

SENSIBILIDADE E ESPECIFICIDADE

Ronaldo Rocha Bastos, Ricardo Rocha Bastos

Departamento de Estatística/ICE/UFJF; Departamento de Clínica Médica/Faculdade de Medicina/UFJF

INTRODUÇÃO

O objetivo desta seção é apresentar conceitos, interpretações e reflexões sobre ferramentas da estatística e da epidemiologia que sejam relevantes à prática médica. Espera-se que esta seção funcione como um veículo de formação continuada, onde os textos apresentados discutam as ferramentas mais encontradas na literatura, com ênfase nos seus aspectos práticos, de forma que os leitores possam utilizar o material aqui apresentado como fonte para trabalhos e estudos de natureza prática e também como referência para a melhor compreensão de trabalhos da área de saúde publicados nos periódicos especializados.

Nesta primeira oportunidade é apresentada uma discussão dos conceitos básicos das características dos testes diagnósticos - sob o enfoque de medicina baseada em evidências - e de sua utilidade na prática médica.

PROBABILIDADES DIAGNÓSTICAS

‘O critério fundamental para a utilidade de um teste diagnóstico é se o mesmo agrega informação a um nível além daquele já existente e se essa informação leva a uma mudança de conduta que seja basicamente benéfica ao paciente’¹.

Existe um sem número de testes diagnósticos comumente usados durante a prática médica. A escolha de um teste depende de vários fatores, dentre os quais se destacam: a prevalência da morbidade, a sensibilidade e especificidade do teste, os valores preditivos em diferentes situações e a razão de verossimilhança.

Vamos entender esses termos (também conhecidos como características operacionais dos testes diagnósticos)? Lembre-se que a discussão que se segue refere-se a qualquer ferramenta usada no processo diagnóstico, na triagem e na gestão em saúde: pergunta dirigida à entrevista, manobra de exame físico, observações eletrocardiográficas ou de imagem, resultados de exames laboratoriais, resultados de testes de triagem realizados em larga escala. Quando falamos em “teste”, podemos entender qualquer uma dessas ferramentas.

O processo diagnóstico tem início quando se considera a frequência de uma patologia num determinado meio. Trata-se da PREVALÊNCIA ou, em linguagem estatística, a PROBABILIDADE PRÉ-TESTE. A consideração de um diagnóstico começa com este conceito. Muitas ações médicas apoiam-se nesta idéia. Veja só: diversas comunidades amazônicas são orientadas para iniciar um curso de cloroquina, tão logo alguém apresente febre alta sem manifestações respiratórias óbvias. Isto porque a prevalência de malária é tão alta, que a probabilidade do diagnóstico estar certo, sem qualquer teste adicional, é bastante razoável. Considere ainda um senhor de 70 anos que procure um clínico. A probabilidade dele ter um tumor de próstata com volume superior a 0,5 ml é de 15%², isto independente de qualquer teste, e antes ainda dele apresentar a razão para a visita! A própria Organização Mundial de Saúde (OMS), através de seu protocolo da Abordagem Integrada das Doenças Prevalentes na Infância (AIDPI ou IMCI), reconhece este fato, recomendando antibióticos para toda criança entre 2 meses e 5 anos que se apresentar com febre, tosse e taquipnéia³ - por aceitar que a probabilidade pré-teste de uma pneumonia bacteriana nesta situação é alta o suficiente para dispensar uma radiografia! Conhecer a probabilidade pré-teste, portanto, é o primeiro passo no processo diagnóstico, isto porque quando a preva-

lência é muito baixa (qual a probabilidade de malária num residente urbano de uma cidade de porte médio da Zona da Mata de Minas Gerais, que não saiu de sua região nos últimos 2 anos?) o diagnóstico não deve ser perseguido com exames complementares e, quando a prevalência é muito alta (qual a probabilidade de uma infecção viral numa pessoa imunologicamente intacta que inicia febre baixa, rinorréia e tosse?), o diagnóstico já está firmado (quem se aventuraria a solicitar identificação viral num caso assim?). Para conhecer a probabilidade pré-teste das doenças em geral (uma tarefa nem sempre fácil), é recomendável que se consulte a literatura especializada⁴.

Uma vez definido que a probabilidade pré-teste (prevalência) de um diagnóstico é de nível tal a exigir investigações (lembre-se, se a prevalência for muito baixa, inicialmente o diagnóstico pode ser descartado; se for muito alta, o tratamento deve ser iniciado sem muita preocupação com métodos diagnósticos), os atributos dos testes diagnósticos devem ser considerados. O primeiro deles é a SENSIBILIDADE. Se de 100 pessoas com um diagnóstico, 70 têm positividade para um teste, a sensibilidade deste teste é de 70%.

SENSIBILIDADE: PORCENTAGEM DAS PESSOAS
COM O DIAGNÓSTICO, QUE APRESENTAM
POSITIVIDADE PARA O TESTE

Outro atributo de um teste é a ESPECIFICIDADE. Se de 100 pessoas sem um diagnóstico, 70 têm negatividade para um certo teste, a especificidade deste teste é de 70%.

ESPECIFICIDADE: PORCENTAGEM DAS PESSOAS
SEM O DIAGNÓSTICO QUE APRESENTAM
NEGATIVIDADE PARA O TESTE.

É importante salientar que sensibilidade e especificidade são atributos inerentes ao teste, ou seja, não dependem do ambiente em que se trabalha. Se um teste tem sensibilidade de 70%, isto é sua sensibilidade em qualquer situação, independente da prevalência observada na região de aplicação do teste diagnóstico, ou do grupo ao qual pertence o paciente testado. O mesmo ocorre para a especificidade.

Existe uma grande limitação dos conceitos de sensibilidade e especificidade: dependem do conhecimento prévio se há ou não a doença, ou seja, partem do princípio de que o diagnóstico já existe ou então de que o mesmo já foi descartado. Daí a necessidade de, ao se avaliar as características de um teste, comparar o desempenho do mesmo em relação a uma referência, ou seja, um teste que à luz dos conhecimentos acumulados e da tecnologia disponível apresente os resultados mais fidedignos: o *padrão-ouro* (*gold standard*). Nos casos de diagnóstico de apendicite, por exemplo, o padrão-ouro é considerado a inspeção direta do apêndice pelo cirurgião.

Sensibilidade e especificidade podem ser visualizadas em conjunto

numa tabela 2 X 2, como a apresentada na Figura 1.

Figura 1
Características operacionais de testes diagnósticos

Resultado do teste a ser avaliado	Resultado pelo padrão-ouro		Total
	+ (D)	- (D')	
+ (S)	a	b	a + b
- (S')	c	d	c + d
Total	a + c	b + d	a + b + c + d

D e D': resultado positivo ou negativo, respectivamente, pelo teste padrão-ouro.

S e S': resultado positivo ou negativo, respectivamente, pelo teste a ser avaliado (presença ou não do sinal diante da manobra)

a: acerto quando o teste avaliado tem resultado positivo

b: resultado falso positivo do teste avaliado

c: resultado falso negativo do teste avaliado

d: acerto quando o teste avaliado tem resultado negativo

a + b + c + d: total de casos avaliados

- Prevalência: porcentagem de pessoas com a doença em relação ao total de casos avaliados (ou, em linguagem estatística: $P(D)$, a probabilidade da doença). Calculada na tabela pelo quociente:

$(a + c) / (a + b + c + d)$

- Sensibilidade: porcentagem das pessoas com a doença que têm o teste positivo (ou, em linguagem estatística: $P(S/D)$, a probabilidade do teste positivo dada a presença de doença). Calculada na tabela pelo quociente: $a / (a + c)$

- Especificidade: porcentagem das pessoas sem a doença que têm o teste negativo (ou, em linguagem estatística: $P(S'/D')$, a probabilidade do teste negativo dada a ausência de doença). Calculada na tabela pelo quociente: $d / (b + d)$

- Valor Preditivo Positivo: a proporção de pessoas com o teste positivo que, efetivamente têm a doença (ou, em linguagem estatística, $P(D/S)$, a probabilidade de doença dado um teste positivo). Calculado na tabela pelo quociente: $a / (a + b)$

- Valor Preditivo Negativo: a proporção de pessoas com o teste negativo que, efetivamente não têm a doença (ou, em linguagem estatística, $P(D'/S')$, a probabilidade de não se ter a doença dado um teste negativo). Calculado na tabela pelo quociente: $d / (c + d)$

É importante considerarmos agora, que sensibilidade e especificidade não andam necessariamente juntas. Em outras palavras, o aumento da sensibilidade costuma ser acompanhado de uma queda da especificidade. Da mesma maneira, o aumento da especificidade costuma ser acompanhado de uma queda da sensibilidade.

É muito difícil, na prática médica, um teste diagnóstico que tenha, ao mesmo tempo, sensibilidade e especificidade de 100%. Daí, a necessidade de conhecermos bem as limitações da utilização dos resultados de um teste, destacadas nos pontos abaixo:

- Um teste de grande sensibilidade só é importante quando dá negativo (efetivamente descarta o diagnóstico!), visto que a probabilidade de resultados falso-negativos (teste negativo em pessoas com o diagnóstico) é muito baixa. A grande sensibilidade de um teste, sem levarmos em conta o valor de sua especificidade, pode vir acompanhada de alguns resultados falso-positivos (teste positivo em pessoas sem o diagnóstico).

- Os bancos de sangue, por exemplo, utilizam esta dimensão da sensibilidade para testar seus doadores. Não lhe parece claro que um teste muito sensível, em detrimento da especificidade, é ideal para, quando dá negativo, afastar a possibilidade de doença no doador? Mesmo que isto cause alguma ansiedade naqueles casos em que o teste dará positivo e não haverá a doença (falso-positivos).

- Uma manobra de grande especificidade, por outro lado, só é importante quando positiva (efetivamente dando o diagnóstico!), visto que a probabilidade de resultados falso-positivos (teste positivo em pessoas sem o diagnóstico) é muito baixa. A grande especificidade de um teste, sem levarmos em conta o valor de sua sensibilidade, pode vir acompanhada de alguns resultados falso-negativos (teste negativo mas com o diagnóstico presente).

- Os serviços especializados (considere um ambulatório de vasculites num Hospital Universitário) utilizam testes com alta especificidade, pois precisam ter precisão em seus diagnósticos, antes de submeterem os pacientes

a abordagens que podem ser muito agressivas (imunossupressores, por exemplo), mesmo cientes de que algumas pessoas terão o diagnóstico ainda que os resultados dos testes sejam negativos (os falso-negativos).

- Num ambiente de atenção primária, os testes de alta sensibilidade são geralmente os preferidos, mesmo sabendo que incluiremos algumas pessoas sem o diagnóstico em questão (os falso-positivos). O trabalho principal dos serviços especializados (de nível secundário para cima) é identificar e definir o significado desses falso-positivos.

Esta necessidade de se saber se há ou não doença, para a definição da sensibilidade e especificidade, vai na contramão da prática médica, onde a questão é outra: qual a probabilidade de doença face a um teste positivo; ou qual a probabilidade de não doença face a um teste negativo. É claro, então, que há a necessidade de se avaliar outros atributos.

Considere novamente a tabela 2 X 2, apresentada na Figura 1, e observe a grandeza correspondente à porcentagem, dentre todos os com teste positivo, daqueles que têm a doença. Pronto, você acabou de encontrar um novo atributo de um teste: VALOR PREDITIVO POSITIVO (por definição, a proporção de pessoas com o teste positivo que, efetivamente têm a doença, ou, em linguagem estatística, $P(D/S)$, a probabilidade de doença dado um teste positivo).

Na mesma tabela, observe a porcentagem, dentre todos com o teste negativo, daqueles que não têm a doença. Agora, você encontrou o VALOR PREDITIVO NEGATIVO (por definição, a proporção de pessoas com o teste negativo que, efetivamente, não têm a doença, ou, em linguagem estatística, $P(D'/S')$, a probabilidade de não se ter a doença dado um teste negativo).

Numa comparação, observa-se que o conceito de valor preditivo é mais adequado à real prática médica, em que o diagnóstico não é conhecido a priori.

A informação do resultado de um teste (S ou S') pode vir a modificar a

SENSIBILIDADE E ESPECIFICIDADE DEPENDEM DO CONHECIMENTO PRÉVIO SOBRE A EXISTÊNCIA OU AUSÊNCIA DA DOENÇA.

VALOR PREDITIVO POSITIVO E VALOR PREDITIVO NEGATIVO NÃO DEPENDEM DO CONHECIMENTO PRÉVIO SOBRE A EXISTÊNCIA OU AUSÊNCIA DA DOENÇA MAS, PELO CONTRÁRIO, INDICAM A PROBABILIDADE DE DOENÇA CASO O TESTE SEJA POSITIVO OU A PROBABILIDADE DE NÃO SE TER A DOENÇA CASO O TESTE SEJA NEGATIVO.

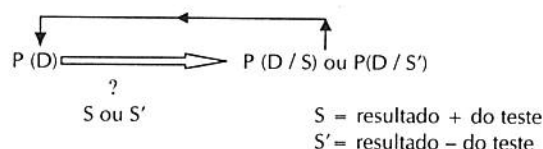
probabilidade a priori, denominada pré-teste (prevalência, representada por $P(D)$). Neste caso, passamos a ter uma probabilidade a posteriori, denominada pós-teste (valor preditivo positivo, representado por $P(D/S)$ ou o valor $P(D'/S')$, que corresponde a 1 - especificidade). Se não houver uma modificação da probabilidade a priori ao termos conhecimento do resultado do teste, isto indica que o teste em questão não acrescenta informação e, portanto, não deveria ter sido realizado. Este processo dinâmico da influência dos resultados dos testes sobre o diagnóstico é representado na Figura 2.

Se fôssemos calcular os valores preditivos (positivo ou negativo) de um mesmo teste (caracterizado por sua sensibilidade e especificidade) em situações de prevalência distintas, iríamos observar que estes valores iriam se modificar. Chegamos então a uma importante conclusão sobre os valores preditivos:

OS VALORES PREDITIVOS VARIAM DE ACORDO COM A PREVALÊNCIA DA DOENÇA NA POPULAÇÃO.

Figura 2

A dinâmica do processo diagnóstico



Isto significa que o desempenho de um teste varia de acordo com o ambiente onde ele é executado. Em outras palavras, ao se observar algo no exame físico, a possibilidade disso significar uma doença em particular, será diferente se a manobra for executada num ambiente de atenção primária ou num serviço especializado num nível terciário. Portanto:

DEVEMOS ESCOLHER TESTES DIAGNÓSTICOS QUE TENHAM BOM VALOR PREDITIVO PARA AS POPULAÇÕES COM AS QUAIS ESTAMOS TRABALHANDO.

A grande maioria dos sinais de exame físico foram descritos a partir de observações em ambiente hospitalar. Isto significa que, a despeito de boas sensibilidade e especificidade que possam ter, seu valor preditivo positivo num ambiente de clínica geral pode ser muito baixo. A moderna semiologia se preocupa com o ambiente onde o exame físico se dá⁵ (e, portanto, com a prevalência das doenças nesses ambientes e o consequente valor preditivo das manobras a serem executadas).

Outra forma de chegarmos ao valor da probabilidade pós-teste sem calcularmos os valores preditivos é através do cálculo do quociente: probabilidade de teste positivo em alguém com a doença / probabilidade de teste positivo em alguém sem a doença. Este quociente é conhecido por RAZÃO DE VEROSSIMILHANÇA (LIKELIHOOD RATIO) POSITIVA. No quadro 2 X 2, ainda na Figura 1, observe que: a probabilidade de um resultado positivo em alguém com a doença, $P(S/D)$, é calculada por $a / (a + c)$. Da mesma forma, a probabilidade de um resultado positivo em alguém sem a doença, $P(S/D')$, é calculada por $b / (b + d)$. Logo, o quociente entre estas duas probabilidades é a RV+.

Volte ao quadro 2 X 2 da Figura 1 e observe que: a probabilidade de um resultado negativo em alguém com a doença, $P(S'/D)$, é calculada por $c / (a + c)$. Da mesma forma, a probabilidade de um resultado negativo em alguém sem a doença, $P(S'/D')$, é calculada por $d / (b + d)$. Calcule agora o quociente: probabilidade de teste negativo em alguém com a doença / probabilidade de teste negativo em alguém sem a doença. Este quociente é conhecido por RAZÃO DE VEROSSIMILHANÇA (LIKELIHOOD RATIO) NEGATIVA.

Como as razões de verossimilhança são o quociente de probabilidades, elas não são probabilidades, mas um número absoluto.

Uma outra maneira de enxergar as razões de verossimilhança é compará-las à sensibilidade e à especificidade. Portanto, $RV+ = \text{sensibilidade} / (1 - \text{especificidade})$ e $RV- = (1 - \text{sensibilidade}) / \text{especificidade}$.

Mas como usar as razões de verossimilhança? Simples, as razões de verossimilhança são fatores usados para transformar probabilidade pré-teste em probabilidade pós-teste. Foi desenvolvido um nomograma para que, localizando a probabilidade pré-teste e a RV, e unindo-se os dois pontos por uma reta, a interseção desta reta com a coluna da direita dará a probabilidade pós-teste⁶.

Se, no exame físico, o achado do exame físico não fosse encontrado, trabalharíamos com o valor de sua RV-. Com o mesmo processo poderíamos calcular a probabilidade pós-teste.

Esta é uma grande utilidade da RV: calcular a probabilidade pós-teste a partir da probabilidade pré-teste, para um paciente individual.

Anteriormente havíamos comentado que sensibilidade e especificida-

de eram atributos limitados, por partirem de um conhecimento prévio de presença ou ausência de doença (exatamente o que se quer definir quando se pratica medicina). Se a RV é calculada a partir de sensibilidade e especificidade, ela não traz consigo os mesmos problemas desses conceitos? Não! E aí está a maior utilidade da RV. Por ser apenas um valor numérico absoluto, um fator que vai transformar probabilidade pré-teste em probabilidade pós-teste, a RV atende exatamente o que se quer saber na consulta médica: nesta pessoa, com tal probabilidade de doença (pré-teste), um achado de exame físico (com tal RV+) aumenta para quanto esta probabilidade (pós-teste)? E ainda, nesta pessoa, com tal probabilidade de doença (pré-teste), a ausência de um achado de exame físico (com tal RV-) diminui para quanto esta probabilidade (pós-teste)?

Numa prática ideal, ao atendermos uma pessoa, deveríamos saber sua probabilidade individual de uma doença em particular (nem sempre teremos este dado com precisão), precisaríamos conhecer as RV de todas as ferramentas que empregamos (e tal informação não está disponível para a maioria das ferramentas), para então calcularmos a probabilidade pós-teste (aplicando o nomograma, que nem sempre estará à mão). Tal prática não existe mas, é certo, que a literatura nos orienta cada vez mais quanto às probabilidades pré-teste de doenças e quanto às RV de diversas manobras (os trabalhos do grupo CARE, por exemplo⁷, e a série Exame Clínico Racional, do JAMA⁸). Quanto ao nomograma, que nem sempre estará à mão, existe uma alternativa: o conceito de CHANCE.

Trabalhamos, até agora, o conceito de probabilidade (pré e pós teste). No entanto, a matemática tem um outro conceito para prever eventos: é o conceito de CHANCE, que pode ser obtido assim:

Assim, se a probabilidade de uma doença é de 0,75 a chance desta doença será de 0,75 / 0,25, ou seja, igual a 3. Observe que chance é um número absoluto, que varia de 0 até + infinito.

Se trabalhamos com chances (em vez de probabilidades), não precisa-

CHANCE É A PROBABILIDADE DE QUE ALGO ACONTEÇA DIVIDIDO PELA PROBABILIDADE DE QUE NÃO ACONTEÇA, OU SEJA, $P(A) / P(A')$.

remos do nomograma na hora de utilizar as RV. Como já vimos, se tivermos uma determinada probabilidade pré-teste e uma RV+, elas podem ser integradas através do nomograma para o cálculo da probabilidade pós-teste. Ao trabalharmos com chances, tudo o que temos a fazer é multiplicar a chance pré-teste pela RV, e obteremos a chance pós-teste!

Se você quiser reverter um valor em chance para probabilidade, lembre-se que: Probabilidade = Chance / (Chance + 1).

O uso do conceito de chance permite novas definições para:

RAZÃO DE VEROSSIMILHANÇA (POSITIVA OU NEGATIVA): um fator que, ao ser multiplicado pela chance pré-teste de uma doença, fornece sua chance pós-teste. Fica claro que quando a RV é igual a 1, a manobra é inútil (pois não muda a chance pré-teste). Quanto mais a RV de uma manobra se afastar de 1, mais útil ela será.

Outra grande vantagem da utilização do conceito de RV é que, ao usarmos várias manobras, elas podem ser facilmente integradas para o cálculo da chance pós-teste:

Assim, identificamos os dois grandes atributos do conceito de razão de verossimilhança para as manobras de exame físico:

- Permite abordar a chance pré-teste de um paciente individual e transformá-la em chance pós-teste.
- Permite integrar os resultados de várias manobras do exame físico para cálculo da chance pós-teste.

(CHANCE PRÉ-TESTE) X (RV DA PRIMEIRA MANOBRA) X (RV DA SEGUNDA MANOBRA) X (RV DA TERCEIRA MANOBRA) X (RV DA MANOBRA DE ORDEM "N") = CHANCE PÓS-TESTE.

Tenha certeza de que sua compreensão deste assunto é fundamental para uma prática com a melhor evidência científica possível. Devemos ser muito criteriosos ao solicitar ou realizar um teste. Escolha e privilegie

aqueles cujos atributos sejam adequados ao que você quer, ou seja:

- Boa sensibilidade (a probabilidade de um resultado positivo num universo de pessoas com a doença). Cuidado: alta sensibilidade pode levar a alguns falso-positivos.
- Boa especificidade (a probabilidade de um resultado negativo num universo de pessoas sem a doença). Cuidado: alta especificidade pode levar a alguns falso-negativos.
- Bom valor preditivo (positivo ou negativo) para o contexto em que você está atendendo o paciente. Cuidado: o valor preditivo de uma manobra num ambiente de atenção primária nunca é igual ao valor preditivo da mesma manobra num ambiente especializado.
- Razão de verossimilhança (positiva e negativa) conhecida e diferente de 1.

Resumindo, as características pré-teste (sensibilidade e especificidade) são inerentes ao mesmo, sendo utilizadas para decidir qual teste devemos realizar. Já as características pós-teste (valores preditivos) são conhecidas também como características clínicas, dependentes da prevalência, sendo utilizadas para decidir o que fazer com o resultado obtido.

Resta lembrar que muitos testes, em especial as manobras de exame físico, podem sofrer variações inter-observador, o que leva à necessidade de medidas que indiquem a reprodutibilidade da manobra, como a estatística Kappa⁹. As manobras que dependem de inspeção apresentam melhor valor para a estatística Kappa que aquelas que dependem de palpação, percussão ou ausculta¹⁰. Uma discussão detalhada sobre medidas de concordância inter-observadores será apresentada neste espaço em outra oportunidade.



REFERÊNCIAS BIBLIOGRÁFICAS

- 1 - Jaeschke R, Guyatt GH, Sackett DL. Users' Guides to the Medical Literature. III. How to Use an Article about a Diagnostic Test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994;271:703-7.
- 2 - Coley CM, Barry MJ, Fleming C, Mulley AG. Early Detection of Prostate Cancer. Part I: Prior Probability and Effectiveness of Tests. *Annals of Internal Medicine* 1997;126:394-406.
- 3 - IMCI 1997. Integrated Management of Childhood Illness: a WHO/UNICEF initiative. Supplement no. 1 to volume 75 of the Bulletin of the World Health Organization, Geneva, Switzerland.
- 4 - Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. 2000. Evidence-Based Medicine – How to Practice and Teach EBM, 2nd edition. London: Churchill Livingstone.
- 5 - McGee S. 2001. Evidence-Based Physical Diagnosis. Philadelphia: Saunders.
- 6 - Friedland DJ et al 2001. Medicina Baseada em Evidências – Uma Estrutura para a Prática Clínica. Rio de Janeiro: Guanabara Koogan.
- 7 - Straus SE, McAlister FA, Sackett DL, Deeks JJ. *JAMA* 2000; 283:1853-1857.
- 8 - Sackett DL. The Rational Clinical Examination: a Primer on the Precision and Accuracy of the Clinical Examination. *JAMA* 1992; 267: 2638-44.
- 9 - Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977; 33:159-174.
- 10 - Kahigwa E, Schellenberg D, Armstrong Schellenberger J, Aponte JJ, Alonso PL. Inter-observer variation in the Assessment of Clinical Signs in Sick Tanzania Children. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2002; 96: 162-166.