

Revista de Saúde Pública

JOURNAL OF PUBLIC HEALTH

Confiabilidade das aferições de estudo sobre violência familiar e desnutrição severa na infância*

Measurement reliability in a study on family violence and severe acute malnutrition

Maria Helena Hasselmann, Claudia S. Lopes e Michael E. Reichenheim

Departamento de Nutrição Social do Instituto de Nutrição da Universidade do Estado do Rio de Janeiro (UERJ). Rio de Janeiro, RJ - Brasil (M.H.H.), Departamento de Epidemiologia do Instituto de Medicina Social da UERJ. Rio de Janeiro, RJ - Brasil (C.S.L., M.E.R.)

HASSEKMANN Maria Helena, Claudia S. Lopes e Michael E. Reichenheim *Confiabilidade das aferições de estudo sobre violência familiar e desnutrição severa na infância** Rev. Saúde Pública, 32 (5): 437-46, 1998

Confiabilidade das aferições de estudo sobre violência familiar e desnutrição severa na infância*

Measurement reliability in a study on family violence and severe acute malnutrition

Maria Helena Hasselmann, Cláudia S. Lopes e Michael E. Reichenheim

Departamento de Nutrição Social do Instituto de Nutrição da Universidade do Estado do Rio de Janeiro (UERJ). Rio de Janeiro, RJ - Brasil (M.H.H.), Departamento de Epidemiologia do Instituto de Medicina Social da UERJ. Rio de Janeiro, RJ - Brasil (C.S.L., M.E.R.)

Resumo

Introdução

Como parte do programa de investigação sobre violência familiar e desnutrição severa na infância, especificamente um estudo do tipo caso-controle foi avaliada a confiabilidade dos instrumentos utilizados no processo de obtenção de dados. Estudou-se a confiabilidade de quatro componentes do instrumento principal: (a) as Escalas sobre Táticas de Conflitos (Conflicts Tactics Scales - CTS) que medem o grau de conflito/violência familiar; (b) o instrumento CAGE (Cut-down; Annoyed; Guilty & Eye-opener) usado para indicar suspeição de alcoolismo; (c) o NSDUQ (Non-student Drugs Use Questionnaire) que visa aferir o uso de drogas ilícitas; e (d) a medida antropométrica de comprimento.

Método

Para os três primeiros componentes citados foram avaliadas a estabilidade (confiabilidade intra-observador ou teste-reteste) e a equivalência (confiabilidade inter-observador), usando-se os 50 primeiros sujeitos captados no estudo caso-controle de fundo. Para a análise, usou-se o índice Kappa (k) com ajustamento (pseudo-Bayes) para lidar com problemas de estimabilidade. Em relação ao componente "d", foi estudada somente a equivalência (n=73), usando-se o Coeficiente de Correlação Intraclasse (Intra-class Correlation Coefficient - ICC) como estimador.

Resultados

Todos os componentes mostraram estabilidade e equivalência aceitáveis. Quanto à estabilidade das CTS, CAGE e NSDUQ, as estimações de k foram em torno de 0,70, 0,78 e 0,85, respectivamente. Em relação à equivalência, encontrou-se os valores de 1,0 para as CTS e NSDUQ e 0,75 para CAGE. A equivalência estimada através do ICC para comprimento foi de 0,99. Algumas situações desviantes foram observadas. Os resultados apontam para uma adequada padronização dos observadores e refletem a boa qualidade do processo de aferição referente ao estudo de fundo, encorajando a equipe de pesquisa a prosseguir com maior segurança.

Violência doméstica. Transtornos nutricionais. Estudos de casos e controles. Reprodutibilidade de resultados.

*Parcialmente subvencionado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico/CNPq (Processo nº 200 122-82/7).

Correspondência para/Correspondence to: Michael E. Reichenheim - Rua São Francisco Xavier, 524 - 20599-090 Rio de Janeiro, RJ - Brasil. E-mail: michael@ims.uerj.br

Recebido em 21.10.1997. Reapresentado em 15.3.1998. Aprovado em 6.4.1998.

Abstract**Introduction**

This study is nested within a research program related to family violence and severe childhood malnutrition. Its aim is to evaluate the reliability of the data collection process in a case-control study. Four components of the main instrument are addressed: (a) CTS (Conflict Tactics Scales) used to measure violence at the family level; (b) CAGE (Cut-down; Annoyed; Guilty & Eye-opener) questionnaire used to gauge suspicion of drinking problems; NSDUQ (Non-student Drugs Use Questionnaire) used to indicate illicit drug consumption; and (d) height/length measurements.

Method

Stability (intra-observer or test-retest reliability) and equivalence (inter-observer reliability) were evaluated for the cited components (a), (b) and (c). Information was replicated among the first 50 subjects selected for the underlying case-control study. The Kappa index (k) was used in the analysis. A pseudo-Bayes adjustment was carried out in order to handle estimation problems. Regarding (d), only equivalence was evaluated (n=73), using the Intraclass Correlation Coefficient as the estimator.

Results

By and large, all components showed acceptable stability and equivalence. Regarding stability, the estimates of k were around 0.70, 0.78 and 0.85, for CTS, CAGE e NSDUQ, respectively. With respect to equivalence, k was 1.0 for CTS and NSDUQ and 0.75 for CAGE. Equivalence for height/length estimated through the ICC was 0.99. Nevertheless, some deviant situations were detected and are further discussed. The results point to an adequate standardization of observers and reflect the good quality of the data collection procedure concerning the main study, encouraging the research team to press forward with greater assurance.

**Domestic violence. Nutrition disorders. Case-control studies.
Reproducibility of results.**

INTRODUÇÃO

O presente estudo reflete a preocupação de avaliar a qualidade das mensurações em investigações epidemiológicas. Avaliações formais da qualidade das aferições, embora essenciais, raramente são conduzidas de forma explícita. Mesmo que um instrumento já tenha sido validado previamente e esteja bem estabelecido no âmbito de um programa de investigação, há a necessidade de um escrutínio do processo de aferição toda vez que este é utilizado num novo contexto. Ao contrário da validade que, uma vez estudada, passa a ser considerada uma propriedade imanente do instrumento, a confiabilidade diz respeito ao processo de aferição específico de um estudo. Pelo seu caráter particular, a confiabilidade necessita ser investigada continuamente e seus resultados são intransferíveis^{27,35}.

Os conceitos de estabilidade e equivalência são importantes ao se abordar a qualidade de informações em investigações epidemiológicas. Ambos estão atrelados à questão da reproducibilidade das aferições

e, conseqüentemente, à confiabilidade da informação. A estabilidade é a capacidade do instrumento produzir os mesmos resultados em diferentes momentos^{2,11}. A equivalência é a capacidade de um mesmo instrumento medir uma certa característica quando aplicado concorrentemente por diferentes observadores^{2,11,36}.

O presente estudo insere-se no bojo de um programa de investigação sobre violência familiar e desnutrição aguda severa²⁶. Um dos projetos específicos do programa concerne um estudo de caso-controlado que postula a violência no nível da família como fator de risco no processo de desnutrição severa aguda na criança²⁵. Além dos eventos desnutrição e violência, o modelo teórico subjacente contempla, entre outros corolários, o alcoolismo e uso de drogas. Por se tratar de eventos de difícil apreensão, decidiu-se aprofundar estes 4 componentes do instrumento de aferição do estudo de fundo, estudando-se a confiabilidade do processo de mensuração sob a perspectiva da estabilidade e equivalência. A seguir, são apresentados os componentes em questão e os procedimentos relativos à replicação.

MÉTODO

Componentes Replicados

As Escalas sobre Táticas de Conflitos (Conflict Tactics Scales - CTS) foram desenvolvidas por Straus³¹ objetivando mensurar as estratégias utilizadas pelos membros da família para resolver possíveis conflitos e, indiretamente, captar um estado de violência no domicílio. Com o intuito de “mapear” os níveis de conflito, o instrumento identifica quatro diferentes táticas: (a) de argumentação que consiste no uso de discussão racional; (b) de agressão verbal quando um indivíduo apela para insultos e ameaças com a intenção de, simbolicamente, machucar e agredir o outro; (c) de violência física não abusiva quando o uso de força física é explícita mas de intensidade moderada, possivelmente sem chegar a danos corporais severos; e (d) de violência física abusiva na forma de espancamentos ou atentados podendo levar a danos corporais severos. As CTS têm sido estudadas desde a sua concepção. Vários estudos apontam para uma baixa taxa de recusa³², indicam uma boa confiabilidade, principalmente para as escalas de violência física^{3,31}; e mostram uma validade de tipo concorrente^{17,32,34}, de constructo^{3,32,33} e de conteúdo³¹⁻³³.

O instrumento CAGE (Cut-down; Annoyed; Guilty & Eye-opener) é usado para indicar suspeição de alcoolismo. O CAGE é o mais simples dos instrumentos que se baseiam em diagnósticos dicotômicos (presença/ausência de alcoolismo). Seu desenvolvimento e uso clínico foram apresentados e descritos por Ewing¹² na Conferência Internacional sobre Alcoolismo, em 1970. O instrumento consiste de quatro questões inseridas de forma intermitente no instrumento de coleta. Não se encontrou evidências na literatura sobre a confiabilidade do CAGE em investigações realizadas no Brasil. Quanto à validade, no entanto, existem estudos realizados no País e no exterior. Mayfield e col.²³ mostraram que o instrumento possui validade concorrente ($\phi=0,89$). Em um estudo realizado em São Paulo, Masur e Monteiro²² mostraram uma sensibilidade e especificidade de 0,88 e 0,83, respectivamente. Ambos estudos definiam um CAGE positivo quando pelo menos dois itens eram positivos.

Desenvolvido por um grupo de especialistas e testados em várias localidades, o Questionário sobre o Uso de Drogas em Não Estudantes (Non-student Drugs Use Questionnaire - NSDUQ) é utilizado para captar o uso de drogas^{23,30}. Estudos teste-reteste aninhados numa investigação patrocinada pela OMS relatam uma boa confiabilidade³⁰. Na Índia foram replicadas 36 aferições e o escore agregando todos os tipos de drogas mostrou concordância quase completa ($r=0,97$). Resultado similar foi encontrado no Canadá ($r=0,95$) no qual foram reavaliadas 30 pessoas. No Paquistão, o resultado foi um pouco pior, ainda que aceitável ($r=0,78$).

Dentre várias medidas antropométricas utilizadas na avaliação do estado nutricional de indivíduos e populações, o peso e a altura são referidas como aquelas mais sensíveis

e específicas para apreender o processo de crescimento e desenvolvimento^{24,40}. As vantagens da utilização de medidas antropométricas na avaliação do estado nutricional decorrem do fato dos procedimentos requeridos no processo serem de simples execução, não invasivos e de baixo custo¹⁴. Ademais, os procedimentos são precisos e acurados, desde que sejam usadas técnicas padronizadas e os antropometristas adequadamente treinados^{14,37}. No sentido de avaliar a qualidade do treinamento, bem como do processo de aferição, ênfase deve ser dada à medição da altura/comprimento por se tratar do procedimento que oferece maiores dificuldades operacionais, especialmente em infantes e pré-escolares^{1,37}. Por isso, decidiu-se replicar preferencialmente a medida de altura/comprimento, uma vez que o estudo caso-controle de fundo contempla o uso do índice antropométrico peso-para-altura para indicar desnutrição severa aguda.

Replicação dos Componentes CTS, CAGE e NSDUQ

A estabilidade (confiabilidade intra-observadores) dos componentes CTS, CAGE e NSDUQ foi estudada através da abordagem teste-reteste conduzida por 2 entrevistadores, especialmente treinados, e um dos investigadores principais (MHH). Realizada uma primeira aferição, as respondentes (mães) eram aferidas novamente pelo mesmo observador após um intervalo de 3 dias a 2 semanas. Para a avaliação da equivalência (confiabilidade inter-observadores), MHH acompanhava a primeira entrevista sem dela participar ativamente, registrando em um questionário separado os itens respondidos ao entrevistador.

As primeiras 50 entrevistas do estudo caso-controle de fundo foram replicadas. A divergência dos tamanhos amostrais reportados nos resultados é resultante dos critérios adotados para as respectivas aplicações como, por exemplo, no caso de mães separadas do companheiro há mais de um ano às quais não se aplicou a CTS sobre o relacionamento pai-mãe/mãe-pai e os componentes CAGE e NSDUQ referentes ao companheiro.

Utilizou-se o Kappa (k) como estimador de confiabilidade pela sua adequação para o tipo de dados categóricos encontrados na CTS, CAGE e NSDUQ^{4,8,13}. Detalhes dos procedimentos estatísticos relativos às estimativas encontram-se em anexo. Para efeito de interpretação usou-se a proposta de Coutinho¹⁰, uma adaptação dos critérios sugeridos por Landis e Koch²⁰. Coutinho¹⁰ propõe cinco categorias para as estimativas de confiabilidade: (a) muito ruim (<0); (b) ruim (0 a 0,40); (c) regular (0,41 a 0,60); (d) boa (0,61 a 0,74) e (e) excelente (>0,75).

As CTS foram analisadas segundo o tipo de relacionamento familiar e escala tática de conflito. Os tipos de relacionamentos estudados foram (a) pai-criança; (b) mãe-criança; (c) pai-mãe e (d) mãe-pai^{32,34}. Para efeitos de concisão, optou-se por não detalhar as escalas de argumentação e agressão verbal e apresentar somente as escalas de violência física abusiva e não abusiva, além da

violência doméstica total. Quanto à classificação da última, utilizou-se a proposta de Straus³⁴. Uma família foi considerada violenta quando apresentava pelo menos uma resposta positiva nos itens relativos à agressão física^{33,39}. Tentando refletir um gradiente de gravidade, também se usou uma variável resumo em 3 níveis. Considerou-se violência grave se esta era abusiva; moderada se não abusiva; e não-violência se a tática de resolução de conflito restringia-se apenas a argumentação e agressão verbal.

Conforme recomenda a literatura, o critério de suspeição de alcoolismo através do CAGE é o de duas ou mais respostas positivas^{7,12,19,22,23}.

Segundo a Organização Mundial de Saúde, o instrumento NSDUQ deve conter perguntas sobre as drogas mais utilizadas na população a ser investigada³⁰. No presente estudo escolheu-se avaliar somente o uso de tranqüilizantes com ou sem recomendações médicas, uso de maconha, cocaína, cola de sapateiro e um item sobre outros tipos de drogas eventualmente consumidas. A

classificação dos usuários foi feita com base nos critérios propostos pela OMS, a saber, (a) não usuários: aqueles que nunca usaram qualquer tipo de drogas; (b) usuários leves: aqueles que, apesar de já terem utilizado drogas alguma vez, não as usaram nem semanalmente nem diariamente no mês anterior à entrevista; (c) usuários moderados: aqueles que usaram drogas semanalmente mas não diariamente no mês anterior à entrevista; e (d) usuários pesados: aqueles que usaram uma ou mais drogas diariamente no mês anterior à entrevista.

Replicação Antropométrica

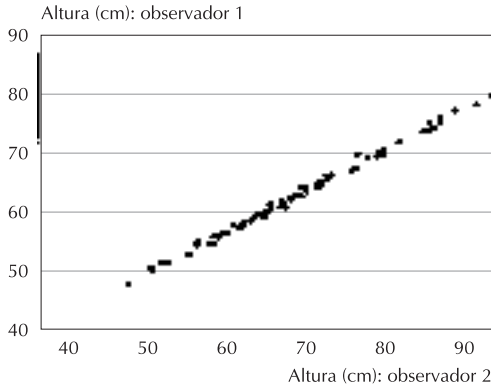
Após uma fase inicial de padronização dos antropometristas (realizada segundo a proposta de Habicht¹⁵), as primeiras 73 crianças captadas no estudo caso-controle de fundo foram aferidas por MHH, concorrentemente aos dois outros observadores. A replicação da altura/comprimento

Tabela 1 - Confiabilidade intra-observador das CTS, segundo tática de conflito e tipo de relacionamento familiar.

Tática de conflito	Tipo de relacionamento*			κ (IC 95%)	Curva de distribuição de κ (densidade)**
		++	+-		
		-+	--		
Violência não abusiva	PC	11 2	0 6	0,776 (0,338 - 1)	
	MC	18 0	2 4	0,750 (0,363 - 1,0)	
	PM	15 1	1 22	0,894 (0,580 - 1,0)	
	MP	10 2	4 23	0,655 (0,343 - 0,967)	
Violência abusiva	PC	8 1	1 9	0,788 (0,339 - 1,0)	
	MC	13 3	2 6	0,545 (0,147 - 0,944)	
	PM	2 1	5 31	0,328 (0,047 - 0,608)	
	MP	6 0	3 30	0,754 (0,450 - 1,0)	
Violência total - I		29 1	2 12	0,840 (0,545 - 1,0)	
Violência total - II	Legenda: G = Viol. Grave M = Viol. Mod. Ñ = Não Viol.	G M Ñ	G M Ñ 19 4 2 2 4 0 0 1 12	0,730 (0,478 - 0,983)	

* Relacionamento: PC=pai-criança; MC= mãe-criança; PM=pai-mãe; MP=mãe-pai (detalhes no texto).

** Linhas verticais representam a média e os limites de credibilidade de 95% (detalhes no Anexo).



ocorreu de forma mascarada e a magnitude estimada pelo Coeficiente de Correlação Intraclasse (Intraclass Correlation Coefficient - ICC)^{4,5}.

RESULTADOS

As Tabelas 1, 2 e 3 sumarizam, respectivamente, os resultados da confiabilidade intra-observador (teste-reteste) dos componentes *CTS*, *CAGE* e *NSDUQ*. Com relação às *CTS*, exceto a relação pai-mãe na escala de violência abusiva em que o *k* foi

Fonte de variação	Soma dos quadrados	GL	Quadrados médios	F	Prob.	CCI (ICC)
Entre-Observadores	1.149.275,74	72	15.962,16			0,99
Intra-Observadores	1.127,50	73	15,44			
Entre-Medidas	56,72	1	6,72	3,81	0,0547	
Resíduo	1.070,78	72	14,87			
Total	1.150.403,24	145	7.933,81			
Grande Média	641,20					

Figura - Confiabilidade inter-observador da medida antropométrica de comprimento: diagrama de dispersão e a tabela de análise de variância com o respectivo coeficiente de correlação intraclasse (CCI).

Tabela 2 - Confiabilidade intra-observador do *CAGE*, segundo informação da respondente sobre si mesma e sobre o companheiro.

Tabela de contingência	κ (IC 95%)	Curva de distribuição de κ (densidade)*									
<table border="1"> <tr><td></td><td>+</td><td>-</td></tr> <tr><td>+</td><td>2</td><td>0</td></tr> <tr><td>-</td><td>0</td><td>48</td></tr> </table>		+	-	+	2	0	-	0	48	1,0 (0,723 - 1,0)	
	+	-									
+	2	0									
-	0	48									
<table border="1"> <tr><td></td><td>+</td><td>-</td></tr> <tr><td>+</td><td>6</td><td>5</td></tr> <tr><td>-</td><td>1</td><td>26</td></tr> </table>		+	-	+	6	5	-	1	26	0,570 (0,265 - 0,874)	
	+	-									
+	6	5									
-	1	26									

* Linhas verticais representam a média e os limites de credibilidade de 95% (detalhes no Anexo).

Tabela 3 - Confiabilidade intra-observador do *NSDUQ*, segundo informação da respondente sobre si mesma e sobre o companheiro.

Categorias (legenda)	Tabela de contingência	κ (IC 95%)	Curva de distribuição de κ (densidade)*																									
Respondente	<table border="1"><tr><td></td><td>P</td><td>M</td><td>L</td><td>Ñ</td></tr><tr><td>P</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>M</td><td>0</td><td>3</td><td>0</td><td>0</td></tr><tr><td>L</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>Ñ</td><td>0</td><td>1</td><td>0</td><td>46</td></tr></table>		P	M	L	Ñ	P	0	0	0	0	M	0	3	0	0	L	0	0	0	0	Ñ	0	1	0	46	0,847 (0,573 - 1,0)	
		P	M	L	Ñ																							
	P	0	0	0	0																							
	M	0	3	0	0																							
L	0	0	0	0																								
Ñ	0	1	0	46																								
Usuário																												
P = Pesado																												
M = Moderado																												
Companheiro	<table border="1"><tr><td></td><td>P</td><td>M</td><td>L</td><td>Ñ</td></tr><tr><td>P</td><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>M</td><td>0</td><td>2</td><td>0</td><td>1</td></tr><tr><td>L</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>Ñ</td><td>0</td><td>0</td><td>0</td><td>34</td></tr></table>		P	M	L	Ñ	P	1	0	0	0	M	0	2	0	1	L	0	0	0	0	Ñ	0	0	0	34	0,864 (0,577 - 1,0)	
		P	M	L	Ñ																							
	P	1	0	0	0																							
	M	0	2	0	1																							
L	0	0	0	0																								
Ñ	0	0	0	34																								
L = Leve																												
Ñ = Não																												

* Linhas verticais representam a média e os limites de credibilidade de 95% (detalhes no Anexo).

baixo (0,33), encontrou-se um k em torno de 0,75. As duas estimativas de violência total revelam k altos de 0,84 e 0,73. Segundo a classificação de Coutinho¹⁰, a estabilidade do *CAGE* quando a respondente se refere ao companheiro é apenas moderada mas passa a ser excelente quando esta se refere a si mesma. Também para o componente *NSDUQ* a estabilidade do pro-

cesso de aferição é classificável como excelente.

A equivalência dos processos de aferição dos entrevistadores em relação componentes *CTS*, *CAGE* e *NSDUQ* encontram-se nas Tabelas 4, 5 e 6. O quadro geral mostra estimações de ponto máximas na grande maioria dos contrastes. As exceções (*CTS-MC* e *CAGE-respondente*) podem ser atribuídas às flutuações amostrais, o que é bem

Tabela 4 - Confiabilidade intra-observador das *CTS*, segundo tática de conflito e tipo de relacionamento familiar.

Tática de conflito	Tipo de relacionamento*	++ +- -+ --	κ (IC 95%)	Curva de distribuição de κ (densidade)**																
Violência não abusiva,	MC	10 0 0 5	1,0 (0,494 - 1,0)																	
	PM	4 0 0 20	1,0 (0,60 - 1,0)																	
	MP	5 0 0 19	1,0 (0,60 - 1,0)																	
Violência abusiva	MC	8 0 2 5	0,727 (0,240 - 1,0)																	
	PM	3 0 0 21	1,0 (0,60 - 1,0)																	
	MP	3 0 0 21	1,0 (0,60 - 1,0)																	
Violência total - I		15 0 0 15	1,0 (0,642 - 1,0)																	
Violência total - II	Legenda: G = Viol. Grave M = Viol. Mod. Ñ = Não Viol.	<table border="1"> <tr><td></td><td>G</td><td>M</td><td>Ñ</td></tr> <tr><td>G</td><td>15</td><td>0</td><td>0</td></tr> <tr><td>M</td><td>0</td><td>5</td><td>0</td></tr> <tr><td>Ñ</td><td>0</td><td>0</td><td>10</td></tr> </table>		G	M	Ñ	G	15	0	0	M	0	5	0	Ñ	0	0	10	1,0 (0,707 - 1,0)	
	G	M	Ñ																	
G	15	0	0																	
M	0	5	0																	
Ñ	0	0	10																	

* Relacionamento: PC=pai-criança; MC=mãe-criança; PM=pai-mãe; MP=mãe-pai (detalhes no texto).

** Linhas verticais representam a média e os limites de credibilidade de 95% (detalhes no Anexo).

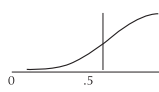
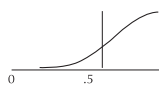
Tabela 5 - Confiabilidade inter-observador do *CAGE*, segundo informação da respondente sobre si mesma e sobre o companheiro.

Tabela de contingência	κ (IC 95%)	Curva de distribuição de κ (densidade)*
Respondente	0,5** (0,12 - 0,89)	
Companheiro	1,0 (0,60 - 1,0)	

* Linhas verticais representam a média e os limites de credibilidade de 95% (detalhes no Anexo).

** Estimação via ajuste pseudo-Bayes (detalhes no Anexo).

Tabela 6 - Confiabilidade intra-observador do NSDUQ, segundo informação da respondente sobre si mesma e sobre o companheiro.

Categorias	Tabela de contingência				κ (IC 95%)	Curva de distribuição de κ (densidade)*	
	P	M	L	Ñ			
Respondente	P	0	0	0	1,0 (0,642 - 1,0)		
	M	0	0	0			
	L	0	0	1			0
	Ñ	0	0	0			29
	Usuário						
Companheiro	P	1	0	0	1,0 (0,636 - 1,0)		
	M	0	1	0			0
	L	0	0	0			0
	Ñ	0	0	0			22
	Usuário						

* Linhas verticais representam a média e os limites de credibilidade de 95% (detalhes no Anexo).

captado pelos largos intervalos de credibilidade.

A Figura mostra os detalhes da análise relativa à replicação da aferição de comprimento das crianças. O coeficiente de correlação intraclassa (*ICC*) foi de 0,99.

DISCUSSÃO

De uma maneira geral, o estudo sugere uma adequação do processo de aferição no âmbito da estabilidade e equivalência das informações colhidas, mesmo levando em conta o tamanho amostral relativamente pequeno e, ocasionalmente, a acentuada assimetria entre eventos positivos e negativos. Estas características e a decorrente indesejável rarefação de dados não são raras em estudos cuja função precípua é avaliar a qualidade do processo de aferição de uma investigação epidemiológica maior. Como apontado na metodologia em anexo, usou-se o estimador pseudo-Bayes para o ajuste do *k* no intuito de lidar com situações de inestimabilidade. Todavia, o procedimento pode revelar também algo sobre o efeito da rarefação, sua repercussão sobre as inferências e, principalmente, sobre os processos de decisão operacionais e gerenciais concernentes ao estudo principal. Ao se imputar probabilidades *a priori* informativas na faixa de 0,02 a 0,05 para as caselas das tabelas de contingência contendo zeros - o que pressupõe distribuições subjacentes onde existe alguma discordância, ainda que esta não tenha se manifestado na amostra única em mãos - o decréscimo dos estimadores ajustados do *k* não passa de 10%. Diante do espectro dos valores encontrados, esta diminuição raramente altera o julgamento qualitativo (segundo a escala de Coutinho), o que é assegurador.

Com poucas exceções (e.g., relacionamento pai-mãe da escala de violência abusiva na Tabela 1), os resultados mostram uma confiabilidade satisfatória das *CTS* tanto no âmbito da estabilidade quanto da equivalência do processo de aferição. É mister apontar que esta qualificação e outras adiante, tomam como base inferencial os estimadores de ponto. No entanto, não pode ficar sem menção o fato de que praticamente todos os limites inferiores de credibilidade encontrados estão abaixo de 0,5, o que demanda, portanto, uma certa cautela à utilização do estimador de ponto na apreciação dos resultados.

Tanto a estimativa de violência total proveniente da variável dicotomizada quanto a da forma tricótoma revelam uma boa estabilidade (0,84 e 0,73, respectivamente) ou, no mínimo, regular, ao se tomar os valores do limites inferiores dos intervalos de credibilidade (0,54 e 0,48, respectivamente). Esses achados estão de acordo com Straus³² que acena para a pertinência do instrumento quando o objetivo é separar as famílias violentas das não violentas. Em contrapartida, deve ser mencionado que ao se analisar o *k* referente às escalas de argumentação e agressão verbal (informação não detalhada nos resultados), encontram-se valores em torno de 33%, o que também vem ao encontro aos achados do autor³². No entanto, no âmbito do estudo caso-controle no qual esta avaliação se aninha, são as escalas de melhor confiabilidade (violência total I e II) que realmente importam, uma vez que estas definem a exposição de interesse central (estado de violência intra-familiar)¹⁶.

Distinto do processo de aferição sobre consumo de drogas, destaca-se o fato da estabilidade do componente *CAGE* ser marcadamente menor quando

a respondente relata os hábitos étlicos do companheiro do que os seus próprios. Em relação à equivalência o padrão observado é inverso, ainda que se conceda que a evidência peca pela imprecisão, sugerindo que a diferença tenha decorrido de um problema amostral. Este panorama parece refletir a sensibilidade da questão do alcoolismo, o que Masur e col.²¹ discutem a partir da ótica de uma “invasão de privacidade” no que tange à admissão do problema. É a impressão dos autores que, ao contrário do esperado, é no primeiro contato que a respondente se dispõe a “denunciar” o companheiro, passando a negar o fato no segundo encontro.

Em relação ao *NSDUQ* é preciso ser mais cauteloso ao se discutir a sensibilidade da admissão do consumo. A detectada concordância intra- e interobservador pode tanto representar uma disposição à admissão quanto uma relutância sistemática. De toda forma, existe uma consistência. Isto garante uma confiabilidade da aferição, ainda que nada se possa afirmar sobre a validade da informação. Como colocado na introdução, esta é assumida *prima facie*. Estas questões merecem aprofundamento, pois, sendo de fácil aplicação, os dois instrumentos tendem a ser usados freqüentemente.

Quanto à confiabilidade da mensuração de comprimento, os achados são bastante consistentes com outros estudos^{1,37,38}.

Ainda que um aprofundamento sobre a questão da validade de instrumentos de aferição mereça toda a atenção, é preciso reforçar que os 4 componentes utilizados na presente investigação são assumidos como válidos *a priori* e que o artigo explicitamente enfoca o processo de aferição em si. É necessário, no entanto, que se distinga entre a validade própria de um instrumento e a da informação gerada. A última depende também da confiabilidade, uma vez que um instrumento aceito como válido (*a priori* ou estudado anteriormente) pode ter, circunstancialmente, uma precária estabilidade e/ou equivalência. Esta deficiência implicará em má qualidade do processo de captação da informação, podendo comprometer, conseqüentemente, a validade da informação²⁷.

Alguns aspectos metodológicos gerais também merecem comentários quando se examina a confiabilidade de um processo de aferição, principalmente a de tipo interobservador. O primeiro

diz respeito aos entrevistadores, que, estando conscientes do processo de avaliação de confiabilidade em curso, tendem a proceder com maior rigor do que em circunstâncias habituais. Isto pode levar a uma superestimação da confiabilidade e, aceitando-se a possibilidade do problema ter ocorrido no presente estudo, deve ser considerado ao se apreciar os resultados¹⁸.

O segundo aspecto é a tendência do observador em alterar a sua abordagem com o tempo. A aplicação continuada de um instrumento pode fazer com que o observador paulatinamente aprimore sua maneira de perguntar e de se relacionar com os respondentes. Estas mudanças tendem a ocorrer principalmente no início do período de contato com o instrumento. Assim, desde que se consiga precocemente evitar problemas com os observadores, pode se esperar que a qualidade e consistência da aferição melhorem progressivamente. Nesta perspectiva, implementou-se um ativo treinamento e um pré-teste precedendo o estudo, o que parece ter ajudado.

Os resultados aceitáveis em relação a estabilidade e equivalência indicaram uma adequada padronização dos observadores, permitindo prosseguir com a mesma equipe em segurança. Isto é particularmente relevante em estudos nos quais o processo de aferição é delegado a muitos entrevistadores. Uma outra questão importante é reconhecer que, mesmo tendo-se constatado uma boa confiabilidade em alguma fase da coleta de dados, é sempre possível ocorrer uma queda de qualidade da aferição ao longo do processo. Para garantir uma boa qualidade dos dados é decisivo que jornadas de replicações sejam planejadas e implementadas periodicamente ou, eventualmente, à troca de membros da equipe de aferidores. No caso do extenso estudo caso-controle de fundo, esta foi a primeira; outras estão contempladas no curso da investigação principal.

AGRADECIMENTOS

Aos entrevistadores de campo e aos funcionários do Hospital Municipal Salles Netto, Hospital Municipal Jesus e IPPMG/UFRJ pela colaboração na coleta de dados.

REFERÊNCIAS

1. AHMED, M.L.; YUDKIN, P.L.; MACFARLANE, J.A.; MCPHERSON, K.; DUNGER, D.B. Are measurement of height made by health visitors sufficiently accurate for routine screening of growth? *Arch. Dis. Child.*, **65**:1345-8, 1990.
2. ALMEIDA FILHO, N. *Epidemiologia sem números: uma introdução crítica à ciência epidemiológica*. Rio de Janeiro, Campus, 1989.
3. ASSIS, S.G. *Crianças, violências e comportamentos: um estudo em grupos sociais distintos*. Rio de Janeiro, Escola Nacional de Saúde Pública/FIOCRUZ, 1992.
4. BARTKO, J.J. On various intraclass correlation reliability coefficients. *Psychol. Bull.*, **83**:762-5, 1976.
5. BARTKO, J.J. General methodology II. Measures of agreement: a single procedure. *Stat. Med.*, **13**:737-45, 1994.
6. BISHOP, Y.M.M.; FIENBERG, S.E.; HOLLAND, P.W. *Discrete multivariate analysis: theory and practice*. Cambridge, The MIT Press, 1975.
7. CAPRIGLIONE, M.J.; MONTEIRO, M.G.; MASUR, J. Aplicação do questionário CAGE para detecção da síndrome de dependência do álcool em 700 adultos na cidade de São Paulo. *Rev. Ass. Bras. Psiq.*, **7**:50-3, 1985.
8. COHEN, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, **20**:37-46, 1960.
9. COHEN, J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol. Bull.*, **70**:213-20, 1968.
10. COUTINHO, E. Confiabilidade do diagnóstico psiquiátrico em hospitais do Rio de Janeiro. Rio de Janeiro, 1987. [Dissertação de Mestrado - Escola Nacional de Saúde Pública, FIOCRUZ].
11. CRONBACH, L.J. Coefficient alpha and the internal structure of tests. *Psychometrika*, **16**:297-334, 1951.
12. EWING, J.A. Detecting alcoholism: the CAGE questionnaire. *JAMA.*, **252**:1905-7, 1984.
13. FLEISS, J.L. *Statistical methods for rates and proportions*. 2nd ed. New York, John Wiley & Sons, 1981.
14. GIBSON, R.S. *Principles of nutritional assessment*. New York, Oxford University Press, 1990.
15. HABICHT, J.P. Estandarização de métodos epidemiológicos quantitativos sobre el terreno. *Bol. Ofic. Sanit. Panam.*, **76**:375-84, 1974.
16. HASSELMANN, M.H. Violência familiar e desnutrição severa na infância: modelo teórico e estudo de confiabilidade dos instrumentos. Rio de Janeiro, 1996. [Dissertação de Mestrado - Instituto de Medicina Social, UERJ].
17. JOURILES, E.N. & O'LEARY, K.D. Interspousal reliability of reports of marital violence. *J. Consul. Clin. Psychol.*, **53**:419-21, 1985.
18. KAZDIN, A.E. Artifact, bias and complexity of assessment: validation of the ABCs of reliability. *J. Appl. Behav. Anal.*, **10**:141-50, 1977.
19. KING, M. At risk drinking among general practice attenders: validation of the CAGE questionnaire. *Psychol. Med.*, **16**:213-7, 1986.
20. LANDIS, J.R. & KOCH, G.G. The measurement of observer agreement for categorical data. *Biometrics*, **33**:159-74, 1977.
21. MASUR, J.; CAPRIGLIONE, M.J.; MONTEIRO, M.G.; JORGE, M.R. Detecção precoce do alcoolismo em clínica médica através do questionário CAGE. *J. Bras. Psiq.*, **34**:31-4, 1985.
22. MASUR, J. & MONTEIRO, M.G. Validation of the "CAGE" alcoholism screening test in a Brazilian psychiatric inpatient hospital setting. *Brazilian J. Med. Biol. Res.*, **16**:215-8, 1983.
23. MAYFIELD, D.; MCLEOD, G.; HALL, P. The CAGE questionnaire: Validation of a new alcoholism screening instrument. *Am. J. Psychiatry*, **131**:1121-3, 1974.
24. MONTEIRO, C.A. Critérios antropométricos no diagnóstico da desnutrição em programas de assistência à criança. *Rev. Saúde Pública*, **18**:56 - 63, 1974.
25. REICHENHEIM, M.E. & HASSELMANN, M.H. Violência doméstica: uma questão no processo de determinação da desnutrição na infância? Rio de Janeiro, 1994 [Projeto de Pesquisa, Instituto de Medicina Social e Instituto de Nutrição da UERJ].
26. REICHENHEIM, M.E. & HASSELMANN, M.H. Um novo perfil da criança desnutrida. A violência doméstica pode estar por trás de casos residuais de desnutrição. *Folha de São Paulo*, 18 de maio de 1997.
27. REICHENHEIM, M.E. & MORAES, C.L. Alguns pilares para a apreciação da validade de estudos epidemiológicos. *Rev. Bras. Epidemiol.*, **1**(2) 1998. [no prelo].
28. SALGADO-UGARTE, I.H.; SHIMITZU, M; TANIUCHI, T. Exploring the shape of univariate data using kernel density estimators. *Stata Tech. Bull.*, **16**:8-19, 1993.
29. SILVERMAN, B.W. *Density estimation for statistics and data analysis*. London, Chapman and Hall, 1986.
30. SMART, R.G.; ARIF, A.; HUGHES, P.; MEDINA MORA, M.E.; NAVARATNAM, V.; VARMA, V.K.; WADUD, K.A. *Drugs use among non-student youth*. Geneva, World Health Organization, 1981. (WHO Offset Publication, No. 60).

31. STRAUS, M.A. Measuring intra-familial conflict and violence: the conflict tactics (CT) scale. *J. Marriage Fam.*, (41):75 - 88, 1979.
32. STRAUS, M.A. *Physical violence in american families: risk factors and adaption to violence in 8145 families*. New Brunswick, Transaction Publisher, 1990.
33. STRAUS, M.A. *Beating the devil out of them: corporal punishment in american families*. New York, Cambridge University Press, 1994.
34. STRAUS, M.A.; GELLES, R.J.; STEINMETZ, S.K. *Behind closed doors: violence in the american family*. New York, Anchor Press, 1981.
35. STREINER, D.L. & NORMAN, G.R. *Health measurement scales: a practical guide to their developemnt and use*. Oxford, Oxford University Press, 1989.
36. VERAS, R.P. *País jovem de cabelos brancos: a saúde do idoso no Brasil*. Rio de Janeiro, Relume Dumará, 1994.
37. VOSS, L.D. & BAILEY, B.J.R. Equipping the community to maesure children's height: the reliability of portable instruments. *Arch. Dis. Child.*, **70**:469-71, 1994.
38. VOSS, L.D.; BAILEY, B.J.R.; CUMMING, K.; WILKIN, T.J.; BETTS, P.R. The reliability of height measurement (the Wessex Growth Study). *Arch. Dis. Child.*, **65**:1340-4, 1990.
39. WISSOW, L.S.; WILSON, M.E.H.; ROTER, D.; LARSON, S.; BERMAN, H.I. Family violence and the evaluation of behavioral concerns in a pediatric primary health clinic. *Med. Care*, **30 (Suppl)**:150-65, 1992.
40. WORLD HEALTH ORGANIZATION. *Physical status: the use and interpretation of anthropometry*. Geneva, 1995. (HO Technical Report Series, No. 854).

ANEXO

O estimador de ponto do Kappa (κ) $p_j = \sum_i p_{ij}$ é dado por:

$$\hat{\kappa} = (p_o - p_e) / (1 - p_e)$$

onde $p_o = \sum_{i=1}^k \sum_{j=1}^k w_{ij} p_{ij}$ e, $p_e = \sum_{i=1}^k \sum_{j=1}^k w_i p_j$ sendo

$$p_i = \sum_j p_{ij} \text{ e } p_j = \sum_i p_{ij}.$$

Os pesos para concordância são:

$$w_{ij} (i=, \dots, k; j=, \dots, k) = 1 - (i-j) / (c-1),^8$$

sendo i e j , respectivamente, os índices de linhas e colunas da tabela de contingência e c correspondendo ao número de categorias da variável aferida. Esta ponderação pretende refletir a importância relativa de cada discordância possível.⁹

Os intervalos de credibilidade reportados são dados por¹³ $\hat{\kappa} \pm 1,96\hat{\delta}_\kappa$, onde

$$\hat{\delta}_\kappa = [(1-p_e)\sqrt{n}]^{-1} \sqrt{(\sum_i p_i p_j [w_{ij} - (\bar{w}_i + \bar{w}_j)]^2) - p_e^2}$$

é o erro padrão e $\bar{w}_i = \sum_j p_j w_{ij}$ e $\bar{w}_j = \sum_i p_i w_{ij}$.

Aqui, assume-se uma distribuição gaussiana mas necessariamente truncada à direita para respeitar o limite superior dos valores próprios do estimador. Optou-se pelo uso de um ajuste baseado no estimador pseudo-Bayes que permite lidar adequadamente com os zeros não-estruturais

(amostrais) encontrados.⁶ Nesta perspectiva, confere-se uma massa (densidade) a qualquer casela de concordância, mesmo que a realização nos dados tenha se expressado por zero em algumas delas. Para obter o estimador pseudo-Bayes é necessário selecionar um conjunto de probabilidades *a priori* λ_{ij} e computar as estimativas (posteriores) para cada casela

$$m_{ij}^* = [n / (n + \hat{v})] (x_{ij} + \hat{v} \lambda_{ij}),$$

onde $\hat{v} = [n^2 - \sum_{ij} x_{ij}^2] / [\sum_{ij} (x_{ij} - n\lambda_{ij})^2]$ é um fator de ponderação. O novo κ e seus limites de credibilidade são, então, calculados a partir de $p_{ij}^* = m_{ij}^* / n$, substituindo p_{ij} . Uma vez que não existe qualquer conhecimento *a priori* para imputar plausíveis probabilidades a cada casela λ_{ij} , preferiu-se usar *a priores* guiados pelos próprios dados. Esta conduta deixa praticamente inalteradas as estimações obtidas com o κ tradicional (não ajustado) mas permite uma estimação em configurações onde este é inestimável (e.g., se todos os indivíduos aferidos são positivos ou negativos).

Nas curvas de distribuição aninhadas nas tabelas mostradas nos resultados, utilizou-se a função de Kernel de Epanechnikov²⁸ e uma largura de banda sugerida por Silverman²⁹ dada por $b = 1,06 \min(\sigma, R/1,34)n^{-1/5}$ onde σ é o desvio-padrão e $R = [Centil 75 - Centil 25]$ calculado a partir da distribuição normal.