

CONFIABILIDADE DE ESCALAS DE COMPROMETIMENTO NEUROLÓGICO EM PACIENTES COM ACIDENTE VASCULAR CEREBRAL

Marco Aurélio Gralha de Caneda¹, Jefferson Gomes Fernandes²,
Andrea Garcia de Almeida³, Fabiana Eloisa Mugno⁴

RESUMO - Escalas de comprometimento neurológico são utilizadas na avaliação de novos tratamentos, no acompanhamento da evolução clínica e em decisões terapêuticas. Estudamos a confiabilidade de três escalas: a Escala de Rankin (ER), o Índice de Barthel (IB) e a National Institute of Health Stroke Scale (NIHSS) versadas em português, em 51 indivíduos com acidente vascular cerebral, estimando a concordância entre resultados de diferentes examinadores. Utilizamos os coeficientes de concordância Kappa e Coeficiente de Correlação Intraclass. A ER teve coeficientes moderados, substanciais ou excelentes. No IB estes foram substanciais ou excelentes nos totais e nos itens, assim como quando foi estratificado em agrupamentos prognósticos e funcionais. A NIHSS apresentou coeficientes excelentes nos totais, substanciais nos itens e moderados e excelentes nas estratificações em grupos de pontos e de itens afins. Estes resultados indicam que as versões em português da ER, o IB e a NIHSS apresentam adequada confiabilidade.

PALAVRAS-CHAVE: escala de Rankin, índice de Barthel, NIHSS, confiabilidade, acidente vascular cerebral.

Reliability of neurological assessment scales in patients with stroke

ABSTRACT - Scales for the assessment of neurological impairment are employed in clinical trials, outcome evaluation and in therapeutic decisions. We evaluated the reliability of the Portuguese version of the Rankin Scale (RS), Barthel Index (BI) and the NIHSS (National Institute of Health Stroke Scale) in 51 stroke patients, estimating the agreement of the results for examiners. The interrater reliability was measured using the Kappa Coefficient and Intraclass Correlation Coefficient. The RS showed moderate, substantial and excellent coefficients of agreement. The BI showed from substantial to excellent coefficients of agreement as a whole and also for its constituents. This scale was stratified in prognostic and functional groups, both presenting coefficients of agreement from substantial to excellent. The NIHSS presented excellent agreement on its total and substantial agreements on its individual items. The NIHSS was also stratified in groups of points, which presented moderate or excellent coefficient, and in groups of congruous items, which showed a moderate to excellent agreement. These results suggest that the Portuguese version of the RS, BI and the NIHSS present adequate reliability.

KEY WORDS: Rankin scale, Barthel index, NIHSS, reliability, stroke.

O acidente vascular cerebral (AVC) é importante problema de saúde pública no mundo. Suas consequências podem ser de grande impacto, gerando enormes demandas de recursos em diagnóstico, tratamento e reabilitação¹. É importante causa de mortalidade, sendo a primeira em algumas regiões e ficando atrás apenas das cardiopatias e neoplasias em outras². No Brasil, o AVC também tem sido patologia responsável por altas taxas de mortalidade, invalidez e de elevado custo social e econômico³. Nos últimos

anos têm sido desenvolvidas estratégias de intervenção preventiva e novas terapêuticas medicamentosas para a fase aguda. O advento da trombólise na isquemia é o corolário destes avanços⁴, assim como a tentativa de obtenção de medicamentos com ação neuroprotetora realmente eficaz. A eficácia destas novas modalidades de tratamento é comprovada pelo efeito clínico de melhora ou de redução do comprometimento neurológico dos pacientes. A mensuração desta eficácia é obtida através da aplicação de escalas

¹Mestre em Medicina e Ciências da Saúde pelo CPG da Faculdade de Medicina da Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brasil (FAMED-PUCRS) e Médico Neurologista do Centro Neurovascular do Hospital Moinhos de Vento, Porto Alegre, Brasil; ²Professor Titular de Medicina Interna da FAMED-PUCRS e Diretor do Instituto de Educação e Pesquisa da Associação Hospitalar Moinhos de Vento, Porto Alegre, Brasil; ³Mestranda em Clínica Médica, Área de Neurociências na FAMED-UFRGS; ⁴Mestranda em Medicina e Ciências da Saúde no CPG da FAMED-PUCRS.

Recebido 10 Outubro 2005, recebido na forma final 19 Janeiro 2006. Aceito 3 Abril 2006.

Dr. Marco Aurélio G. Caneda - Rua Dr. Alcides Cruz 81 / 603 - 90630-160 Porto Alegre RS -Brasil. E-mail: mcaneda@terra.com.br

que dimensionam este comprometimento. Estas são úteis também no acompanhamento da evolução do quadro clínico dos pacientes e na antecipação de seu prognóstico^{5,6}. Uma escala pode mensurar o dano neurológico: perda ou anormalidade de uma função psicológica, fisiológica ou anatômica; pode abordar a incapacidade: restrição ou carência resultante na habilidade de executar uma tarefa dentro de padrões considerados normais; ou ainda avaliar a desvantagem ou invalidez: prejuízos que afetam o indivíduo em seu contexto social após seu acometimento⁷.

Dentre os critérios clinimétricos que uma escala deve respeitar encontra-se a confiabilidade, ou a qualidade de um determinado método em reproduzir seus resultados em repetidas aplicações com a menor variabilidade possível⁸⁻¹⁰. O grau de confiabilidade de uma escala pode ser estimado mensurando-se a concordância entre resultados obtidos por diferentes examinadores em sua aplicação a um mesmo paciente. A concordância entre-examinadores é obtida através do cálculo de coeficientes de concordância⁸. Estes fornecem as proporções de concordâncias que são esperadas ocorrer ao acaso ou por real concordância entre-examinadores. Os coeficientes mais empregados na literatura são o Kappa (α) e Kappa Ponderado ($\alpha\alpha$) e o Coeficiente de Correlação Intraclassa (ICC). Estes são expressos matematicamente em valores que variam de -1 a 1, que indicam a discordância ou concordância totais, respectivamente. A concordância ao acaso é indicada pelo valor 0. É possível classificar a concordância entre resultados, e, assim, inferir a confiabilidade. Coeficientes variando de 0 a 0,20 são de concordância fraca; de 0,21 a 0,40, regular; de 0,41 a 0,60, moderada; 0,61 a 0,80, substancial e de 0,81 a 1,0, excelente. Coeficientes maiores que 0,60 indicariam uma confiabilidade clinicamente satisfatória⁹⁻¹².

As escalas mais utilizadas para avaliar o comprometimento do AVC são: Escala de Rankin^{13,14}, Índice de Barthel^{14,15} e a National Institute of Health Stroke Scale (NIHSS)¹⁶. O presente estudo teve por finalidade avaliar a confiabilidade de versões em português destas três escalas em pacientes com AVC.

MÉTODOS

Estudo transversal da confiabilidade da Escala de Rankin (ER), do Índice de Barthel (IB) e da National Institute of Health Stroke Scale (NIHSS). A base populacional deste estudo foi hospitalar, com pacientes provindos da região metropolitana de Porto Alegre, RS, e que foram atendidos no Hospital São Lucas da PUCRS.

Foram incluídos consecutivamente 51 pacientes (n mínimo necessário de 40 indivíduos, para margem de erro de

0,20 e α de 0,05)¹⁷, maiores de 18 anos com diagnóstico de AVC em fase aguda ou subaguda. Este diagnóstico foi realizado clinicamente por neurologista, ou médico residente em neurologia, e complementado por tomografia computadorizada ou ressonância magnética de encéfalo. Foram excluídos pacientes com outras patologias neurológicas que não tivessem origem vascular, aqueles com ataques isquêmicos transitórios e menores de 18 anos de idade.

As escalas empregadas, traduzidas e adaptadas para o idioma português (podem ser obtidas diretamente com o autor do trabalho), foram aplicadas por dois examinadores, A.G.A. (examinador 1 - E1) e M.A.G.C. (examinador 2 - E2), em observação direta dos pacientes. Não permitiu-se intervalo maior que 24 horas entre as aplicações, nem aplicações imediatamente consecutivas ou simultâneas, objetivando evitar flutuações no quadro clínico, causar fadiga ou estimular um "esforço em acertar" dos pacientes. Os examinadores receberam treinamento prévio em seminários conjuntos, simulações em pacientes e em fitas de videotape com material de treinamento específico.

A concordância entre-examinadores foi estimada pelos coeficientes α , $\alpha\alpha$ e ICC, calculados nos resultados obtidos em cada item das escalas, em suas estratificações ou nos escores totais. Na ER os coeficientes α , $\alpha\alpha$ e ICC foram calculados nos escores totais. No IB os coeficientes α e $\alpha\alpha$ foram aplicados nos resultados dos itens individuais da escala, nos escores totais e nas estratificações, o ICC foi aplicado somente nos escores totais. Estratificamos o IB de duas maneiras: (a) Quanto à classificação prognóstica - Grupo I: 0 a 45 pontos=incapacidade severa; Grupo II: 50 a 70 pontos=moderada; Grupo III: 75 a 95 pontos=leve e Grupo IV: 100 pontos=independência funcional; (b) Quanto a agrupamentos funcionais - Grupo A: auto-cuidados (itens 1 a 7: alimentação, banho, apresentação pessoal, vestir, cuidados com intestinos e bexiga, e uso do banheiro) e Grupo B: mobilidade (itens 8 a 10: deambulação, transferência do leito para cadeira e subir escadas).

Na NIHSS os coeficientes α e $\alpha\alpha$ foram calculados nos itens individuais e nas estratificações. Nos escores totais foi calculado somente o ICC. Estratificamos a NIHSS da seguinte forma: (a) Segundo o escore total: em agrupamentos de quatro pontos, um marcador de alterações clínicas⁴, tendo o último seis pontos - Grupo 1=0 a 4; Grupo 2=5 a 8; Grupo 3=9 a 12; Grupo 4=13 a 16; Grupo 5=17 a 20; Grupo 6=21 a 24; Grupo 7=25 a 28; Grupo 8=29 a 32; Grupo 9=33 a 36 e Grupo 10=37 a 42 pontos; (b) Em agrupamentos funcionais: Grupo I=Consciência (itens 1 a, b e c: nível de consciência, nível de consciência-perguntas e nível de consciência-

Tabela 1. Pacientes e diagnóstico.

Item	n=51
Média de Idade	64,1 anos (d.p.* 13,9)
Homens : Mulheres	26(52%) : 25(48%)
Diagnóstico (AVC I : H)**	47(92%) : 4(8%)
Investigação (TC : RM)***	51(100%) : 4(8%)

*desvio padrão; **AVC I, acidente vascular cerebral isquêmico; H, hemorrágico; ***tomografia computadorizada e/ou ressonância magnética de encéfalo.

Tabela 2. Escala de Rankin: coeficientes de concordância entre-examinadores e comparação com estudos prévios.

Estudo	Nosso estudo	(20) [§]	(22) [§]	(21) [§]	(19) [§]
Ano	2005	1988	1991	1999	2002
n	51	65	50	43	58
α^*	0,45	0,56	-	-	0,44
IC**	0,31-0,60	-	-	-	-
$\alpha\alpha^{***}$	0,70	0,91	0,85	0,76	0,78
IC	0,58-0,90	-	0,47-1,00	-	-
ICC****	0,83	-	-	-	-
IC	0,72-0,90	-	-	-	-

*coeficiente Kappa; **intervalo de confiança 95%; ***coeficiente Kappa ponderado; ****coeficiente de correlação intraclass; [§]Número da referência bibliográfica.

Tabela 3. Índice de Barthel: coeficientes de concordância e comparação com estudos prévios.

Estudo		Nosso Estudo	(23) [†]	(22) [†]	(24) [†]	(26) [†]	(15) [†]	(25) [†]
Ano		2005	1988	1991	1993	1995	2001	2001
n		51	7	50	60	54	22	121
Item da Escala								
1	α^* (IC**)	0,91 (0,81-1,00)	1	1	-	0,81	-	-
	$\alpha\alpha^{***}$ (IC)	0,94 (0,87-1,00)	-	-	-	-	0,62	-
2	α (IC)	0,92 (0,81-1,00)	0,88	0,88	-	0,92	0,81	-
	$\alpha\alpha$ (IC)	-	-	-	-	-	-	-
3	α (IC)	0,96 (0,88-1,00)	0,47	0,47	-	0,42	0,78	-
	$\alpha\alpha$ (IC)	0,84 (0,73-0,94)	-	-	-	-	-	-
4	α (IC)	0,75 (0,60-0,90)	0,72	0,72	-	0,71	-	-
	$\alpha\alpha$ (IC)	0,84 (0,73-0,94)	-	-	-	-	0,78	-
5	α (IC)	0,84 (0,70-0,97)	1	1	-	-	-	-
	$\alpha\alpha$ (IC)	0,88 (0,77-0,99)	-	-	-	0,67	-	-
6	α (IC)	0,87 (0,75-0,99)	1	1	-	-	-	-
	$\alpha\alpha$ (IC)	0,92 (0,84-1,00)	-	-	-	0,52	-0,09	-
7	α (IC)	0,85 (0,72-0,97)	0,74	0,74	-	-	-	-
	$\alpha\alpha$ (IC)	0,88 (0,76-0,99)	-	-	-	0,63	-	-
8	α (IC)	0,71 (0,56-0,86)	0,59	0,59	-	-	-	-
	$\alpha\alpha$ (IC)	0,84 (0,73-0,94)	-	-	-	0,60	0,71	-
9	α (IC)	0,76 (0,62-0,90)	0,66	0,66	-	-	0	-
	$\alpha\alpha$ (IC)	0,88 (0,80-0,96)	-	-	-	0,91	-	-
10	α (IC)	0,87 (0,75-0,99)	1	1	-	-	-	-
	$\alpha\alpha$ (IC)	0,88 (0,76-1,00)	-	-	-	0,79	0,73	-
Ítems - médias	α	0,84	-	-	0,88	-	0,53	0,73
	$\alpha\alpha$	0,88	-	0,61	0,86	0,69	0,44	-
Escores totais	α	0,65	0,80	0,96	-	-	-	-
	$\alpha\alpha$	0,89	-	-	-	-	-	-
	ICC [¶]	0,95	-	-	0,94	-	-	0,99

*coeficiente Kappa; **intervalo de confiança 95%; ***coeficiente Kappa ponderado; [¶]coeficiente de correlação intraclass; [†]número da referência bibliográfica.

comandos); Grupo II=Sensitivo-sensorial (itens 3, 8 e 11: campos visuais, sensibilidade e extinção-inatenção); Grupo III=Motor (itens 2, 4, 5, 6, 7 e 10: movimentação ocular, paralisia facial, motricidade de membros superiores e inferiores, ataxia e disartria) e Grupo IV=Linguagem (item 9: linguagem).

Para a obtenção dos coeficientes de concordância utilizamos os programas Kappa Version 3.0 com PABAK (kappa ajustado para viés de prevalência) e ICC- SPSS Version 11 para Windows.

Este estudo foi aprovado pelo Comitê de Ética em

Pesquisa da Pontifícia Universidade Católica do Rio Grande do Sul e foram obtidos termos de consentimento livre e informado dos pacientes incluídos.

RESULTADOS

Entre os pacientes incluídos, 26 (52%) eram mulheres e 25 (49%), homens. A média de idade foi 64,1 (25 a 89 anos). O diagnóstico clínico foi corroborado por método de imagem em 92% dos casos, sendo estabelecido diagnóstico puramente clínico em 8%. Em

Tabela 4. NIHSS: coeficientes de concordância em comparação com estudos prévios.

Estudo	Nosso Estudo	(16) [¶]	(29) [¶]	(28) [¶]	(30) [¶]	(27) [¶]	(33) [¶]	(34) [¶]
Ano	2004	1989	1989	1994	1994	1999	2001	2002
n	51	24	20	6 ^a	11 ^a	31	59 ^b	45
Item	Coefic.							
1a	α^* (IC**) 0,91(0,79-1,00)	0,49	0,50	0,48	0,52(0,32-0,69)	0,83	-	-
	$\alpha\alpha^{***}$ (IC) 0,95(0,89-1,00)	-	-	-	-	0,58(0,29-0,88)	0,66(0,34-0,96)	0,45(0,16-0,74)
1b	α (IC) 0,78(0,62-0,94)	0,80	0,64	0,67	0,79(0,75-0,81)	0,62	-	-
	$\alpha\alpha$ (IC) 0,84(0,71-0,97)	-	-	-	-	0,56(0,28-0,81)	0,50(0,09-0,92)	0,93(0,64-1,22)
1c	α (IC) 0,76(0,57-0,95)	0,58	0,41	0,74	0,46(0,06-0,82)	0,87	-	-
	$\alpha\alpha$ (IC) 0,80(0,63-0,97)	-	-	-	-	0,70(0,42-0,97)	0,53(0,10-0,91)	0,94(0,65-1,23)
2	α (IC) 0,84(0,70-0,98)	0,82	0,33	0,74	0,26(-0,01-0,54)	0,81	-	-
	$\alpha\alpha$ (IC) 0,88(0,77-0,98)	-	-	-	-	0,44(0,11-0,83)	0,60(0,14-0,94)	0,66(0,36-0,95)
3	α (IC) 0,52(0,30-0,73)	0,81	0,57	0,80	0,87(0,80-0,92)	0,79	-	-
	$\alpha\alpha$ (IC) 0,61(0,39-0,84)	-	-	-	-	0,71(0,49-0,93)	0,43(0,25-0,66)	0,87(0,58-1,16)
4	α (IC) 0,29(0,09-0,49)	0,57	0,22	0,29	0,29(0,25-0,31)	0,65	-	-
	$\alpha\alpha$ (IC) 0,42(0,25-0,59)	-	-	-	-	0,70(0,53-0,87)	0,71(0,39-0,92)	0,74(0,46-1,02)
5a	α (IC) 0,69(0,53-0,85)	0,85	0,77	0,95	0,86(0,76-0,93)	0,75	-	-
	$\alpha\alpha$ (IC) 0,85(0,76-0,94)	-	-	-	-	0,81(0,66-0,95)	0,77(0,60-0,91)	0,97(0,67-1,26)
5b	α (IC) 0,79(0,65-0,93)	¥	¥	0,79	0,85(0,79-0,90)	0,81	¥	-
	$\alpha\alpha$ (IC) 0,89(0,80-0,98)	-	-	-	-	0,81(0,60-0,99)	-	0,95(0,66-1,24)
6a	α (IC) 0,73(0,59-0,88)	0,83	0,78	0,80	0,68(0,63-0,74)	0,64	-	-
	$\alpha\alpha$ (IC) 0,86(0,77-0,95)	-	-	-	-	0,66(0,50-0,83)	0,75(0,53-0,96)	0,94(0,65-1,24)
6b	α (IC) 0,78(0,63-0,93)	¥	¥	0,51	0,87(0,82-0,91)	0,49	¥	-
	$\alpha\alpha$ (IC) 0,87(0,77-0,98)	-	-	-	-	0,51(0,24-0,78)	-	0,97(0,68-1,26)
7	α (IC) 0,20 ¹ (-0,22-0,62)	0,57	-0,16	0,47	0,39(0,30-0,61)	0,52	-	-
	$\alpha\alpha$ (IC) 0,24(-0,21-0,68)	-	-	-	-	0,17(-0,15-0,68)	0,81(0,57-1,0)	0,69(0,40-0,97)
8	α (IC) 0,62(0,43-0,81)	0,60	0,50	0,58	0,87(0,80-0,92)	0,51	-	-
	$\alpha\alpha$ (IC) 0,69(0,52-0,85)	-	-	-	-	0,54(0,31-0,78)	0,70(0,42-0,95)	0,89(0,60-1,18)
9	α (IC) 0,63(0,45-0,80)	0,64	0,79	0,65	0,48(0,44-0,51)	0,61	-	-
	$\alpha\alpha$ (IC) 0,78(0,67-0,90)	-	-	-	-	0,62(0,35-0,90)	0,67(0,36-0,96)	0,84(0,55-1,12)
10	α (IC) 0,78(0,63-0,92)	0,55	0,32	0,51	0,57(0,49-0,62)	0,68	-	-
	$\alpha\alpha$ (IC) 0,86(0,76-0,96)	-	-	-	-	0,63(0,38-0,86)	0,76(0,41-1,0)	0,28(0,05-0,58)
11	α (IC) 0,32 ² (0,06-0,58)	0,58	0,61	0,67	0,53(0,47-0,58)	0,74	-	-
	$\alpha\alpha$ (IC) 0,35(0,06-0,63)	-	-	-	-	0,65(0,35-0,94)	0,77(0,40-1,0)	0,89(0,59-1,18)
α (médio - itens)	0,64	0,69	0,48	0,64	0,61	0,68	-	-
$\alpha\alpha$ (médio -itens)	0,82	-	-	-	-	0,60	0,66	0,80
ICC****(IC)	0,96(0,94-0,98)	-	-	0,95	-	0,92/ 0,96	0,76(0,61-0,90)	-

*coeficiente Kappa ;**intervalo de confiança 95%; ***coeficiente Kappa ponderado; ****coeficiente de correlação intraclassa; [¶]Número da referência bibliográfica; ^apacientes em videotape; ^bestudo retrospectivo; ¹Kappa com "viés de prevalência". Corrigido por PABAK, $\alpha=0,82$; ²Kappa com "viés de prevalência". Corrigido por PABAK, $\alpha=0,62$; [¥]Não publicado. Consta "coeficiente motor dos membros".

Tabela 5. NIHSS: coeficientes de concordância entre-examinadores nas estratificações.

Estratificação	α^*	IC** α	$\alpha\alpha^{***}$	IC $\alpha\alpha$
A) Por pontuação				
Total	0,52	(0,38-0,67)	0,84	(0,77-0,90)
B) Grupos de itens				
I) Consciência	0,81	(0,71-0,90)	0,86	(0,79-0,94)
II) Sensitivo-sensorial	0,58	(0,47-0,69)	0,61	(0,49-0,73)
III) Motor	0,69	(0,64-0,75)	0,82	(0,78-0,86)
IV) Linguagem	0,63	(0,45-0,80)	0,78	(0,67-0,90)
Média	0,67	–	0,76	–

*coeficiente Kappa; **intervalo de confiança 95%; ***coeficiente Kappa ponderado.

47 pacientes (92,1%) o AVC foi isquêmico e em 4 (7,8%) foi hemorrágico (Tabela 1). Em média, as aplicações foram finalizadas pelos examinadores no 1,8º dia de internação (dia 0 ao 14º dia). O intervalo médio entre as aplicações do E1 e E2 foi 11 horas e 14 minutos (60 minutos à 24 horas).

Escala de Rankin – Foi obtido coeficiente α indicativo de concordância moderada entre-examinadores, um $\alpha\alpha$ substancial e um ICC excelente (Tabela 2).

Índice de Barthel – Nos escores totais foram encontrados coeficiente α 0,65, concordância substancial e $\alpha\alpha$ 0,89, excelente. Os resultados nas estratificações também foram satisfatórios: na estratificação por agrupamentos prognósticos o α obtido foi 0,76, concordância substancial e o $\alpha\alpha$ 0,84, excelente. Na estratificação por agrupamentos funcionais os resultados foram semelhantes: no Grupo A (auto-cuidados), o coeficiente α foi 0,88 e o $\alpha\alpha$ 0,91, ambos de excelente concordância. No Grupo B (mobilidade), o α foi 0,79, concordância substancial, e o $\alpha\alpha$ 0,87, excelente. O coeficiente α dos itens individuais e o ICC dos escores totais estão expostos na Tabela 3.

NIHSS – Nos escores totais o coeficiente ICC obtido foi 0,96, concordância excelente. As Tabelas 4 e 5 apresentam, respectivamente, os coeficientes obtidos nos itens individuais e nas estratificações da escala.

DISCUSSÃO

De maneira geral as escalas avaliadas mostraram, em suas versões em português, confiabilidade adequada para uso em pacientes com AVC em nosso meio. Os intervalos entre as aplicações foram satisfatórios, pois permitiram boa recuperação aos pacientes não estafando-os com repetições, amenizando

possível compenetração ou maior esforço voluntário destes e, fundamentalmente, evitando flutuações clínicas.

O emprego de diferentes coeficientes de concordância deveu-se ao interesse em compará-los com dados da literatura. Uma das limitações do coeficiente α é que ele não contempla em que magnitude ocorre a discordância. O coeficiente $\alpha\alpha$ reduz este viés ponderando esta discrepância^{10,11,16}. O ICC considera as proporções de variância total, devida a verdadeira variabilidade entre os pacientes, nos examinadores e ao erro residual do método^{12,18}. Assim, possivelmente exprime com maior fidedignidade a concordância entre-examinadores.

Escala de Rankin – Poucas publicações anteriores abordaram a concordância entre-examinadores. O α de nosso estudo foi 0,45, praticamente igual ao de um estudo britânico, em que foi 0,44¹⁹, porém abaixo do estudo pioneiro, que obteve um α 0,56 (Tabela 2)²⁰. Estes coeficientes são de moderada concordância. É importante frisar que o coeficiente α pode apresentar distorções geradas por altas concentrações de respostas em uma categoria, um viés de “prevalência”. Em nosso estudo 66,6% das concordâncias ocorreram no escore “5”. Provavelmente houve influência da fase aguda da doença, pois, sendo esta uma escala de desvantagem, com categorias intermediárias abordando a marcha, tais classificações poderiam ter sido geradas pela restrição dos pacientes ao leito. Além disso, a pontuação extrema apresenta menor potencial de variabilidade, as intermediárias são mais suscetíveis à subjetividade, o que poderia representar um problema da escala quando aplicada em fases precoces de doença.

O coeficiente $\alpha\alpha$ foi substancial, 0,70, como o de um estudo de 2002, quando foi 0,78¹⁹. Em 1999 um

grupo alemão, testando a confiabilidade de uma adaptação da ER a seu idioma em 43 pacientes, obteve α 0,76, semelhante ao de nosso estudo²¹. No início dos anos 90, um trabalho encontrou um α 0,85, excelente. No entanto, o número de pacientes de cada série de resultados comparados não era igual, o que pode ter afetado o valor deste coeficiente²². Outra publicação, do final da década de 80, apresentou α 0,91, excelente, porém contava com pacientes em fase aguda e crônica, o que pode, pela estabilidade clínica, gerar melhor concordância²⁰. Não encontramos publicação anterior utilizando o coeficiente ICC na ER.

Índice de Barthel – Mostrou-se mais objetivo que a ER, o que implicou em seus coeficientes mais elevados. Os coeficientes dos escores totais, α 0,65, substancial, e α 0,89, excelente, são de classificação semelhante aos encontrados na literatura. Um estudo canadense apresentou α 0,80, também substancial²³. Outro trabalho publicou coeficiente α 0,96 nos totais²². O coeficiente ICC nos totais gerais da escala, 0,95, excelente, é comparável aos resultados publicados anteriormente de 0,94 e 0,99 (Tabela 3)²⁴⁻²⁶.

Os coeficientes k da maioria dos itens foram excelentes, sendo que apenas três (itens 4, 8 e 9), foram substanciais. A média foi 0,84. Um estudo do início dos anos 90 apresentou α médio de 0,8824, bastante semelhante ao nosso. Recentemente uma publicação chinesa apresentou α médio dos itens de 0,73, substancial²⁵, e um estudo inglês, de 0,53, moderado¹⁵. Estas variadas classificações seriam atribuíveis a diferentes metodologias destes estudos em relação ao nosso, sobretudo no intervalo de aplicação das escalas pelos examinadores e na fase de doença dos indivíduos.

Os coeficientes α foram excelentes, com média 0,88. Um estudo anterior, apresentou coeficientes mais discretos, com α médio 0,61, substancial²⁶. Neste estudo as aplicações não foram diretas, mas com coleta de informações junto a cuidadores ou familiares, o que parece ter acarretado este coeficiente reduzido. Pior resultado obteve um estudo de 2001, onde o α foi apenas moderado, de 0,44¹⁵. Um terceiro estudo apresentou α médio de seus itens de 0,69²⁵, também substancial.

Um aspecto inusitado de nosso estudo foi a estratificação do IB. Na primeira modalidade, com itens divididos em grupos de auto-cuidados ou mobilidade, a finalidade foi identificar uma possível diferença de concordância "intraescala". Teoricamente, itens de auto-cuidados, que envolvem utilização de funções

cognitivas para sua execução, seriam de realização mais complexa que itens abordando a mobilidade. Discrepâncias marcantes em algum dos grupos prejudicariam a homogeneidade da escala, não sendo possível, por exemplo, atribuir igual pontuação a todos seus componentes. No entanto, houve pouca diferença entre-examinadores, visto que os coeficientes foram excelentes, com exceção do " α mobilidade", substancial. Um único estudo anterior abordou superficialmente este dado, mas diferentemente de nossos resultados, apenas afirmava que itens avaliando tarefas motoras teriam tido uma concordância mais elevada que os de auto-cuidados²⁶. A segunda modalidade de estratificação objetivou saber se examinadores diferentes classificariam os indivíduos em iguais patamares prognósticos. Os coeficientes nesta estratificação foram substanciais ou excelentes.

NIHSS – O coeficiente ICC 0,96, excelente, nos totais gerais é comparável àqueles descritos na literatura. Um estudo de 1999 apresentou coeficientes ICC variando de 0,92 a 0,96²⁷, com cerca da metade dos indivíduos apresentando-se em fase crônica. Entre os examinadores deste estudo havia um profissional não-médico, cuja presença, no entanto, não alterou os coeficientes. Dois outros estudos, também com concordâncias excelentes, ICC de 0,95 e 0,96^{5,28}, foram realizados entre vários examinadores, também com não-médicos ou médicos não-neurologistas, mas em pacientes apresentados em fitas de videotape, o que também não causou impacto na concordância.

Em nosso estudo, 76% dos coeficientes dos itens tiveram concordância substancial ou excelente. Em contrapartida, 16% dos coeficientes foram fracos ou regulares. Isto ocorreu nos itens "4. paralisia facial" (α 0,29 e α 0,42), "7. ataxia" (α 0,20 e α 0,24) e "11. extinção-inatenção" (α 0,32 e α 0,35). É importante frisar que a concordância fraca à regular nos itens ataxia e extinção e inatenção, deu-se a partir de um fenômeno estatístico de concentração de concordâncias, ou "viés de prevalência", em uma só categoria. Como há uma frequência aguardada ao acaso em cada categoria, a ausência ou baixa prevalência gera grande distorção nos coeficientes kappa¹¹. Quando corrigidos pelo PABAK, o item 7 teve coeficiente 0,82 e o item 11, 0,62. Restaria saber se os baixos coeficientes registrados neste item pela literatura seriam também fruto deste viés estatístico^{16,27,29}.

Mesmo em estudos com pacientes apresentados em videotape, em que a variabilidade natural do quadro clínico está eliminada, estes itens foram problemáticos^{28,30}. Possivelmente sua própria natureza im-

põe dificuldades, como é atestado por estudos avaliando a concordância na execução do exame neurológico clássico³¹. Outros dois itens têm sido alvo de críticas por seus baixos coeficientes de concordância: consciência e disartria. Estudos anteriores, realizados em pacientes ou videotape, apresentaram coeficientes de fraca, regular ou no máximo moderada concordância^{16,28-30}. No nosso estudo esta dificuldade não se repetiu. Nos itens consciência os coeficientes foram excelentes. É possível que este efeito tenha ocorrido pela pequena variabilidade nos pacientes, com 75% destes classificados na pontuação "0", ou seja, na normalidade. Este argumento não justificaria os coeficientes, também satisfatórios, no item disartria. Aqui, acreditamos ficar evidente o que fora demonstrado em estudo anterior: o treinamento prévio dos examinadores exerce grande papel e pode ser decisivo nos coeficientes de concordância³². Também é importante destacar que o item movimentação ocular obteve excelente concordância em nosso estudo em oposição a alguns estudos anteriores^{29,30}. Os coeficientes médios dos itens foram substanciais ou excelentes estando em acordo com coeficientes descritos em estudos anteriores (Tabela 4)^{27,30,33}.

Um aspecto inovador deste estudo foi a análise de estratificações da NIHSS. Na primeira modalidade os escores totais foram divididos em grupos de quatro pontos com finalidade de avaliar se havia concordância clinicamente aceitável na classificação do dano neurológico. O coeficiente α foi moderado, 0,52. Como as discrepâncias existentes foram em grau pequeno, o coeficiente $\alpha\alpha$ foi excelente, 0,84. Em outra modalidade os itens foram divididos em grupos de funções afins, com finalidade de identificar possíveis reduções "intraescala" de concordância, representando focos de distorção na sua homogeneidade. No Grupo I - consciência, a concordância foi excelente, α 0,81 e $\alpha\alpha$ 0,86. Não se justificaria, a nosso ver, a retirada destes itens numa virtual NIHSS modificada, como anteriormente proposto na literatura³⁴. O pior rendimento foi do Grupo II - sensitivo-sensorial, que, mesmo assim, apresentou coeficientes moderados. Provavelmente a dificuldade seja inerente à natureza dos componentes deste grupo, os quais foram até mesmo desaconselhados para emprego em escalas⁶. Entretanto, sua abolição representaria perdas na sensibilidade, a qual também é propriedade fundamental de uma escala. Os Grupos III - motor e IV - linguagem, apresentaram coeficientes satisfatórios de concordância (Tabela 5).

As versões das três escalas mostraram-se compre-

ensíveis aos pacientes. Uma pequena dificuldade surgiu na identificação de algumas figuras da NIHSS, em ocasiões chegando a suscitar dúvidas a alguns pacientes, mas sem prejudicar seu desempenho. A ER, em que pese seu caráter subjetivo, exibiu confiabilidade substancial. Entretanto, houve impressão de que se empregada em momento de maior estabilidade da doença, poderia obter classificação maior de concordância. É possível que aliada ao IB seja um importante instrumento de prognóstico na doença cerebrovascular. O IB exibiu maior concordância, mostrando ser instrumento de ótima confiabilidade, simetricamente, não havendo discrepâncias entre seus grupos de itens. Parece ser não só importante na predição prognóstica, como ficou demonstrado por uma de suas estratificações, mas também valioso na avaliação e planejamento de terapias auxiliares na reabilitação dos pacientes. Um aspecto identificado neste estudo foi o melhor desempenho desta escala quando em observação direta do paciente em comparação àqueles com avaliações indiretas, achado que talvez merecesse uma abordagem específica em futuros ensaios. Por último, a NIHSS, que teve tradução e adaptação original para o português, apresentou concordância similar a de sua versão em inglês, inclusive em suas debilidades. Foi interessante que um de seus itens, considerado como um ponto fraco da escala, tenha entre nós apresentado-se também problemático, porém às custas de um viés estatístico. Esclarecimentos em futuros estudos seriam providenciais e oportunos.

Em conclusão, a ER, o IB e a NIHSS, em versões e adaptações para o nosso idioma, mostraram-se como escalas de avaliação de comprometimento neurológico com evidências de confiabilidade clinicamente aceitável e de boa aplicabilidade em nosso meio. Os resultados deste estudo provêm o suporte para a utilização das versões destas escalas em pesquisas, tais como ensaios clínicos, e para padronização da avaliação diagnóstica e prognóstica na assistência a pacientes com AVC no Brasil.

REFERÊNCIAS

1. Dobkin B. The economic impact of stroke. *Neurology* 1995;45(Suppl 1):S6-S9.
2. Adams HP, Brott TG, Crowell RM. Guidelines for the management of patients with acute ischemic stroke. *Stroke* 1994;25:1901-1914.
3. Lessa I. Epidemiologia das doenças cerebrovasculares no Brasil. *Rev Soc Cardiol Estado de São Paulo* 1999;9:509-518.
4. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med* 1995;333:1581-1587.
5. Goldstein LB, Samsa GP. Reliability of the National Institute of Health Stroke Scale. *Stroke* 1997;28:307-310.

6. Orgogozo JM. Advantage and disadvantage of neurological scales. *Cerebrovasc Dis* 1998;8(Suppl 2):S2-S7.
7. Orgogozo JM. The concepts of impairment, disability and handicap. *Cerebrovasc Dis* 1994;4(Suppl 2):S2-S6.
8. Côté R, Batista RN, Wolfson CM, Hachinski V. Stroke assessment scales: guidelines for development, validation and reliability assessment. *Can J Neurol Sci* 1988;15:261-265.
9. Hantson L, Keyser JD. Neurologic scales in the assessment of cerebral infarction. *Cerebrovasc Dis* 1994;4(Suppl 2):S7-S14.
10. Lyden PD, Lau GT. A critical appraisal of stroke evaluation and rating scales. *Stroke* 1991;22:1345-1352.
11. Altman DG. *Practical statistics for medical research*. London: CRC Press, 1990:396-409.
12. Everitt, BS. *Statistical methods for medical investigations*. London: Oxford University Press, 1989:16-28.
13. Haan DR, Limburg M, Bossuyt P, Meulen JVD, Aaronson N. The clinical meaning of rankin handicap grades after stroke. *Stroke* 1995;26: 2027-2030.
14. Sociedade Brasileira de Doenças Cerebrovasculares – Grupo de Trabalho de Patologia Vascular Cerebral da ABN. Escala de Rankin de incapacidade modificada e índice de Barthel modificado. *Newsletter* 1999;6:7.
15. Green J, Foster A, Young J. A Test-retest reliability study of the Barthel index, the Rivermed mobility index, the Nottingham extended activities of daily living scale and the Frenchay activities index in stroke patients. *Disabil Rehabil* 2001;23:670-676.
16. Brott TG, Adams HP, Olinger CP, et al. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke* 1989;20:864-870.
17. Streiner DL, Norman GR. *Health measurements scales: a practical guide to their development and use*. Reliability, 2nd ed. London: Oxford University Press, 1995:104 -127.
18. Fleiss JL. *Statistical methods for rates and proportions*. 2nd ed. London: John Wiley & Sons, 1981:218.
19. Wilson JT, Hareendran A, Grant M, et al. Improving the assessment of outcome in stroke. *Stroke* 2002;33:2243-2246.
20. Swieten JCV, Koudstaal PJ, Visser MC, Schouten HJA, Gijn JV. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke* 1988;19:604-607.
21. Berger K, Weltermann B, Kolominsky PR, et al. The reliability of stroke scales. The German version of NIHSS, ESS and Rankin scale (Abstr.). *Fortschr Neurol Psychiatr* 1999;67:81-93.
22. Wolfe CDA, Taub NA, Woodrow BA, Burney PGJ. Assessment of scales of disability and handicap for stroke patients. *Stroke* 1991;22:1242-1244.
23. Lowen SC, Anderson BA. Reliability of the modified motor assessment scale and the Barthel index. *Phys Ther* 1988;68:1077-1081.
24. Haan DR, Limburg M, Schuling J. Clinimetric evaluation of the Barthel index, a measure of limitations in daily activities (Abstr.). *Ned Tijdsch Geneesk* 1993;137:917-921.
25. Hsueh IP, Lee MM, Hsieh CL. Psychometric characteristics of the Barthel index of daily living index in stroke patients (Abstr.). *J Formos Med Assoc* 2001;100: 526-532.
26. Wyller TB, Sveen U, Bautz-Holter E. The Barthel ADL index one year after stroke: comparison between relatives and occupational therapist's scores. *Age and Ageing* 1995;24:398-401.
27. Dewey HM, Donnan GA, Freeman EJ, et al. Interrater reliability of the National Institute of Health Stroke Scale: rating by neurologists and nurses in a community-based stroke incidence study. *Cerebrovasc Dis* 1999;9:323-327.
28. Albanese MA, Clarke WR, Adams HP, Woolson RF, et al. Ensuring reliability of outcome measures in multicenter clinical trials of treatments for acute ischemic stroke. *Stroke* 1994;25:1746-1751.
29. Goldstein LB, Bertels C, Davis JN. Interrater reliability of the NIH stroke scale. *Arch Neurol* 1989;46:660-662.
30. Lyden P, Brott T, Tilley B, et al. Improved reliability of the NIH stroke scale using video training. *Stroke* 1994;25:2220-2226.
31. Shinar D, Gross CR, Mohr JP, et al. Interobserver variability in the assessment of neurologic history and examination in the stroke data bank. *Arch Neurol* 1985;42:557-565.
32. Schmülling S, Grönd M, Rudolf J, Kiencke P. Training as a prerequisite for reliable use of NIH stroke scale. *Stroke* 1998;29:1258-1259.
33. Bushnell CD, Johnston DC, Goldstein LB. Retrospective assessment of initial stroke severity. *Stroke* 2001;32:656-660.
34. Meyer BC, Hemmen TM, Jackson CM, Lyden PD. Modified national institute of stroke scale for use in stroke clinical trials. *Stroke* 2002;33: 1261-1266.