

# Modelo hierarquizado: uma proposta de modelagem aplicada à investigação de fatores de risco para diarreia grave\*

## *Hierarchical model: a proposal for model to be applied in the investigation of risk factors for dehydrating diarrhea*

Sandra C Fuchs, Cesar G. Victora e Jandyra Fachel

*Departamento de Medicina Social da Faculdade de Medicina, Universidade Federal do Rio Grande do Sul. Porto Alegre, RS - Brasil (S.C.F.); Departamento de Medicina Social da Faculdade de Medicina da Universidade Federal de Pelotas. Pelotas, RS - Brasil (C.G.V.); Departamento de Estatística do Instituto de Matemática da Universidade Federal do Rio Grande do Sul. Porto Alegre, RS - Brasil (J.F.)*

### Resumo

Os estudos epidemiológicos de doenças que acometem as crianças geralmente envolvem grande número de variáveis. As associações entre potenciais fatores de risco e doença frequentemente são avaliadas através de modelagem estatística sem descrição das estratégias empregadas. No estudo realizado apresenta-se uma abordagem hierarquizada aplicada à avaliação de fatores de risco para diarreia grave. As variáveis foram hierarquicamente agrupadas em características socioeconômicas, ambientais, reprodutivas maternas, nutricionais e demográficas. Na análise univariada todas as variáveis associaram-se com diarreia grave. Em cada bloco selecionaram-se fatores de confusão através de um algoritmo, utilizando-se o processo retrógrado de seleção, através do módulo em passos, segundo um  $p=0,10$ . Os fatores de risco foram avaliados através de regressão logística após o ajuste para fatores de confusão de cada conjunto e para aqueles hierarquicamente superiores. As variáveis incluídas no modelo permitiram identificar corretamente uma proporção elevada de casos ( $\gamma=0,74$ ) e todos os blocos contribuíram significativamente para a modelagem.

**Diarreia infantil, epidemiologia. Fatores de risco. Modelos estatísticos.**

### Abstract

*In epidemiological investigations of infant diseases, data are usually collected on a large set of variables. The associations between presumptive risk factors and the outcome is commonly evaluated through statistical modeling, but the model-building strategies are seldom described. In the project, data collected in a case-control study of risk factors for dehydrating diarrhea in infants have*

\* Apresentado como Tese de Doutorado intitulada: "Fatores de risco para diarreia complicada por desidratação moderada a grave: um estudo de casos e controle" em 1993, junto ao Curso de Pós-Graduação em Medicina Clínica Médica da Universidade Federal do Rio Grande do Sul. Financiado pela Organização Mundial da Saúde, Projeto n° C6/181/304 (WHO registry file number) e 91023 EDP (project ID number).

**Correspondência para / Correspondence to:** Sandra Costa Fuchs - Departamento de Medicina Social da Faculdade de Medicina da Universidade Federal do Rio Grande do Sul. Rua Ramiro Barcelos, 2600, 4° andar, sala 415 - 90035-003 Porto Alegre, RS - Brasil. Fax (55 051 330).1380  
Recebido em 4.5.1995. Aprovado em 11.11.1995.

*been used to present a hierarchized approach to the assessment of risk factors. The variables were grouped into a hierarchy of categories, ranging from distal determinants to proximate ones. These categories included, in this order, socio-economic, environmental, reproductive maternal, nutritional and demographic sets. According to the univariate analyses all variables were associated with the outcome. As the purpose was to identify a parsimonious model to explain the data, in each set the confounders were selected through backward elimination, according to an alpha level of 0.10. The risk factors were evaluated through logistic regression after adjustment for confounders in the same set or in hierarchically superior sets. This approach allows researchers to quantify the contribution of each level of adjustment, to understand the model-building strategy as well as interpret the independent associations. The goodness-of-fit assessed at each set showed significant improvements in the model. The gamma coefficient of association was employed to quantify the proportion of cases and controls correctly identified by comparing the observed value with that predicted by the variables in the model. The final model resulted in a gamma of 0.74. The children's ages did not improve the prediction of cases and controls, but they have been kept in the model as they affect some exposures such as breastfeeding.*

**Infantile diarrhea, epidemiology. Risk factors. Statistical models.**

## INTRODUÇÃO

A saúde infantil, particularmente nos países em desenvolvimento, é determinada por grande número de características relacionadas a condições socio-econômicas, ambientais, nutricionais, de cuidados com a saúde, entre outras<sup>3,11</sup>. As doenças infecciosas, por exemplo, têm se associado com menor nível social e econômico, freqüentemente aferido através de renda, escolaridade, tipo de habitação, disponibilidade de água encanada e de esgoto cloacal. Essas condições não são responsáveis diretas pela ocorrência de doença, contudo, favorecem a proximidade de alguns determinantes<sup>3,11</sup>. Assim, por exemplo, uma mãe pobre e com baixa escolaridade iniciará sua vida reprodutiva mais precocemente e estará menos apta a se beneficiar da estrutura assistencial durante a gestação. O estado nutricional materno prévio, acentuado pelo ganho reduzido de peso durante a gestação, propicia um ambiente intra-uterino desfavorável ao crescimento fetal. O baixo peso de nascimento resultante é um fator de risco para a morbimortalidade durante toda a infância. A mãe, com ingesta alimentar inadequada e sem suporte familiar, tem menor chance de iniciar e manter o aleitamento. Número excessivo de pessoas em um domicílio insalubre, com água e esgoto inadequados, favorece maior contaminação do ambiente onde vive a criança. A presença de fatores biológicos e ambientais, intensificados pela ausência do efeito

protetor do leite materno, determinam maior frequência e gravidade da doença diarreica que impedem a recuperação nutricional posterior.

A investigação de fatores de risco para uma doença infecciosa envolve a coleta de número considerável de variáveis. Considerando-se a possibilidade de que uma ou mais variáveis de confusão possam explicar parcial ou totalmente o efeito detectado, torna-se indispensável o emprego de análise multivariada na identificação de fatores de risco independentes. Os critérios mais freqüentemente empregados para incluir variáveis na modelagem são: a) ajustar-se para todas as variáveis coletadas; b) ajustar-se para aquelas cujo controle represente uma diferença importante na medida de efeito; c) controlar-se para as variáveis que se associarem significativamente com o desfecho; d) selecionar-se para o modelo apenas a variável cuja medida de efeito ajustada for estatisticamente significativa; e) controlar-se para as variáveis a menos que a mudança na estimativa de exposição-efeito, produzida pelo controle da variável, seja equivalente a zero<sup>9</sup>.

Algumas críticas a esses processos de modelagem incluem a ineficiência de modelo ajustado para variáveis que não são fatores de confusão, a não consideração da variabilidade estatística ou a aplicação de um teste estatístico como medida do grau de confusão presente em uma associação. Uma vez que os investigadores freqüentemente não possuem conhecimento prévio sobre os fatores de confusão, apli-

cam algum tipo de estratégia aos dados coletados para decidir quais as variáveis que devem entrar na modelagem<sup>9</sup>.

A modelagem hierarquizada, aqui proposta, é uma alternativa aplicável a estudos epidemiológicos com um grande número de covariadas<sup>5, 12</sup>. A escolha de critérios para a seleção de variáveis de confusão ultrapassa o aspecto puramente estatístico, criticado em alguns processos anteriormente citados<sup>2, 4, 14</sup>. A hierarquização das variáveis independentes é estabelecida no marco conceitual e mantida durante a análise dos dados permitindo a seleção daquelas mais fortemente associadas com o desfecho de interesse. Assim, gera-se um modelo com um número de termos necessários para testarem-se as associações, mas não saturado pelo excesso de variáveis.

Além da seleção de variáveis de confusão, outra questão refere-se à verificação dos pressupostos do modelo. Uma vez que a validade das inferências depende de quão bem o modelo descreve os dados observados, torna-se necessário saber se a modelagem contém as variáveis que deveria conter e se estas entraram em sua forma correta<sup>7</sup>. Portanto, realizada a análise multivariada, deveria avaliar-se o ajuste da modelagem.

No presente estudo utilizou-se a modelagem hierarquizada para se analisar os fatores de risco para diarreia grave (cujos detalhes estão descritos na tese de Fuchs<sup>5</sup>), discriminou-se a contribuição dos blocos de variáveis para o ajuste do modelo, e avaliaram-se os aspectos estatísticos que poderiam influenciar a modelagem.

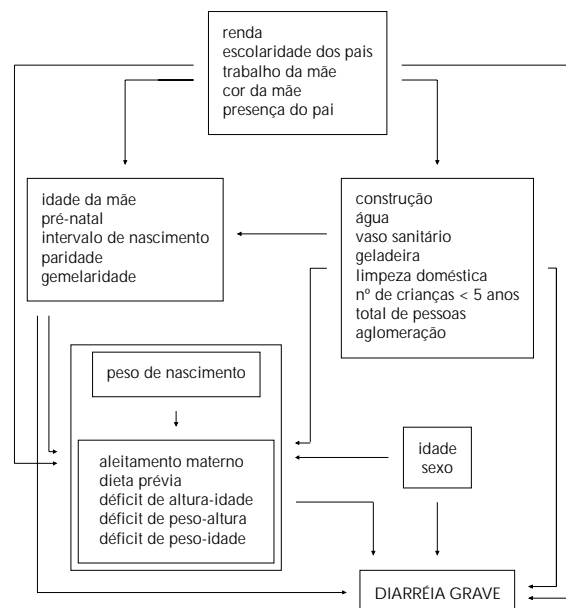
## MATERIAL E MÉTODO

Um estudo de casos e controles foi delineado para investigar os fatores de risco para diarreia complicada por desidratação moderada a grave (diarreia grave). Os casos constituíram-se de crianças, com idade entre 3 dias e 23 meses, com diarreia de duração inferior a oito dias, hospitalizadas com desidratação moderada a grave em dois hospitais pediátricos de Porto Alegre, entre dezembro de 1987 e março de 1988. Recrutou-se uma amostra consecutiva de todas as crianças com diarreia grave, internadas em um dos hospitais. No segundo hospital, recrutou-se uma amostra consecutiva de pacientes hospitalizados em três dias da semana, aleatoriamente determinados. Definiu-se diarreia pela presença de mais de três evacuações com consistência líquida ou semilíquida durante um período de 24 h, em crianças com três meses ou mais<sup>17</sup>. Para as crianças menores de três meses aceitou-se a definição da mãe sobre diarreia, ou seja, número aumentado de evacuações, com consistência reduzida, que deixava a mãe preocupada<sup>15, 17</sup>. A presença ao exame físico de prega cutânea persistente, associada a pelo menos um outro sinal

(fontanela deprimida, língua seca ou com umidade diminuída, olhos encovados, volume urinário reduzido ou nulo, pulso radial fraco ou ausente, psiquismo deprimido ou comatoso), caracterizou a desidratação como moderada a grave. Estabeleceu-se o grau de desidratação através de exame físico padronizado realizado por pediatra treinado.

Selecionaram-se como controles crianças residentes na mesma vizinhança e emparelhadas individualmente pelo grupo etário do caso, 0 a 11 meses ou 12 a 23 meses, sem diarreia nos últimos sete dias. Excluíram-se casos e controles com doenças como pneumonia, doença infecto-contagiosa, irmãos e crianças institucionalizadas.

Estruturou-se um marco teórico com as variáveis associadas com a mortalidade e a incidência de diarreia, discriminando-as em blocos hierarquizados (Fig.), uma vez que inexistiam dados relativos à diarreia grave. O marco teórico priorizou o determinismo econômico sobre as características sociais e a cronologia de suas relações na hierarquização dos blocos<sup>16</sup>. O padrão social e econômico estabelece as condições de habitação e do ambiente onde nasce e cresce a criança. O ambiente, permeando a aproximação entre os agentes causais e a criança, possibilita a instalação de um quadro de diarreia. A gravidade da diarreia é facilitada por características maternas e da criança, como por exemplo, mãe adolescente, baixo peso de nascimento, ausência de aleitamento materno e desnutrição.



**Figura** - Marco teórico para a investigação de fatores de risco para diarreia grave estruturado em blocos hierarquizados.

Investigou-se a associação entre diarreia grave e os seguintes fatores de risco:

- \* *demográficos*: sexo e idade (0-1, 2-3, 4-5, 6-8, 9-11, 12-17 e 18-23 meses),
- \* *socioeconômicos*: renda familiar (< 3,6 e ≥ 3,6 salários-mínimos), presença e escolaridade do pai

(0 ou ausente e  $\geq 1$  ano), escolaridade da mãe ( $< 8$  e  $\geq 8$  anos), cor da mãe biológica e trabalho materno (não trabalha ou licença, ocasional, diário parcial e turno integral),

- \* *ambientais*: número de pessoas na casa (2-3, 4-5,  $\geq 6$ ), número de crianças menores de 5 anos (1-2, 3-6), aglomeração- número de pessoas por peça- ( $\leq 2$ ,  $> 2$ ), tipo de construção (tijolo ou madeira, maloca), geladeira em casa, vaso sanitário com descarga, água disponível (dentro de casa, pátio da casa, fora do terreno, poço ou rio) e ambiente doméstico limpo (índice composto pela observação dos seguintes itens: limpeza da casa, disponibilidade de sabão para lavar as mãos e toalha para secar, presença de panelas com comida a vista, panelas destampadas, moscas na cozinha, lixo a vista, presença de animais, fezes e água empoçada no pátio),
- \* *reprodutivos maternos*: idade materna ( $< 20$  e  $\geq 20$  anos), consulta no pré-natal, ordem de nascimento ( $1^{\circ}$ - $2^{\circ}$ ,  $3^{\circ}$  e  $\geq 4^{\circ}$ ), intervalo de nascimento (não tem irmão,  $\geq 48$ , 47-30, 29-18 e  $< 18$  meses) e gemelaridade,
- \* *nutricionais antropométricos*: peso de nascimento ( $\geq 3.000$ , 2.999-2.500 e  $< 2.500$  gramas), déficit altura-idade, déficit peso-idade e déficit peso-altura ( $> -1$ ,  $-1$  a  $-1,9$  e  $\leq -2$ ) foram estudados em relação ao número de desvios-padrão de mediana, comparativamente ao padrão do índice de referência do "National Center for Health Statistics" (NCHS).
- \* *nutricionais dietéticos*: tipo de leite (leite materno, leite materno+leite de vaca reconstituído, leite materno+leite de vaca em pó, leite de vaca reconstituído, leite de vaca em pó), história de amamentação (continua, parou e nunca mamou) e dieta prévia (leite materno/leite materno+água, leite materno+leite artificial, leite materno+sólidos, leite materno+ leite artificial+sólidos, leite artificial, leite artificial+sólidos/sólidos).

As variáveis foram obtidas através de entrevista, utilizando-se um questionário padronizado e pré-codificado, previamente testado e as variáveis ambientais foram coletadas através de observação. A confiabilidade das informações foi avaliada através de nova visita domiciliar (10%), de segundo contato no hospital (20%) e supervisão de 15% das entrevistas. O peso foi aferido com a criança desnuda, deitada (para menores de 6 meses) ou sentada, em balança Salter com aproximação para 100 gramas. Os casos foram pesados na admissão e após a reidratação, quando não havia sinais de desidratação, enquanto que os controles foram pesados no domicílio. O comprimento foi aferido com a criança deitada, em infantômetro similar ao modelo desenvolvido pelo AHRTAG ("Appropriate Health Resources and Technologies Action Group Ltda., London").

Consideraram-se como fatores de risco independentes aqueles cuja razão de produtos cruzados diferiu de um, sendo estatisticamente significativa ( $p < 0,05$ , bicaudal), no limiar da significância ( $0,05 < p < 0,10$ , bicaudal) ou

não significativa ( $p > 0,10$ ). Empregou-se a expressão risco ou proteção provável para as associações que alcançaram significância limítrofe. A amostra conferiu um poder de 80% para detectar-se uma razão de produtos cruzados de 1,5 como estatisticamente significativa a nível de 5% (bicaudal), para prevalência de exposição aos fatores de risco para os controles variando entre 20% e 65% .

## Análise dos Dados

Investigou-se a associação de cada uma das variáveis com diarreia grave, através da razão de produtos cruzados empregando-se método que estima a razão de máxima verossimilhança condicional, do módulo PECAN do programa EGRET. As variáveis categóricas e discretas foram mantidas como tal, utilizando-se os pontos de corte adotados na literatura. Utilizou-se o marco teórico hierarquizado como o referencial para a construção de um modelo que permitisse o ajuste para fatores de confusão<sup>12</sup>. O modelo proposto discriminou os fatores de risco hierarquicamente superiores como exercendo sua ação através daqueles situados inferiormente. Por exemplo, se o efeito de baixo nível socioeconômico se expressa através de moradia em condições insalubres e da ocorrência de desnutrição, estas características não devem ser consideradas fatores de confusão, mas fatores intermediários. Assim, as características ambientais e desnutrição não deveriam entrar no modelo.

A seguir, identificou-se a melhor escala para a inclusão de cada variável no ajuste de fatores de confusão. Empregou-se a razão de máxima verossimilhança, medindo-se a diferença entre a função desvio ("deviance") do modelo vazio e aquela incluindo a variável. Considerou-se como mais adequada a escala que com menor número de graus de liberdade manteve a significância estatística, sem desviar-se da linearidade. Ao estudarem-se múltiplas exposições, o modelo hierarquizado indica a ordem em que devem entrar as variáveis. Assim, identifica-se que o risco de desenvolver diarreia grave associado à desnutrição deve ser quantificado após o ajuste para nível socioeconômico, condições ambientais, características maternas, idade da criança e peso de nascimento. Uma vez que muitos fatores em estudo poderiam estar confundindo as associações, utilizou-se a hierarquização das variáveis em blocos para selecionarem-se as variáveis de confusão mais fortemente associadas com diarreia grave. Empregou-se um algoritmo e utilizou-se regressão logística condicional, com o processo retrógrado ("backward") de seleção, através do módulo em passos ("stepwise"), do programa . Consideraram-se como potenciais fatores de confusão as variáveis selecionadas segundo um p alfa de 0,10 dentro de cada bloco. Para as análises subsequentes, mantiveram-se aquelas variáveis que permaneceram associadas com diarreia grave após o ajuste para as variáveis de confusão do mesmo bloco e para daquelas hierarquicamente superiores.

Testou-se o ajuste da modelagem através dos componentes sistemático e de erro. A construção do modelo se concentra no componente sistemático que é função

matemática das variáveis independentes que caracterizam o “valor médio” de Y. Na verdade, para o modelo logístico não se trata de média e sim da transformação logística de y para um dado vetor x. Analisou-se o componente sistemático através da medida ordinal de associação gama entre os valores observados para a variável dependente e aqueles preditos pelo modelo<sup>1</sup>. Calculou-se o gama através da diferença entre as probabilidades de concordância e discordância, variando de -1 a +1. O componente de erro é avaliado através do exame individual das observações, cujo valor individual se afasta da “média”, podendo ser problemático sob o modelo atual, e pela medida coletiva do erro. O componente de erro foi mensurado coletivamente pelo teste de ajustamento (“goodness of fit”), analisando-se a função desvio para o acréscimo de cada bloco (apenas com as variáveis selecionadas). Calculou-se a diferença da função desvio entre os dois modelos, testando-se a significância pela distribuição do qui-quadrado de Pearson, com o número de graus de liberdade calculados pela diferença entre o número de parâmetro nos dois modelos<sup>10</sup>.

## RESULTADOS

No período estudado incluíram-se 200 casos. Entre estes, perdeu-se 4% (8 casos) devido a não localização do endereço e a recusas. Assim, estudaram-se 192 casos e 192 controles, sendo que 85% eram menores de um ano e aproximadamente 50% do sexo masculino. No grupo menor de um ano (Tabela 1), identificaram-se riscos de maior magnitude para as crianças entre dois e cinco meses, e para aquelas no segundo ano de vida detectou-se maior risco para as menores de dezoito meses.

**Tabela 1** - Associação entre fatores de risco demográficos e diarreia grave.

Variável	Razão de produtos cruzados (IC 95 %)
<b>Idade (meses)</b>	
0-1	2,6 (1,3-5,5)
2-3	7,1 (3,0-16,5)
4-5	3,5 (1,6-7,5)
6-8	2,4 (1,2-4,8)
9-11	1,0
p	< 0,001
12-17	3,7 (1,0-13,1)
18-23	1,0
p	0,03
<b>Sexo</b>	
Masculino	1,0
Feminino	0,9 (0,6-1,4)
p	0,7

## Fatores de Risco Socioeconômicos

Na Tabela 2 exemplificam-se com as variáveis socioeconômicas as diferentes escalas das variáveis que poderiam entrar no ajuste para fatores de confusão. Escolaridade dos pais e renda, coletadas como variáveis contínuas, poderiam ser mantidas como tal ou categorizadas, diferentemente de trabalho materno e cor da pele. Renda familiar, presença e escolaridade do pai e escolaridade da mãe foram examinadas como potenciais fatores de confusão dicotomizadas. A escolaridade do pai categorizadas em cinco níveis, desviou-se significativamente da linearidade.

No bloco de características socioeconômicas, selecionaram-se renda familiar, presença e escolaridade do pai e cor da pele da mãe como fatores de confusão. As

**Tabela 2** - Seleção da escala das variáveis socioeconômicas para inclusão na modelagem de fatores de confusão.

Variável	Função desvio	Razão de verossimilhança	Graus de liberdade	p
<b>Renda familiar (SM)*</b>				
Contínua	262,56	3,61	1	0,06
**2 Categorias	262,09	4,08	1	0,04
Quartis	259,11	7,06	3	0,07
5 Categorias	261,22	4,95	4	0,29
Desvio da linearidade		1,58	1	0,21
<b>Presença e escolaridade do pai (anos)</b>				
**2 Categorias	253,19	12,98	1	< 0,001
5 Categorias	248,83	17,34	4	0,001
Desvio da linearidade		5,51	1	0,02
<b>Escolaridade da mãe (anos)</b>				
Contínua	262,59	3,58	1	0,06
**2 Categorias	261,84	4,33	1	0,04
4 Categorias	261,61	4,56	3	0,21
Desvio da linearidade		2,45	1	0,12

\* SM: salários-mínimos

\*\* Escala selecionada para a modelagem de fatores de confusão

**Tabela 3** - Associação entre diarreia grave e fatores de risco socioeconômicos.

Variável	Razão de produtos cruzados (IC 95%) a	Razão de produtos cruzados (IC 95%) b
<b>Renda familiar (SM)*</b>		
≥ 3,6	1,0	1,0
< 3,6	1,9 (1,1-3,2)	1,8 (1,0-3,1)
p	0,01	0,04
<b>Trabalho materno</b>		
Não/licença gestante	1,0	1,0
Ocasional	1,5 (0,7-3,2)	1,2 (0,5-2,6)
Turno parcial diário	1,3 (0,6-2,7)	1,1 (0,5-2,4)
Turno integral	1,3 (0,7-2,2)	1,1 (0,7-2,2)
p	0,7	1,0
<b>Presença e escolaridade do pai (anos)</b>		
≥ 1	1,0	1,0
0 ou ausente	2,5 (1,5-4,3)	2,7 (1,6-4,8)
p	<0,001	<0,001
<b>Escolaridade da mãe (anos)</b>		
≥ 8	1,0	1,0
< 8	1,9 (1,0-3,4)	1,5 (0,8-2,8)
p	0,04	0,20
<b>Cor da mãe biológica</b>		
Branca	1,0	1,0
Negra	1,7 (0,9-3,0)	1,7 (0,9-3,1)
Mista	2,3 (1,3-4,0)	2,2 (1,2-4,1)
p	0,01	0,02

(a) Não ajustada

(b) Ajustada para: renda familiar, presença e escolaridade do pai e cor da mãe biológica

\* SM: salários mínimos

variáveis selecionadas mantiveram-se associadas com diarreia grave, mesmo após o ajuste para os demais fatores de confusão dentro do bloco, permanecendo, portanto, para a análise dos blocos seguintes.

As características socioeconômicas de famílias com renda inferior a 3,6 salários-mínimos, pai ausente ou analfabeto e mãe com cor da pele diferente da branca foram identificadas como fatores de risco para diarreia grave (Tabela 3). O ajuste para os fatores de confusão reduziu substancialmente o risco representado pela pouca escolaridade materna. O trabalho materno manteve-se não associado ao risco de desenvolver diarreia grave, mesmo controlando-se para os fatores de confusão.

### Fatores de Risco Ambientais

Entre as variáveis ambientais: número total de pessoas que residem na casa, número de crianças menores de cinco anos, aglomeração e índice de limpeza do ambiente, todas foram incluídas no modelo sob a forma contínua, exceto o índice de limpeza do ambiente doméstico que foi dicotomizado.

A modelagem dos fatores de confusão ambientais identificou o índice de limpeza do ambiente doméstico, o número de menores de cinco anos residindo

na casa, a disponibilidade de água e o tipo de construção como as variáveis mais fortemente associadas com diarreia grave. Calculadas as razões de produtos cruzados ajustadas para estas variáveis e para as socioeconômicas, todos os fatores de confusão permaneceram no modelo. Embora tenha se reduzido a magnitude da associação para a disponibilidade de água, houve tendência à significância preenchendo o critério para manter-se na modelagem.

Observa-se, na Tabela 4, através das razões de produtos cruzados emparelhadas e não ajustadas (primeira coluna) que a grande maioria das variáveis ambientais associaram-se significativamente com diarreia grave. Ao acrescentar-se o ajuste para os fatores de confusão ambientais (segunda coluna), a aglomeração, a disponibilidade de geladeira e de vaso sanitário com descarga tornaram-se estatisticamente não significativas. O acréscimo das variáveis socioeconômicas (terceira coluna) no modelo não modificou marcadamente as associações.

A coabitação de mais do que duas crianças menores de cinco anos, a moradia em maloca e a sujeira do ambiente doméstico associaram-se com maior risco de desenvolver diarreia grave. A ausência de água encanada dentro de casa, por outro lado, mostrou apenas tendência a maior risco.

**Tabela 4** - Associação entre diarreia grave e fatores de risco ambientais.

Variável	Razão de produtos cruzados (IC 95%) a	Razão de produtos cruzados (IC 95%) b	Razão de produtos cruzados (IC 95%) c
<b>Nº pessoas da casa</b>			
2-3	1,0	1,0	1,0
4-5	1,6 (0,9-2,8)	1,6 (0,8-3,2)	1,6 (0,8-3,3)
≥ 6	1,9 (1,0-3,6)	1,6 (0,8-3,4)	1,4 (0,6-3,1)
p	0,13	0,4	0,4
<b>Nº crianças &lt; 5 anos</b>			
1-2	1,0	1,0	1,0
3-6	3,6 (1,7-7,4)	2,7 (1,2-5,9)	2,4 (1,0-5,6)
p	< 0,001	0,01	0,03
<b>Aglomeracão</b>			
≤ 2	1,0	1,0	1,0
> 2	1,7 (1,1-2,6)	1,0 (0,6-1,8)	1,0 (0,5-1,7)
p	0,03	0,8	0,8
<b>Tipo de construção</b>			
Tijolo ou madeira	1,0	1,0	1,0
Maloca	3,4 (1,8-6,3)	2,3 (1,1-4,6)	2,3 (1,1-5,0)
p	< 0,001	0,02	0,03
<b>Geladeira em casa</b>			
Sim	1,0	1,0	1,0
Não	1,9 (1,2-3,0)	1,1 (0,6-2,0)	1,1 (0,6-2,2)
p	0,006	0,8	0,7
<b>Vaso sanitário com descarga</b>			
Sim	1,0	1,0	1,0
Não	2,3 (1,4-3,9)	1,0 (0,5-2,0)	1,0 (0,5-2,1)
p	< 0,001	0,9	0,9
<b>Água disponível</b>			
Dentro de casa	1,0	1,0	1,0
Pátio da casa	2,6 (1,4-4,7)	1,9 (1,0-3,7)	2,1 (1,1-4,3)
Fora do terreno	1,9 (0,9-4,3)	1,0 (0,4-2,5)	1,0 (0,4-2,6)
Poço ou rio	4,8 (1,3-17,8)	3,4 (0,8-14,2)	3,0 (0,7-12,5)
p	< 0,001	0,08	0,08
<b>Ambiente doméstico limpo</b>			
Sim	1,0	1,0	1,0
Não	4,1 (2,3-7,2)	3,1 (1,7-5,9)	2,9 (1,5-5,6)
p	< 0,001	< 0,001	< 0,001

(a) Razão de produtos cruzados não ajustada

(b) Razão de produtos cruzados ajustada para os fatores de confusão ambientais

(c) Razão de produtos cruzados ajustada para os fatores de confusão socioeconômicos e ambientais

## Fatores de Risco Reprodutivos Maternos

As características reprodutivas maternas, Tabela 5, associaram-se com diarreia grave e foram selecionadas como potenciais fatores de confusão. Idade materna foi testada como variável contínua, com duas e quatro categorias, sendo selecionada a dicotômica para a modelagem. As demais variáveis maternas foram mantidas com a escala apresentada na Tabela 5. O ajuste para as variáveis de confusão do mesmo bloco (segunda coluna) modificou a magnitude de algumas associações. Contudo, ao acrescentar-se o ajuste para os fatores de confusão ambientais e socioeconômicos (terceira coluna), houve alteração substancial, sendo que consulta no período

do pré-natal, ordem e intervalo de nascimento tornaram-se estatisticamente não significativos. Mãe adolescente e gemelaridade caracterizaram riscos independentes, respectivamente, 3,8 e 11,6 vezes maior da criança desidratar.

## Fatores de Risco Nutricionais

As características nutricionais identificadas como fatores de confusão incluíram o peso de nascimento, o déficit peso-idade e a dieta prévia. As características antropométricas entraram no ajuste de fatores de confusão como variáveis contínuas (tendência); o tipo de leite consumido e as outras características dietéticas entraram como variáveis categóricas.

**Tabela 5** - Associação entre diarreia grave e fatores de risco maternos.

Variável	Razão de produtos cruzados (IC 95%) a	Razão de produtos cruzados (IC 95%) b	Razão de produtos cruzados (IC 95%) c
<b>Idade materna (anos)</b>			
≥ 20	1,0	1,0	1,0
< 20	2,5 (1,4-4,6)	5,0 (2,3-11,1)	3,8 (1,6-8,9)
p	< 0,001	< 0,001	< 0,001
<b>Consulta no pré-natal</b>			
Sim	1,0	1,0	1,0
Não	2,6 (1,4-4,9)	2,3 (1,1-4,5)	1,6 (0,7-3,4)
p	< 0,001	0,02	0,2
<b>Ordem de nascimento</b>			
1º-2º	1,0	1,0	1,0
3º	2,2 (1,2-3,6)	2,4 (1,2-5,0)	2,1 (0,9-4,8)
≥ 4º	1,6 (1,0-2,6)	1,3 (0,7-2,5)	1,2 (0,5-2,4)
p	0,02	0,04	0,2
<b>Intervalo de nascimento (meses)</b>			
Não tem irmão	1,0	1,0	1,0
≥ 48	0,8 (0,4-1,6)	1,2 (0,5-2,9)	1,6 (0,6-4,5)
47-30	1,2 (0,6-2,5)	1,3 (0,6-3,1)	0,9 (0,3-2,6)
29-18	1,7 (1,0-3,1)	2,5 (1,1-5,6)	1,7 (0,6-5,2)
< 18	4,6 (2,0-10,3)	5,2 (1,9-14,3)	3,6 (1,0-12,5)
p	< 0,001	< 0,001	0,17
<b>Gemelaridade</b>			
Não	1,0	1,0	1,0
Sim	5,5 (1,2-24,8)	6,1 (1,1-32,5)	11,6 (0,9-143,7)
p	0,009	0,02	0,03

(a) Razão de produtos cruzados não ajustada

(b) Razão de produtos cruzados ajustada para os fatores de confusão maternos

(c) Razão de produtos cruzados ajustada para os fatores de confusão socioeconômicos, ambientais e maternos.

O peso de nascimento foi analisado (Tabela 6) ajustando-se apenas para a idade materna (primeira coluna) e para os demais fatores de confusão (terceira coluna). O controle para variáveis de confusão não modificou substancialmente o risco representado pelo peso inferior a 2.500 gramas.

As demais variáveis antropométricas ajustadas para idade da criança (primeira coluna) associaram-se marcadamente com diarreia grave, modificando-se a magnitude das associações ao adicionar-se o controle para os fatores de confusão nutricionais (segunda coluna). A inclusão das variáveis de confusão maternas, ambientais e socioeconômicas reduziu a maior parte das razões de produtos cruzados. Portanto, as características antropométricas, associadas com o desenvolvimento de diarreia grave, foram o peso de nascimento, os déficits altura-idade e peso-idade.

Na Tabela 7, identificam-se com maior risco de apresentar diarreia grave o consumo de outro leite que não o materno, a história de ter parado a amamentação e a dieta prévia à doença. As associações não se alteraram substancialmente após a adição dos fatores de confusão no modelo. A dieta prévia ao episódio caracterizou-se como apresentando risco quando composta por leite artificial ou sólidos.

O pequeno número absoluto de crianças desnutridas recebendo leite artificial gerou instabilidade no modelo e intervalos de confiança amplos.

### Avaliação do Ajuste da Modelagem

Teoricamente, o modelo se estrutura nas relações entre a variável dependente (Y) e o conjunto de variáveis independentes (X), analisadas estatisticamente pelo modelo de regressão logística. As características socioeconômicas, exclusivamente, permitiriam identificar 29% dos casos, elevando-se para 49% ao incluírem-se as ambientais, mantendo-se em 49% ao adicionar-se a idade, e elevando-se para 74% com a inclusão das variáveis nutricionais. Quando todas as variáveis significativas foram incluídas na modelagem, uma proporção maior de casos foi identificada (74%). A Tabela 8 apresenta o teste de ajustamento para cada bloco de características estudadas. A inclusão dos blocos de variáveis socioeconômicas, ambientais, reprodutivas maternas e nutricionais contribuiu significativamente para a modelagem. A inclusão da variável idade da criança alcançou significância limítrofe, contudo este ajuste torna-se necessário visto que as exposições diferem segundo a idade.



**Tabela 6** - Associação entre diarreia grave e fatores de risco nutricionais antropométricos.

Variável	Razão de produtos cruzados (IC 95%) a	Razão de produtos cruzados (IC 95%) b	Razão de produtos cruzados (IC 95%) c
<b>Peso de nascimento (gramas)</b>			
≥ 3.000	1,0 @		1,0 #
2.999-2.500	1,8 (1,1-3,1)		2,4 (1,2-4,9)
< 2.500	3,5 (1,8-6,8)		3,6 (1,5-8,8)
p	< 0,001		0,002
<b>Déficit altura-idade*</b>			
> -1	1,0	1,0	1,0
-1 a -1,9	3,5 (2,0-6,2)	2,1 (1,0-4,1)	1,4 (0,6-3,4)
≤ -2	7,1 (3,6-14,2)	3,6 (1,6-8,0)	4,5 (1,3-15,5)
p	<0,001	0,002	0,04
<b>Déficit peso-idade*</b>			
> -1	1,0	1,0	1,0
-1 a -1,9	2,7 (1,5-4,8)	2,8 (1,3-5,9)	2,4 (1,0-5,9)
≤ -2	8,0 (3,6-17,7)	6,5 (2,5-16,9)	3,9 (1,1-14,1)
p	<0,001	<0,001	0,04
<b>Déficit peso-altura*</b>			
>-1	1,0	1,0	1,0
-1 a -1,9	1,5 (0,8-2,7)	1,9 (1,3-5,9)	1,5 (0,5-4,2)
≤ -2	5,4 (1,5-19,3)	7,6 (2,5-24,8)	1,9 (0,2-16,9)
p	0,007	0,007	0,7

(a) Razão de produtos cruzados ajustada para idade da criança

(b) Razão de produtos cruzados ajustada para os fatores de confusão nutricionais

(c) Razão de produtos cruzados ajustada para os fatores de confusão socioeconômicos, ambientais, maternos, demográficos e nutricionais

@ Ajustado apenas para a idade materna

# Ajustada para as variáveis mencionadas em c, exceto idade da criança

\* Escores-z

**Tabela 7** - Associação entre diarreia grave e fatores de risco nutricionais dietéticos.

Variável	Razão de produtos cruzados (IC 95%) a	Razão de produtos cruzados (IC 95%) b	Razão de produtos cruzados (IC 95%) c
<b>Tipo de leite</b>			
LM	1,0	1,0	1,0
LM+LV	1,7 (0,7-4,4)	1,9 (0,7-5,4)	1,3 (0,3-4,9)
LM+LPO	4,2 (0,9-20,2)	1,7 (0,3-9,4)	2,2 (0,3-17,2)
LV	5,8 (2,5-13,5)	6,2 (2,4-16,6)	6,0 (1,8-19,8)
LPO	10,1 (3,3-31,2)	8,2 (2,3-28,8)	6,9 (1,4-33,3)
p	< 0,001	< 0,001	0,006
<b>História de amamentação</b>			
Contínua	1,0	1,0	1,0
Parou	4,3 (2,3-8,2)	4,9 (2,3-10,6)	6,4 (2,3-17,3)
Nunca mamou	3,7 (1,2-11,1)	2,8 (0,8-9,8)	0,7 (0,1-3,7)
p	< 0,001	< 0,001	< 0,001
<b>Dieta prévia</b>			
LM/LM+H <sub>2</sub> O	1,0	1,0	1,0
LM+LA	5,3 (1,4-19,7)	4,5 (1,1-18,6)	5,5 (0,7-43,5)
LM+Sólidos	0,9 (0,2-4,5)	1,2 (0,2-6,7)	2,2 (0,3-28,4)
LM+LA+Sólidos	1,0 (0,2-4,2)	1,1 (0,2-5,6)	1,5 (0,1-16,3)
LA	40,0 (9,0-178,3)	35,6 (7,0-182,6)	54,7(4,9-609,4)
LA+Sól./Sólidos	2,8 (0,8-10,4)	3,2 (0,7-14,0)	5,4 (0,6-48,4)
p	< 0,001	< 0,001	< 0,001

(a) Razão de produtos cruzados ajustada para idade da criança

(b) Razão de produtos cruzados ajustada para os fatores de confusão nutricionais

(c) Razão de produtos cruzados ajustada para os fatores de confusão socioeconômicos, ambientais, maternos, demográficos e nutricionais

LM: leite materno, LV: leite de vaca reconstituído, LPO: leite de vaca em pó LA: leite de vaca

Sólidos: alimentos não líquidos

**Tabela 8** - Avaliação do componente de erro através da função desvio, analisada a contribuição de cada bloco para o ajuste do modelo.

Bloco de variáveis	Função desvio	Graus de liberdade	Bloco adicionado		
			Qui-quadrado	Graus de liberdade	p
Modelo vazio	266,17				
Nível socioeconômico	229,50	7	36,67	7	< 0,001
Nível socioeconômico + ambientais	192,14	11	37,36	4	< 0,001
Nível socioeconômico + ambientais + maternas	179,42	13	12,72	2	0,002
Nível socioeconômico + ambientais + maternas + demográficas	166,70	19	12,72	6	0,05
Nível socioeconômico + ambientais + maternas + demográficas + nutricionais	115,89	26	50,81	7	< 0,001

## DISCUSSÃO

A análise hierarquizada permitiu identificar as características que elevam o risco da criança desidratada e cujo efeito é independente de idade, nível socioeconômico, condições ambientais, história reprodutiva materna e fatores nutricionais. As condições ambientais exercem efeito além daquele propiciado pelo padrão socioeconômico. A história reprodutiva materna afeta o risco de desenvolver diarreia grave, além do efeito representado pelo nível socioeconômico e pelas condições ambientais. Os fatores de risco nutricionais atuam independentemente da idade da criança e das características maternas, ambientais e socioeconômicas sobre a ocorrência de diarreia grave.

A caracterização de fatores de risco independentes baseou-se na identificação e ajuste para fatores de confusão, através da modelagem. O objetivo desta é identificar um modelo que com um número suficiente de variáveis permita a interpretação adequada dos dados<sup>7</sup>. A seleção de variáveis de confusão mantém-se como um processo quase subjetivo que consiste, parcialmente, em incluir variáveis pertinentes adequadamente medidas<sup>4</sup>.

O modelo hierarquizado, aplicado à investigação de fatores de risco para diarreia grave parece sobrepujar esses problemas. Em primeiro lugar, determinaram-se as variáveis de interesse a partir do conhecimento vigente sobre diarreia. A seguir, estruturou-se um modelo hierarquizado para discriminar as relações entre as variáveis e destas com diarreia grave. Assim, sabe-se que baixa renda familiar não é causa direta de doença mas, através da moradia em condições insalubres, pode determinar que maior carga de bactérias enteropatogênicas elevem a frequência de episódios de diarreia, que, por sua vez, pioram o estado nutricional. Portanto, ao investigar-se o efeito deletério do ambiente sobre a ocorrência

de diarreia grave, deveriam ser levados em conta fatores de confusão, como nível socioeconômico.

Em estudos como este, que se propõe a avaliar uma série de fatores de risco e trabalha com um grande número de covariáveis, torna-se difícil determinar quantos e quais são os termos necessários para uma descrição adequada dos dados<sup>7, 10</sup>. A inclusão de todo o conjunto de variáveis explanatórias, como propõe Rothman<sup>13</sup>, poderia significar uma regressão com aproximadamente 30 termos. Além de haver algum grau de colinearidade entre algumas variáveis, o excesso de termos para uma amostra limitada geraria instabilidade do modelo. Assim, em cada bloco determinaram-se as covariáveis mais fortemente associadas com diarreia grave usando-se um nível de significância flexível ( $p < 0,10$ ). Alguns autores preconizam um nível de significância ainda mais flexível ( $p < 0,20$ )<sup>7</sup>. Este, se empregado, levaria a inclusão de mais uma variável (déficit de peso-altura).

Os métodos empregados na criação de modelos são sujeitos a erros e não há um método único para identificar-se o melhor modelo<sup>6</sup>. A significância estatística na análise univariada, que frequentemente rege a identificação dos fatores de confusão, decorre tanto do tamanho da amostra em estudo quanto da magnitude da associação, enquanto que a presença de confusão depende apenas da magnitude da associação entre fator de confusão, fator em estudo e desfecho clínico<sup>6</sup>. Uma das vantagens da análise hierarquizada é que os critérios de seleção são previamente estabelecidos e visualiza-se o desenvolvimento do processo. Assegurou-se a adequação das variáveis modeladas através de sua avaliação em escala linear, dicotômica ou com maior número de categorias<sup>8</sup>. A seleção dos fatores de confusão socioeconômicos exemplifica como foi identificada a melhor escala para cada variável entrar na modelagem.

A avaliação do ajustamento da modelagem, através do componente sistemático (associação gama) mostra que uma proporção elevada de casos e controles seriam corretamente identificados através dos fatores de risco investigados. Contudo, mesmo a idade estando significativamente associada com diarreia grave, não seria capaz de melhorar a predição do desfecho clínico. Ao avaliar-se o componente de erro poderiam ser analisadas as observações individuais que se afastam da “média”<sup>7</sup>. Optou-se, contudo pela verificação da contribuição de cada bloco uma vez que a análise hierarquizada centra-se na medida de efeito a partir da inclusão de blocos. Uma vez que idade foi incluída no modelo e não representou um ganho substancial para o ajuste, deveria ser removida da modelagem caso o critério fosse puramente estatístico<sup>4</sup>. O exemplo a seguir torna evidente a importância biológica desta variável na mo-

delagem<sup>15</sup>. Crianças menores têm maior chance de serem amamentadas, mas quando sob aleitamento artificial têm maior risco de desidratar.

Concluindo, a modelagem hierarquizada representa uma alternativa aos métodos tradicionais de análise. Contempla os aspectos “biológicos” e estatísticos, permitindo estruturar a investigação dos fatores de risco e facilitar a interpretação dos resultados. A verificação do ajuste da modelagem mostrou o ganho representado pela inclusão de cada bloco.

## AGRADECIMENTOS

Ao Dr. José Martinez, coordenador do Programa de Controle da Doença Diarréica da Organização Mundial de Saúde, por seus comentários críticos na estruturação do projeto.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. AGRESTI, A. *Categorical data analysis*. New York, Wiley-Interscience Publication, 1989.
2. DAY, N.E.; BYAR, D.P.; GREEN, S.B. Overadjustment in case-control studies. *Am. J. Epidemiol.*, **112**:696-706, 1980.
3. EGBUONU, L. & STARFIELD, B. Child health and social status. *Pediatrics*, **69**:550-7, 1982.
4. EVANS, S.J. Uses and abuses of multivariate methods in epidemiology. *J. Epidemiol. Community Health*, **42**:311-5, 1988.
5. FUCHS, S.C. Fatores de risco para diarreia complicada por desidratação moderada a grave: um estudo de casos e controles. Porto Alegre, 1993 [Tese de Doutorado - Faculdade de Medicina da UFGRS].
6. GREENLAND, S. Modeling and variable selection in epidemiologic analysis. *Am. J. Public Health*, **79**:340-9, 1989.
7. LEMESHOW, S. & HOSMER, D.W. *Applied logistic regression*. New York, Wiley-Interscience Publication, 1989.
8. LEW, R.A.; DAY, C.L.; HARRIST, T.J.; WOOD, W.C.; MIHM JR., M.C. Multivariate analysis. *JAMA*, **249**:641-3, 1983.
9. MALDONADO, G. & GREENLAND, S. Simulation study of confounder-selection strategies. *Am. J. Epidemiol.*, **138**:923-36, 1993.
10. MICKEY, R.M. & GREENLAND, S. The impact of confounder selection criteria on effect estimation. *Am. J. Epidemiol.*, **129**:125-37, 1989.
11. MOSLEY, W.H. & CHEN, L.C. An analytical framework for the study of child survival in developing countries. *Popul. Develop. Rev.*, **10** (Supl):25-48, 1984.
12. OLINTO, M.T. Epidemiologia da desnutrição infantil em Pelotas. Pelotas, 1993. [Dissertação de Mestrado - Faculdade de Medicina da UFPel].
13. ROTHMANN, K.J. *Modern epidemiology*. Boston, Little Brown, 1986.
14. VANDENBROUCKE, J.P. Should we abandon statistical modeling altogether? *Am. J. Epidemiol.*, **126**:10-3, 1987.
15. VICTORA, C.G.; SMITH, P.G.; VAUGHAN, J.P.; NOBRE, L.C.; LOMBARDI, C.; TEIXEIRA, A.M.B.; FUCHS, S.C.; MOREIRA, L.B.; GIGANTE, L.P.; BARROS, F.C. Evidence for protection by breast-feeding against infant deaths from infectious diseases in Brazil. *Lancet*, **2**:319-21, 1987.
16. VICTORA, C.G.; FACCHINI, L.A.; BARROS, F.C.; LOMBARDI, C. Pobreza e saúde: como medir nível socioeconômico em estudos epidemiológicos de saúde infantil? In: Congresso Brasileiro de Epidemiologia, 10. Campinas, 1990. *Anais*, Campinas, 1990.p.302-15.
17. WORLD HEALTH ORGANIZATION. *The treatment and prevention of acute diarrhoea: practical guidelines*. Geneva, World Health Organization, 1989.