário. Os dados foram registrados em formulário adaptado para esta pesquisa, segundo os domínios da classificação de diagnósticos da NANDA-I⁽⁷⁾. As duas autoras formularam, independentemente, os diagnósticos dos pacientes avaliados e, a seguir, os discutiram, definindo, por consenso, os mais acurados. Obteve-se, assim, uma lista de diagnósticos correspondente aos registros de entrevista e exame físico de cada paciente. Essa lista continha os diagnósticos definidos como os mais acurados e também alguns dos diagnósticos aventados e definidos pelas autoras como pouco acurados ou nada acurados. Para cada caso elaborou-se um *gabarito* da acurácia (alta/baixa) dos diagnósticos listados. O *padrão-ouro* serviu de base para as estimativas de validade do instrumento. No total dos cinco casos havia 43 diagnósticos listados.

Procedimentos analíticos

Para testar as propriedades psicométricas do instrumento elaborado, foi solicitado a 12 enfermeiras especialistas em diagnósticos de enfermagem para aplicar a EADE aos dados escritos dos cinco pacientes avaliados pelas autoras. Os critérios de escolha das especialistas foram direcionados pela experiência prática e teórica na área de diagnósticos de enfermagem, que corresponde ao público-alvo para o qual a EADE foi desenvolvida.

A pesquisa seguiu os dispositivos da resolução 196/96 do Conselho Nacional de Saúde para pesquisas que envolvem seres humanos (Proc. nº 539/2006-CEP-EEUSP).

A validade de um novo instrumento pode ser avaliada comparando-se as respostas que ele fornece com as respostas que resultaram de um instrumento reconhecidamente válido, considerado padrão-ouro para comparações⁽¹¹⁾. Como não há outro instrumento ao qual comparar a EADE, o padrão-ouro assumido neste estudo foram os gabaritos dos diagnósticos dos cinco pacientes.

Para estimar a validade da EADE, testou-se a hipótese de que os itens do instrumento (presença de pistas, relevância, especificidade e coerência) predizem diagnósticos acurados (diagnósticos corretos segundo o padrão-ouro definido). Aplicaram-se Equações de Estimação Generalizadas (EEG) para distribuição binomial⁽¹²⁾. Definiu-se como variável dependente (variável resposta) o *acerto* do diagnóstico (coincidência da resposta do especialista com o padrão-ouro) e como variáveis independentes os itens do instrumento (presença de pistas, relevância da pista, especificidade da pista e coerência da pista).

Nas EEG foi considerado que há uma dependência de resposta, pois há repetição de pacientes (05) e de especi-

alistas (12) para os diagnósticos avaliados (admitiu-se que a presença de pistas influenciaria da mesma maneira no acerto do diagnóstico para qualquer um dos diagnósticos avaliados). Em síntese, considerou-se que havia correlação intra-paciente e intra-especialista no modelo, mas que os diagnósticos eram independentes. Ou seja, quando a especialista 1 examinou o paciente 1, ela avaliou vários diagnósticos; consideramos então que havia correlação entre essas avaliações por se tratar da mesma especialista e do mesmo paciente, mas entre diagnósticos iguais em casos diferentes (diagnóstico-paciente) consideramos independência. Todos os cálculos foram realizados no *Statistical Analysis System* (SAS) Versão 8.02.

Apenas para os diagnósticos-pacientes que foram julgados com pista presente avaliaram-se os efeitos dos itens relevância, especificidade e coerência das pistas na acurácia do diagnóstico (acerto segundo o padrão-ouro). Para essa finalidade, testaram-se as associações das respostas de relevância, especificidade e coerência das pistas com o acerto dos diagnósticos-pacientes.

A avaliação foi feita por razão de chances (*OR*), isto é, a razão da chance de acerto do diagnóstico com determinada resposta para cada item da EADE (ex: relevância alta/moderada, especificidade baixa/nula e coerência alta/moderada) com relação à chance de acerto sob outra característica (ex. relevância alta/moderada, especificidade alta/moderada e coerência baixa/nula). Com esse raciocínio foram avaliados os efeitos de cada um dos quatro itens da EADE.

A confiabilidade da EADE foi testada em termos de concordância entre avaliadores, estimada pelo coeficiente *Kappa*, considerando como unidade de informação o diagnóstico-paciente. O coeficiente foi estimado para os quatro itens do instrumento (pistas, relevância, especificidade e coerência). Para todos os testes foi assumido o nível de significância de 5%.

Foram feitas também análises complementares para atribuir escores finais para cada item. Tais escores foram atribuídos com base nos valores aproximados das *Odds Ratio* obtidas nas EEG para efeito combinado de relevância, especificidade e coerência. Esse procedimento foi realizado para que a EADE forneça resultados ordinais, em graus de acurácia. Para definir graus de acurácia os escores totais possíveis da EADE foram interpretados de acordo com as definições dos seus itens.

RESULTADOS

As 12 especialistas que aplicaram a EADE tinham em média 43,8(±9,7) anos de idade e 21,8(±9,5) anos de formação; 83,3% tinham doutorado; 91,7% atuavam predominantemente na área de saúde do adulto e 75,0 % estavam envolvidas com o ensino e/ou pesquisa; 91,6% fazia uso de classificações na prática clínica e 83,2% avaliaram a própria habilidade em formular diagnósticos como acima de 70% acurada.

Conforme descrito no método, aplicaram-se EEG para estimar a validade da EADE. Para essa finalidade testou-se a hipótese de que os itens do instrumento (pistas, relevância, especificidade e coerência) predizem diagnósticos acurados (diagnósticos corretos segundo o padrão-ouro definido).

A variável dependente foi o acerto do diagnóstico. As variáveis independentes foram os itens da EADE quanto às pistas.

Os resultados mostraram grande variação quanto ao efeito da presença de pista para cada especialista, o que quer dizer que houve interação especialista-pista. Uma especialista, por exemplo, teve a chance de acerto com a presença de pista 5 vezes a chance de acerto sem presença de pistas. Outra especialista não apresentou efeito da

presença de pistas, pois a razão de chances foi inferior a 1 e não foi significativa. Para essa especialista, a presença ou não de pista não interferiu em nada no seu acerto de diagnóstico. O efeito geral da presença de pista foi de 2,16 (IC 95% [1,39-3,35], $p=0,001$) em média (para todas as especialistas incluídas neste modelo).

Os resultados sobre o efeito dos itens relevância, especificidade e coerência da pista no acerto dos diagnósticos foram estimados pelas EEG aplicadas apenas aos diagnósticos-pacientes para os quais as especialistas indicaram haver pistas. Do total de 516 julgamentos (12 especialistas X 43 diagnósticos) restaram 418 julgamentos para essa parte das análises, pois em 98 julgamentos as especialistas responderam não haver pistas para o diagnóstico-paciente.

Tabela 1 - Resultados das EEG separadamente para os efeitos dos itens de relevância, especificidade e coerência da EADE - São Paulo - 2006

Item		Intervalo de confiança 95%				p - valor
		OR	EP(OR)	Lim inf	Lim sup	
Relevância*	Alta ou Moderada / Baixa ou Nula	16,20	5,25	8,58	30,58	<,0001
Especificidade**	Alta ou Moderada / Baixa ou Nula	12,94	3,66	7,44	22,52	<,0001
Coerência***	Alta ou Moderada / Baixa ou Nula	19,48	6,04	10,61	35,76	<,0001

* Modelo com efeitos de relevância ($p<0,001$) e controle de especialista ($p=0,006$)

** Modelo considerando especialista ($p=0,004$), especificidade ($p<0,001$) e especialista*especificidade($p=0,007$)

*** Modelo com efeitos de coerência ($p<0,001$) e controle de especialista ($p<0,001$)

Houve efeito da relevância, da especificidade e da coerência no acerto do diagnóstico (Tabela 1). A variável especialista foi considerada nos três modelos para os cálculos dos efeitos dos itens no acerto do diagnóstico, pois as especialistas variaram entre si nas comparações consideradas (alta ou moderada *versus* baixa ou nula). Somente no item especificidade houve interação entre especialista e especificidade, isto é, as OR das especialistas quanto ao efeito de seus julgamentos quanto à especificidade da pista ser alta/moderada ou baixa/nula diferiram significativamente entre elas.

O acerto do diagnóstico para quem julgou haver pistas de relevância alta ou moderada foi 16,2 vezes a chance de acerto com relação a quem julgou haver pistas de relevância baixa ou nula e foi estatisticamente significativa ($p<0,001$). O modelo do efeito da relevância da pista no acerto do diagnóstico foi controlado pela variável 'especialista' ($p<0,001$), o que indica que as especialistas julgaram de forma diferente a relevância das pistas dos diagnósticos-pacientes. O fato de não ter havido interação das especialistas com a relevância indica que o efeito do julgamento da relevância no acerto do diagnóstico não diferiu significativamente entre as enfermeiras.

A chance de acerto de quem identificou pistas de especificidade alta ou moderada foi 12,9 vezes a chance de acerto com relação quem identificou pistas de especificidade

baixa ou nula e foi estatisticamente significativa ($p<0,001$). O modelo do efeito da especificidade da pista no acerto do diagnóstico teve interação entre especialista e especificidade ($p=0,007$), o que indica que as OR das especialistas quanto ao efeito do julgamento da especificidade das pistas diferiu significativamente entre elas. Isto é, houve diferenças entre as especialistas quanto ao efeito do julgamento da especificidade no acerto dos diagnósticos.

A chance de acerto das especialistas que identificaram pistas de coerência alta ou moderada foi de 19,4 vezes a chance de acerto com relação às que identificaram pistas de coerência baixa ou nula e foi também significativa ($p<0,001$). O modelo do efeito da coerência da pista no acerto do diagnóstico foi controlado pela variável 'especialista' ($p<0,001$), o que indica que as especialistas julgaram a coerência das pistas de forma estatisticamente diferente. O fato de não ter havido interação entre especialista e coerência mostra que o efeito do julgamento da coerência das pistas no acerto do diagnóstico não foi estatisticamente diferente entre as especialistas.

Quando analisados separadamente (Tabela 1), todos os itens foram significantes para o acerto do diagnóstico. No entanto, os resultados na Tabela 2 mostram que o modelo considerando os três itens combinados apresentou apenas um fator significativo, a coerência.

Tabela 2 - Resultados da EEG para efeito combinado de relevância, especificidade e coerência - São Paulo - 2006

Item	OR	EP(OR)	IC 95%	p - valor
Relevância: Alta ou Baixa	1,01	0,75	[0,23-4,35]	0,994
Especificidade: Alta ou Baixa	3,41	2,57	[0,78-14,92]	0,103
Coerência: Alta ou Baixa	8,05	5,53	[2,10-30,93]	0,002

GEE: enfermeira (p=0,222), relevância (p=0,994), especificidade (p=0,103) e coerência (p=0,002)

Os resultados da Tabela 3 mostram a concordância entre as especialistas.

Tabela 3 - Resultados do Kappa para os itens da EADE avaliados por 12 especialistas - São Paulo -2006

Itens	Kappa (geral)		
	Coefficiente	p - valor	IC 95%
Presença de pista	0,329	< 0,001	[0,292-0,366]
Relevância da pista	0,257	< 0,001	[0,229-0,284]
Especificidade da pista	0,242	< 0,001	[0,215-0,269]
Coerência da pista	0,276	< 0,001	[0,249-0,303]

Os coeficientes gerais do Kappa variaram de 0,242 (especificidade) a 0,329 (presença de pista). O Kappa pode variar de 0 a 1, onde 0 é ausência total de concordância e 1 é 100% de concordância entre avaliadores.

Para atribuir valores ordinais para a EADE foram usados valores aproximados das OR obtidas nas EEG para efeito combinado (Tabela 2). Ao item *presença de pista* não foi atribuído escore, pois quando a resposta do avaliador é negativa ele é orientado a não responder os outros itens. Para definir os graus de acurácia, os escores totais possíveis, obtidos com as OR, foram interpretados de acordo com as definições dos itens da EADE e depois agrupados com a finalidade de reduzir o número de categorias de escore total:

1) Grau 0 (não há pistas que indiquem o diagnóstico em questão ou as pistas existentes têm baixa relevância, baixa especificidade e baixa coerência) - Categoria de Acurácia: **nula** (não há diagnóstico ou o diagnóstico tem acurácia nula).

2) Graus 1 / 3,5 / 4,5 (há somente pista(s) altamente relevante(s) para o diagnóstico em questão; ou há somente pista(s) altamente específica(s) para o diagnóstico em questão; ou há pista(s) altamente relevante(s) e específica(s) para o diagnóstico em questão) - Categoria de Acurácia: **moderada** (DE pouco acurado ou DE moderadamente acurado).

3) Graus 8 / 9 / 11,5 / 12,5 (há somente pista(s) altamente coerente(s) com o contexto do diagnóstico em questão; ou há pista(s) altamente relevante(s) e coerente(s) com o contexto do diagnóstico em questão; ou há pista(s) altamente específica(s) e coerente(s) com o contexto do diagnóstico em questão; ou há pista(s) altamente relevan-

te(s), específica(s) e coerente(s) com o contexto do diagnóstico em questão - Categoria de Acurácia: **alta** (DE acurado ou DE muito acurado).

Observa-se que as três categorias de acurácia geradas são teoricamente pertinentes. A acurácia *nula* aplica-se aos julgamentos em que não há pistas para o diagnóstico ou de que as pistas existentes são de baixa relevância, baixa especificidade e baixa coerência. A categoria de acurácia *moderada* aplica-se aos julgamentos em que há somente pistas altamente relevantes e/ou altamente específicas. Nessa categoria, o julgamento é também o de que as pistas não são coerentes com o conjunto de dados existentes, apesar de serem relevantes, específicas ou ambas. A categoria *alta acurácia* aplica-se aos julgamentos em que as pistas existentes são altamente coerentes, isoladamente ou junto com qualquer outra combinação quanto a especificidade e relevância das pistas.

A EADE, desenvolvida neste estudo, deve ser aplicada sobre um conjunto de dados escritos (dados de avaliação do paciente e lista de diagnósticos de enfermagem) e é composta por: 1. Orientações para aplicação do instrumento; 2. Tabela de lançamento dos julgamentos; 3. Sistema de pontuação das respostas A e B. No apêndice apresenta-se um exemplo de aplicação da EADE que contém esses elementos, um conjunto de dados escritos e um gabarito.

DISCUSSÃO

Os resultados confirmam que mesmo especialistas chegam a conclusões diagnósticas diferentes diante de um mesmo conjunto de dados. Essa situação não é exclusiva do diagnóstico de enfermagem, pois é identificada

também em outras áreas clínicas. Em todos os casos, observou-se que as especialistas julgaram de forma diferente cada um dos itens, o que foi confirmado pelos coeficientes *Kappa* (Tabela 3).

Todas as estimativas de *Odds Ratio* obtidas pelas EEG na análise univariada foram estatisticamente significantes frente ao alfa estabelecido ($\leq 0,05$) (Tabela 1), o que permite rejeitar a hipótese de nulidade de que os itens da EADE não predizem a acurácia do diagnóstico e atesta a validade da EADE. Nas EEG, combinando-se as respostas aos itens de relevância, especificidade e coerência, só coerência manteve a significância (Tabela 2). Com base na literatura^(2,13), defendíamos que o item *relevância* seria o mais importante para prever a acurácia do diagnóstico enunciado, porque não basta haver pista para o diagnóstico em questão, é necessário que essa pista seja importante (relevante) para enunciá-lo. Acreditávamos também que o item *especificidade* assumiria o segundo maior escore da escala. É comum o compartilhamento de pistas entre diagnósticos de enfermagem^(2,5) e uma pista específica seria fundamental para afirmar um diagnóstico acurado. Um valor menor era previsto para o item *coerência*, visto que esse item não analisa a pista frente ao diagnóstico propriamente dito, mas sim frente ao conjunto de dados em que a mesma está inserida. Os resultados contrariaram as expectativas, pois o item *coerência* teve o maior poder preditivo para o *acerto* do diagnóstico (Tabela 2).

Os dados obtidos por meio da análise multivariada foram usados para atribuir *pesos* para os itens da EADE gerando escores ordinais que expressam o caráter de continuidade da variável. Foram obtidos oito escores. A interpretação desses escores, com base nas definições teóricas e validadas dos itens da EADE, norteou a redução dos oito escores para três categorias de acurácia: acurácia alta, acurácia moderada e acurácia nula.

Quanto à confiabilidade da EADE, os coeficientes gerais do *Kappa* variaram de 0,242 (especificidade) a 0,329 (presença de pista). Os resultados mostraram que a confiabilidade foi baixa (*fair*)⁽¹⁴⁾. No entanto, a estimativa de confiabilidade baixa deve ser analisada considerando-se as características da EADE e o contexto em que ela é aplicada.

A aplicação da EADE exige que cada avaliador refaça parte do raciocínio diagnóstico necessário para estabelecer cada diagnóstico. Nesse sentido, é razoável contrastar os resultados aqui obtidos com os de estudos sobre a acurácia de diagnósticos clínicos. Estudos dessa natureza são realizados com diagnósticos em outros contextos, com a diferença de que esses estudos são realizados sobre apenas um diagnóstico enquanto que, no presente estudo, cada avaliador julgou 43 diagnósticos.

Em estudo sobre a confiabilidade inter-observador do sistema de classificação de úlceras de pressão da *European Pressure Ulcer Advisory Panel*, no qual 1452 enfermeiras de cinco países europeus classificaram 20 fo-

tografias validadas de lesões de pele, o *Kappa* estimado foi de 0,33⁽¹⁵⁾. Em revisão de literatura sobre a acurácia de diagnóstico quanto à qualidade, quantidade e densidade óssea antes e durante a realização de implantes dentários, os autores identificaram valores de *Kappa* entre 0,33 e 0,67⁽¹⁶⁾. Resultados como esses sugerem que os obtidos no presente estudo são compatíveis com o que se observa em termos de concordância entre-observadores nos diagnósticos clínicos.

CONCLUSÃO

Este estudo permitiu criar uma escala com quatro itens (presença, relevância, especificidade e coerência das pistas) para avaliação da acurácia de diagnósticos de enfermagem a partir de dados escritos. A EADE foi desenvolvida para ser aplicada por avaliadores que tenham experiência no uso de classificações de diagnósticos de enfermagem e conhecimento na área clínica da situação do paciente cujos diagnósticos serão avaliados. A avaliação dos diagnósticos de enfermagem é baseada nos registros das pistas (características definidoras) que os mesmos apresentarem.

De posse dos dados escritos da avaliação do paciente e da listagem dos diagnósticos enunciados, o avaliador julga se há pistas para cada diagnóstico formulado. Se houver pistas, julga-as quanto aos graus de relevância e de especificidade das pistas frente ao diagnóstico avaliado e quanto ao grau de coerência frente ao conjunto dos dados disponíveis. As respostas a cada item da EADE, com exceção do item *presença de pistas*, correspondem a um escore (Relevância alta = 1 / Especificidade alta = 3,5 / Coerência alta = 8), e a soma dos mesmos resulta em um escore final que indica o grau de acurácia do diagnóstico avaliado (0 / 1 / 3,5 / 4,5 / 8 / 9 / 11,5 / 12,5). Por fim, com base no grau de acurácia obtido é possível identificar a categoria de acurácia em que se enquadra (alta, moderada ou nula).

Recomenda-se que no uso da EADE sejam relatados os graus de acurácia obtidos e também as categorias em que foram classificados.

A EADE mostrou estimativas razoáveis de validade e confiabilidade. Ela foi testada para ser utilizada com dados escritos e qualquer outra forma de aplicação requer prévias estimativas de validade e confiabilidade.

Este estudo mostra grandes avanços em relação ao seu precursor⁽⁵⁾, em que a versão traduzida da Escala de Acurácia de Diagnóstico de Lunney - LSM⁽²⁾ não mostrou estimativas aceitáveis de confiabilidade no uso com dados escritos.

Das limitações deste estudo, destacam-se a fragilidade do *padrão-ouro* estabelecido e a baixa concordância entre avaliadores produzidas pela EADE. O que motivou a definição do *padrão-ouro* utilizado no estudo foi a inexistência de alternativa melhor. A baixa concordância obtida parece ser consistente com o que se obtém em diagnós-

ticos clínicos, mas a busca de estratégias para melhorar a concordância no uso da EADE é o principal desafio para o seu refinamento. Recomenda-se que, para usar a EADE,

os avaliadores sejam treinados previamente com estudos de caso e que tenham adequada compreensão das definições dos itens.

REFERÊNCIAS

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 Psiqu 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.

APÊNDICE

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

Escore para os itens da EADE				
Itens		Definições	Categorias	Escore 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 cada item e some todos os escores de cada diagnóstico. O escore total é interpretado da seguinte forma:

Categorias de acurácia da EADE		
Escore 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