

Sistemas Especialistas de Apoio ao Diagnóstico em Medicina. Relações com o Teorema de Bayes e com a Lógica do Raciocínio Diagnóstico

Pedro José Negreiros de Andrade

Fortaleza - CE

“Ninguém se iluda pensando que pode competir com a memória de um computador e muito mais que isso com a rapidez de utilização dos dados memorizados... É um mundo novo que desponta aceleradamente e que revolucionará toda estrutura de trabalho médico e, principalmente, a ética médica”¹.

Sistemas especialistas são o fruto da aplicação da chamada engenharia do conhecimento, uma das subespecialidades da inteligência artificial². Tais sistemas empregam técnicas simples de inteligência artificial para simular a atuação de peritos humanos. Uma das características que um sistema de inteligência artificial pode apresentar é a capacidade de adquirir conhecimentos, ou seja, modificar-se com o uso. Isso, em geral, não acontece com os chamados sistemas especialistas que apesar de apresentarem um desempenho, às vezes, comparável ao de especialistas humanos na resolução de problemas específicos, não são capazes, em geral, de aprender, ou seja, de apresentar comportamento realmente inteligente. O maior interesse por estes sistemas data do início dos anos 70, quando o paradigma da inteligência artificial sofreu uma importante mudança, muito bem expressa nas palavras de Goldstein³.

Poderia deduzir-se desta afirmativa que a eficácia de um sistema especialista depende mais da quantidade de conhecimentos neles depositados do que da capacidade de adquirir conhecimentos. O fato é que nenhum método genérico perfeito foi encontrado ainda para criar ambiente inteligente sem o suficiente conhecimento para raciocinar sobre ele. Um avanço importante e recente na área de aquisição de conhecimentos seria os sistemas baseados em redes neurais, inspirados em modelos biológicos, ou seja, nos fundamentos anatomofuncionais do cérebro^{4,5}. Mas, até que esta mudança no paradigma venha a dar frutos maiores, os sistemas especialistas continuarão a depender muito mais da qualidade da sua base de conhecimentos do que da complexidade dos seus algoritmos.

Dentro, pois, desta lógica de aplicação de bases de conhecimento, os sistemas especialistas de diagnóstico em medicina dividiram-se, na sua fase inicial, em dois tipos⁶: a) sistemas baseados em regras; b) sistemas baseados em reconhecimento de padrões.

A maioria dos esforços iniciais, no sentido de aplicar técnicas de inteligência artificial na resolução de problemas reais em medicina, fundamentou-se em sistemas baseados em regras. Tais programas são relativamente fáceis de criar, porque seu conhecimento é catalogado na forma de regras do tipo If...Then... utilizadas em forma de cadeia com o objetivo de atingir uma conclusão. Este tipo de programa facilmente consegue parecer inteligente, permitindo também o fácil uso das chamadas árvores de decisão, muito frequentes na medicina moderna. Em áreas de domínio limitado, tais programas têm se mostrado bastante efetivos. Um exemplo bastante conhecido é o sistema Mycin, desenvolvido na Universidade de Stanford para a seleção de antibióticos em pacientes com infecção severa⁶. Esse programa já foi amplamente testado, mostrando um desempenho semelhante ao de especialistas na área de doenças infecciosas. O problema desses sistemas baseados em regras é que em áreas mais complexas, como o diagnóstico em Medicina Interna, o domínio do conhecimento é de tal forma extenso que os torna de implementação extremamente difícil^{6,7}. Assim, sistemas desse tipo, designados pioneiramente para a área médica, encontraram uma aplicação bem maior em tarefas comerciais ou técnicas tipo “avaliação de linhas telefônicas”. Apesar da mais que evidente utilidade desses sistemas em programas médicos baseados em árvores de decisão, a principal limitação ao seu uso advém da dificuldade de aceitar condutas de um computador que ponham em risco vidas humanas⁶.

Em face da dificuldade de aplicar sistemas baseados em regras em áreas de domínio extenso como o diagnóstico em Medicina Interna, a solução pareceria estar em programas tipo reconhecimento de padrões (pattern matching systems). Ao nível mais elementar, tais programas funcionam da seguinte maneira⁷: 1) para cada possível doença (hipótese diagnóstica) determinam se um achado (sintoma, sinal ou exame complementar) apresentado pelo paciente pode aparecer nela; 2) estabelecem escores para cada doen-

Hospital Universitário Walter Cantídio da UFC
Correspondência: Pedro José Negreiros de Andrade - Rua Tibúrcio Cavalcante,
1445/802 - 60125-100 - Fortaleza, CE
Recebido para publicação em 17/12/98
Aceito em 16/6/99

ça de acordo com o número de achados que ela apresenta em comum com o paciente; 3) ordenam as doenças de acordo com os escores obtidos; 4) perguntam se achados pertencentes à base de dados da doença com maior pontuação e não listados no caso encontram-se presentes no paciente; 4) repetem a análise (1 e 2); 5) repetem o procedimento na(s) doença(s) seguintes (3 e 4).

Se um programa tipo reconhecimento de padrões for desenvolvido dentro desta forma simplista, ele se assemelhará à maneira como muitos iniciantes abordam o processo de diagnóstico. Em consequência falhará por não levar em conta os seguintes fatores: a) quanto fortemente a presença de um achado (sintoma, sinal ou exame complementar) lembra uma determinada doença?; b) quanto fortemente a ausência de um achado afasta a possibilidade de uma doença?; c) qual a prevalência de cada uma das hipóteses na população estudada?; d) quanto fortemente um achado presente no paciente e não presente na hipótese afasta o diagnóstico?; e) o doente apresenta apenas uma ou mais de uma doença? f) se houver mais de uma doença, estarão elas relacionadas?

Sabe-se que o médico experiente, dispondo de uma forte base de conhecimentos na cabeça, emprega com competência estas informações ao longo do processo diagnóstico, sendo esta uma das características que o separa dos iniciantes⁸⁻¹¹.

Por isso mesmo, os programas de computador mais conhecidos na área de diagnóstico diferencial, como o Internist¹², sua variante QMR¹³, o DX Plain, o Iliad além do programa Consultor, por nós desenvolvido, incorporam a um programa tipo reconhecimento de padrões aspectos de raciocínio probabilístico e (no caso do Internist) de encaideamento causal. Para que sistemas especialistas deste tipo sejam melhor entendidos é necessário que discutamos mais detalhadamente os fundamentos lógicos do processo de diagnóstico.

Lógica do processo de diagnóstico

“O avanço da tecnologia coloca o problema diagnóstico no primeiro plano das preocupações. O diagnóstico definitivo saiu da esfera Clínica e passou a depender de uma tecnologia médica de altos custos e potencialmente iatrogênica. Esta realidade obriga-nos a atuar com maior tirocínio, a utilizar métodos científicos na argumentação clínica, partindo da premissa que a iatrogenia e os gastos médicos diminuem na medida dos acertos na solicitação de exames complementares”¹⁴.

Princípios básicos

A palavra diagnóstico tem origem grega, significando discernir ou distinguir. Em termos médicos, foi definido por Mason como “uma série de procedimentos de ordem intelectual e operacional através dos quais se obtém uma resposta a um problema clínico”¹⁴. Neste sentido, há várias palavras que complementam a palavra diagnóstico modificando-lhe o sentido: diagnóstico clínico, diagnóstico anatomopatológico, diagnóstico radiológico, diagnóstico eletrocardio-

gráfico, etc. Já a expressão diagnóstico diferencial foi definido por Harvey¹¹ como a “arte de distinguir uma doença de outras, estabelecendo uma ou mais causas bem definidas para explicar as alterações apresentadas pelo paciente”. O clássico instrumento de ensino desta forma de diagnóstico tem sido as sessões clinicopatológicas em que casos particularmente difíceis são submetidos a internistas ou subespecialistas para efeito de ensino. Na prática clínica, assim como nas referidas sessões, a seqüência pela qual é feito o diagnóstico diferencial pode ser resumida nas seguintes etapas¹¹: 1) colheita dos dados; 1a) história clínica; 1b) exame físico; 1c) exames laboratoriais; 2) análise dos dados; 2a) Avaliação crítica dos dados colhidos; 2b) listagem dos achados em ordem de importância; 2c) seleção de um ou, preferencialmente, dois a três achados centrais; 2d) listagem das doenças em que esses achados centrais se encontram; 2e) busca do diagnóstico final pela seleção da doença que melhor explica os achados do paciente; 2f) revisão de todos os dados positivos, a fim de não deixar nenhum achado considerado importante sem um diagnóstico que o explique.

A seqüência citada não é necessariamente rígida. A análise de um caso usualmente começa durante o próprio processo de colheita dos dados, quando o médico utiliza a sua experiência para detalhar os achados obtidos. Diante da queixa de dispnéia, um cardiologista procurará defini-la como “de esforço ou de repouso”, na tentativa de fazer o diagnóstico de insuficiência cardíaca. Procurará a partir daí, no exame físico, sinais de congestão sistêmica (turgência jugular, estertores de base) ou de sobrecarga do coração (desvio do ápex, impulsão de meso, ritmo de galope), para confirmar a hipótese formulada. Diante de um paciente com febre prolongada que apresenta um sopro mitral, o médico experiente procurará sinais de fenômenos embólicos, palidez, esplenomegalia ou baquetamento de dedos, para confirmar a suspeita de endocardite infecciosa (EI). Estes dois exemplos mostram como a análise dos dados pode começar precocemente, aperfeiçoando a obtenção da história e do exame físico.

Terminada a colheita dos dados, o médico os ordena por importância e inicia-se o processo de diagnóstico diferencial dentro da seqüência proposta anteriormente. Alguns médicos procuram muitas vezes caracterizar combinações de achados como síndromes no intuito de simplificar o processo de diagnóstico diferencial. Expressões como síndrome consumptiva, síndrome diarreica, síndrome de hipertensão pulmonar, síndrome congestiva, síndrome infecciosa, síndrome de derrame pleural e outras são freqüentemente usadas. Representam na realidade formas de expressar em poucas palavras a combinação de sintomas, sinais ou achados laboratoriais enquadrados como de maior relevância.

Apesar do extremo didatismo desta abordagem, alguns problemas podem advir durante o seu uso. O primeiro é a escolha equivocada dos achados centrais. Exige alguma experiência a definição de achados que, pela sua maior especificidade, mereça estar no centro do processo de diagnóstico. A escolha equivocada de um achado central pode tornar a lista de possibilidades levantadas excessivamente longa, dificultando a elaboração do diagnósti-

co. Além disso, a omissão de um achado relevante da lista, é um erro comum, de grave conseqüência para o processo de diagnóstico¹⁵. Outro erro freqüente, mesmo entre os mais experientes, é a conclusão prematura de um diagnóstico definitivo ou síndrome, sem que haja dados indiscutíveis para estabelecê-lo¹⁵.

Para efeito de se entender melhor como é feita a análise diagnóstica, como os erros acima podem ser evitados e como podemos simular o raciocínio médico em um programa de computador é interessante que façamos uma revisão crítica dos três tipos de raciocínio diagnóstico habitualmente utilizados.

Tipos de raciocínio diagnóstico

Segundo Sox¹⁶ existem três tipos básicos de raciocínio diagnóstico: 1) o raciocínio fisiopatológico; 2) o raciocínio por reconhecimento de padrões; 3) o raciocínio probabilístico;

O raciocínio fisiopatológico é o mais difícil de simular em um programa de computador. Quando um paciente descreve história de febre há seis semanas, seguida três semanas depois por dor no quadrante superior direito de crescente severidade, o médico começa a imaginar uma massa que cresce e estabelece pressão em estruturas sensíveis à dor. Começa então a pensar em um abscesso hepático ou em um nódulo maligno com centro necrótico, causando resposta inflamatória.

Já o raciocínio por reconhecimento de padrões é o mais freqüentemente usado tanto por estudantes de medicina quanto por especialistas. É, além disso, o mais fácil de simular em programas de computador. Certos achados ocorrem juntos e sua combinação leva o médico a formular hipóteses. A partir daí ele confronta os dados do paciente com os da doença (hipótese) para verificar em que grau os mesmos se encaixam. Uma das falhas desse tipo de raciocínio consiste na incapacidade de reconhecer doenças quando elas se apresentam de forma diferente das suas manifestações clássicas^{7,14}. As outras são semelhantes às comentadas em relação aos programas de computador que utilizam regras simples de reconhecimento de padrões: a) não levar em conta quanto fortemente a presença de um achado evoca uma doença; b) não levar em conta quanto fortemente ausência de um achado esperado afasta uma determinada hipótese; c) não levar em conta a prevalência da doença; d) não levar em conta a relevância de um achado, ou seja, quanto fortemente o fato dele não ser explicado por uma doença a elimina como hipótese.

O raciocínio probabilístico baseia-se no fato de que os médicos convivem com a incerteza em um grau comparável ao de bem poucos profissionais. Probabilidade, no caso, seria apenas uma maneira de medir essa incerteza. Quando dois médicos dizem, por exemplo, que um doente provavelmente tem embolia pulmonar, um pode estar pensando em uma probabilidade de 30% e o outro de 90%. No manuseio dessa incerteza o médico recorre muitas vezes a exames laboratoriais que a alteram, mas não a eliminam. Na interpretação probabilística desses testes, os médicos experientes usam intuitivamente o chamado Teorema de Bayes^{16,17}, o qual re-

laciona sensibilidade, especificidade, prevalência e probabilidade pós-teste. Obviamente, a aplicação de tais conceitos deve idealmente referir-se não apenas a exames laboratoriais, mas também a achados de história ou exame físico.

Além destas três formas clássicas de raciocínio, duas outras formas de fazer diagnóstico merecem ser citadas: o raciocínio intuitivo e o diagnóstico por aforismas. O raciocínio intuitivo é provavelmente uma mistura inconsciente dos três tipos de raciocínio já citados, além de incorporar outros elementos subjetivos difíceis de definir. O diagnóstico por aforismas é mais especificamente a mistura do raciocínio por reconhecimento de padrões com o raciocínio probabilístico. Tais aforismas constituem-se em uma série de regras baseadas na experiência de clínicos respeitáveis. É conhecido, por exemplo, o sinal de Courvoisier ("icterícia associada à vesícula palpável e indolor significa obstrução carcinomatosa do colédoco"). Outro exemplo é a expressão segundo a qual "icterícia associada palidez e esplenomegalia faz pensar em anemia hemolítica". Tais aforismas, embora memoráveis e abrangendo uma infinidade de combinação de achados clínicos, constituem, na verdade, simplificação de problemas complexos. Além disso, sua memorização pouco contribui ao ensino da lógica do processo de diagnóstico. É interessante lembrar que muita ênfase tem sido dada, em anos recentes, às chamadas regras de predição clínica. As mesmas têm por objetivo reduzir o grau de incerteza no diagnóstico, determinando como usar achados clínicos para fazer predições. Uma análise mais cuidadosa mostra que essas regras de predição clínica são a reedição moderna dos antigos aforismas¹⁸. A diferença é que tais regras estabelecidas através da análise de centenas de pacientes, empregam cálculos de probabilidade e técnicas matemáticas sofisticadas. Por isso, elas têm sido motivo de grande interesse nos últimos anos, em particular pela sua potencial utilização em sistemas de apoio à decisão médica^{18,19}.

Dada a importância do raciocínio probabilístico para os sistemas de apoio ao diagnóstico médico é sobre ele que discorreremos com maior detalhe nas próximas páginas.

Racícioínio probabilístico em medicina

"Quando você pode medir um fenômeno sobre o qual está falando e expressá-lo em números, você sabe alguma coisa sobre ele. Mas, quando você não pode medi-lo, quando você não pode expressá-lo em números, o seu conhecimento é vago e insatisfatório. Pode ser o começo do conhecimento, mas você progrediu muito pouco em direção ao estágio de ciência" (Lord Kelvin (1824-1907)⁶.

"As aparências para mente são de quatro tipos: as coisas são o que parecem ser ou não são e nem parecem ser; ou são e não parecem ser; ou não são e mesmo assim parecem ser. Identificar corretamente todos esses casos é a tarefa do homem sábio" Epictetus (Século II DC)²⁰.

Conceitos básicos

O racióínio médico é sabidamente mais baseado em probabilidades do que em certezas. Sendo o racióínio

probabilístico tão importante no processo de diagnóstico, na definição de regras de predição clínica¹⁸ e mesmo na conduta em casos complexos, é espantoso que quase nada sobre o mesmo seja ensinado no currículo médico.

Para entendermos melhor os aspectos quantitativos da lógica diagnóstica, ou mais especificamente o raciocínio probabilístico e o teorema de Bayes, é necessária uma breve introdução aos conceitos de sensibilidade e especificidade a partir da definição de probabilidade condicional. Dentro deste conceito para dois eventos quaisquer A e B, sendo $P(A) > 0$, definimos a probabilidade condicional de B dado A, como sendo:

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

Consideremos agora a situação abaixo, sintetizada numa tabela 2x2, onde n_1 pessoas sabidamente portadoras de uma doença e n_2 pessoas sabidamente livres da doença foram submetidas a um determinado teste clínico:

Teste/Doença	Positivo	Negativo	Total
(D) Presente	a	b	n_1
(d) Ausente	c	d	n_2
Total	a+c	b+d	N

Temos nessa situação os seguintes resultados possíveis: teste positivo e doença presente: VP (verdadeiro positivo) = a casos; teste negativo e doença ausente: VN (verdadeiro negativo) = d casos; teste negativo e doença presente: FN (falso negativo) = b casos; teste positivo e doença ausente: FP (falso positivo) = c casos.

Ao perguntarmos: se a doença está presente, qual a probabilidade do teste ser positivo, ou melhor, $P(T/D)$? A resposta é dada da seguinte maneira:

$$P(T/D) = \frac{P(T \cap d)}{P(d)} = \frac{a/N}{n_1/N} = \frac{a}{n_1} = \frac{a}{a+b} = \frac{VP}{VP+FN}$$

Esta é a chamada a sensibilidade estimada do teste.

Ao perguntarmos: Se a doença está ausente, qual a probabilidade do teste ser negativo, ou melhor, $P(T/d)$? A resposta no caso é dada da seguinte maneira:

$$P(T/d) = \frac{P(T \cap d)}{P(d)} = \frac{d/N}{n_2/N} = \frac{d}{n_2} = \frac{VN}{VN+FP}$$

Esta é a chamada especificidade estimada do teste.

Estes conceitos de sensibilidade e especificidade são bastante conhecidos, embora frequentemente em um nível superficial ou intuitivo. Começaram a ser usados mais intensamente após a II Guerra Mundial a partir da análise da eficiência de achados radiológicos no diagnóstico de tuberculose²¹.

Um teste altamente sensível raramente deixará de ser positivo em pessoas que realmente tenham a doença. Como exemplo de teste altamente sensível teríamos a presença de febre em pacientes com EI, ou o FAN positivo em paciente

com lúpus eritematoso sistêmico (LES).

Um teste altamente específico raramente deixará de ser negativo em pessoas que não tenham a doença. Como exemplo de teste de alta especificidade teríamos a presença de uma grande vegetação no ecocardiograma para efeito de diagnóstico de EI.

Em resumo, testes altamente sensíveis são úteis para afastar doenças quando negativos e testes altamente específicos são úteis para confirmar doenças quando positivos. Sempre que alterarmos o critério de positividade de um teste no sentido de torná-lo mais específico, estaremos, por outro lado, diminuindo a sua sensibilidade. O oposto é igualmente verdadeiro.

O conceito de prevalência refere-se à frequência da doença na população estudada. Já o conceito de valor preditivo é mais complexo, referindo-se à probabilidade que um indivíduo tem de ter a doença após o resultado alterado de um teste. Só poderemos saber o valor preditivo de um teste se considerarmos conjuntamente informações sobre prevalência, sensibilidade e especificidade desse teste, expressas matematicamente através da fórmula de Bayes.

Teorema de Bayes

Embora os conceitos de sensibilidade e especificidade sejam bastante difundidos, o mesmo não acontece com os conceitos de probabilidade pré-teste e pós-teste e, particularmente, com o teorema de Bayes, que, apesar de colocado nos capítulos introdutórios de livros-texto clássicos⁸ raramente é lido ou ensinado aos estudantes de medicina.

Recentemente, testou-se a capacidade dos estudantes de medicina de interpretar quantitativamente resultados de exames laboratoriais através da questão em uma prova de residência médica: "um teste tem uma sensibilidade de 70% e uma especificidade de 90%. Aplicado em um paciente pertencente a uma população em que a prevalência da doença é de 1% o exame resulta positivo. Qual a probabilidade pós-teste deste paciente ter a doença? Não foi nenhuma surpresa verificar que a maioria dos estudantes situou esta probabilidade entre 70 e 90%, em vez de escolherem a resposta correta, que era 6,6%. Os únicos que acertaram foram os que, tendo rodado pelo Serviço de Cardiologia, nele haviam recebido alguma informação sobre o teorema de Bayes, através da sua aplicação à ergometria²².

Thomas Bayes (1702-1761) foi um filósofo, matemático e religioso inglês, sendo considerado um dos pais do cálculo de probabilidade. Análise bayesiana seria, pois, uma teoria de decisão estatística para cálculo de probabilidade de uma proposição, com base na probabilidade original e nas novas relevâncias. Expresso em termos clínicos é o conceito segundo o qual o valor preditivo de um teste ou achado clínico depende não apenas da sua sensibilidade e especificidade, mas também da probabilidade prévia (ou seja, da prevalência da doença na população estudada).

É de importância histórica ressaltar que o trabalho pioneiro sobre a aplicação do teorema de Bayes em Medicina foi publicado em 1959 por um especialista em informática,

Lested, e um médico radiologista, Lusted, este último previamente envolvido nos trabalhos que deram origem aos conceitos médicos de sensibilidade e especificidade^{20,23}.

A fórmula original de Bayes, usada pelos estatísticos, seria a seguinte:

$$P(C_i/A) = \frac{P(C_i) \cdot P(A/C_i)}{\sum_{j=1}^n P(C_j) \cdot P(A/C_j)}, i=1, \dots, n$$

Uma versão mais simples, aplicável à medicina é a seguinte:

$$VP+ = \frac{Pv \cdot S}{Pv \cdot S + (1-Pv) \cdot Pv(1-E)}$$

$$E \text{ analogamente: } Pv \cdot (1-S)$$

$$VP- = \frac{Pv(1-S)}{Pv(1-S) + (1-Pv) \cdot E}$$

Onde: VP + = valor preditivo (probabilidade pós-teste) de teste positivo; VP - = valor preditivo (probabilidade pós-teste) de teste negativo; Pv = prevalência da doença (ou probabilidade pré-teste); S = sensibilidade do teste ou achado clínico; E = especificidade do teste ou achado clínico.

Um aspecto inerente à fórmula bayesiana é a possibilidade de ser aplicada seqüencialmente. Em outras palavras, após aplicar ao resultado de um teste à fórmula bayesiana, o valor preditivo obtido passa a ser a nova probabilidade de ocorrência da doença. A partir dela novas probabilidades deverão ser calculadas face a outros testes ou achados. O único pressuposto para essa aplicação seqüencial é que os testes sejam independentes entre si, ou seja, que o resultado de um não interfira no do outro. Exemplos de testes não independentes seriam a depressão do ST na prova de esforço e a presença da mesma depressão de ST no Holter. Outro exemplo seria a presença de palidez conjuntival ao exame físico e o achado de hemoglobina diminuída no hemograma.

Até onde pudemos investigar a primeira tentativa prática de aplicar o teorema de Bayes à medicina foi na área de cardiologia, através dos trabalhos de Warner e cols.^{24,25}, que o utilizaram como auxiliar ao diagnóstico de cardiopatias congênitas. Mas, o trabalho de maior repercussão na nossa especialidade foi publicado por Rifkin e Hood²⁶ que conseguiram explicar através dele as curiosas diferenças de credibilidade do eletrocardiograma de esforço quando aplicados em homens ou mulheres, assim como em pacientes sintomáticos ou assintomáticos. Seguiram-se outros trabalhos demonstrando do ponto de vista prático a força do emprego da lógica bayesiana na interpretação da ergometria^{17,27}, levando por consequência a sua grande divulgação entre cardiologistas, logo a seguir estendida a todas as áreas da medicina interna.

O que esses trabalhos mostraram, se tomarmos um teste com sensibilidade de 60% (S=0,6) e uma especificidade de

90% (E=0,9) e o aplicarmos em uma população com prevalência da doença de 1% (PV=0,01), é que a probabilidade de um paciente com teste positivo apresentar doença seria (de acordo com a fórmula de Bayes) a seguinte:

$$VP+ = \frac{0,01 \times 0,6}{(0,01 \times 0,6) + (0,99 \times 0,1)} = 0,056 \text{ (ou } 5,6\%)$$

Esta situação reflete a probabilidade de uma mulher de 43 anos assintomática (ou uma jovem de 33 anos com dor torácica não cardíaca) apresentar doença coronariana face a um teste de esforço positivo^{17,26-28}. A lição básica do emprego da fórmula no caso é que constitui um desperdício de recursos solicitar uma prova de esforço eletrocardiográfica nesta situação, pois dificilmente a mesma alterará a conduta a ser tomada²⁹. Se a probabilidade pré-teste for no entanto de 50% (caso de um homem de 48 anos com angina atípica) o teste de esforço positivo alterará a probabilidade de doença coronariana para 85% se positivo e para 28% se negativo.

Uma outra forma de calcular a probabilidade pós-teste é através do uso de diagramas. Tomemos, por exemplo, uma população com probabilidade de apresentar uma doença de 5%, submetendo-se a um teste que apresenta sensibilidade de 70% e especificidade de 90%. O caso guarda uma certa analogia com o uso de uma urografia excretora em um paciente com uma probabilidade um pouco maior que a habitual de apresentar estenose de artéria renal. Em caso, portanto, de urografia alterada, qual seria a probabilidade de estenose da artéria renal em um paciente com probabilidade prévia de 5%? Para resolver esse problema poderemos construir o seguinte diagrama:

1000 pacientes			
950 sadios		(PV=5%) =>	50 com estenose
95	855	<=(E=90%) S=70%) =>	35 15
FP	VN		(VP) (FN)
VP			
Prob. pós teste + =			
VP + FP			
		35	15
Prob. pós teste + = = 26,9%			
		35+95	130

A lição que se tirou do emprego da fórmula bayesiana nesta situação foi que a prática de solicitar urografia excretora na avaliação de possível estenose de artéria renal, uma regra no passado, deveria ser suspensa, uma vez que os custos e a morbidade do teste seriam maiores que o ganho a ser obtido com ele³⁰. A consequência é que outros testes como o renograma com captopril passaram então a ser utilizados no seu lugar, e ainda assim em pacientes com suspeita significativa de hipertensão renovascular.

Dos exemplos citados, parece claro a grande importância da probabilidade prévia, ou seja, da prevalência da doença na interpretação de um teste ou achado clínico. Fica claro também que testes com sensibilidade e especificidade

limitados alteram significativamente a probabilidade diagnóstica apenas quando aplicados em indivíduos com probabilidade intermediária de ter uma doença¹⁶. Nas situações em que a probabilidade prévia é muito alta ou muito baixa, testes de sensibilidade e especificidade limitados não alteram o diagnóstico. A exceção seria testes de alta especificidade para indivíduos com baixa probabilidade prévia ou testes de alta sensibilidade para indivíduos com alta probabilidade prévia.

Outras formas de utilização do teorema de Bayes têm sido propostas como o uso de nomogramas³¹ e a tabela 2x2 dos estatísticos e epidemiologistas¹⁵, por alguns considerados de mais fácil aplicação³².

Apesar da só recente divulgação do teorema de Bayes entre internistas e cardiologistas, particularmente em função da interpretação de testes ergométricos^{22,32}, é importante enfatizar que os médicos experientes sempre o empregaram, ainda que intuitivamente, em todas as etapas do processo diagnóstico. Quando os antigos clínicos afirmavam ser a clínica soberana, face aos exames laboratoriais, estavam, sem o saber, raciocinando bayesianamente. Quando afirmavam ser melhor pensar em doenças comuns, ainda que com manifestações atípicas, em lugar de manifestações típicas de raridades, estavam, intuitivamente, aplicando o teorema de Bayes. Usar a presença de achados de alta especificidade para confirmar diagnósticos e achados de alta sensibilidade para afastá-los, é aplicar, ainda que não formalmente, a lógica bayesiana. Em uma época de proliferação tecnológica, em que exames laboratoriais são solicitados desnecessariamente e mal interpretados, dar relevância à história do paciente e a dados epidemiológicos é raciocinar bayesianamente²².

Diagnósticos baseados em escores, suas relações com o teorema de Bayes e com sistemas especialistas

O teorema de Bayes pela sua perfeição lógica, exatidão matemática e possibilidade de aplicação sequencial, pareceria, pois, o instrumento ideal para estabelecer uma ponte entre o raciocínio médico e a informática. Apresenta, no entanto, algumas limitações, particularmente no que toca ao seu uso em sistemas especialistas de diagnóstico diferencial: a) dificuldade de definir a probabilidade prévia em situações diferentes. Programas designados para os Estados Unidos da América teriam que ser modificados em regiões como o interior do Ceará. Além disso, a prevalência de certas doenças varia também com a instituição, sendo, às vezes, baixíssima em um ambulatório geral, não tão baixa na enfermaria de um hospital terciário e alta em sessões anatomoclínicas; b) dificuldade de aplicar sequencialmente a fórmula Bayesiana em múltiplos testes e, particularmente, em circunstâncias em que os testes não sejam independentes^{16,33}; c) dificuldade de entendimento da fórmula por parte de especialistas médicos encarregados da elaboração da base de dados.

Estas dificuldades seriam a provável explicação para o fato dos programas baseados em uma aplicação rígida do teorema de Bayes funcionarem melhor em áreas mais limitadas do conhecimento médico. Uma forma de minimizá-las

seria o uso do teorema de Bayes convertido em um sistema de escores, conforme proposta de Rembold e Watson³⁴.

Segundo esses autores, convertendo a probabilidade em chances e linearizando a equação pelo uso do logaritmo natural correspondente, teríamos a conversão dos valores fracionários em números inteiros, que seriam chamados de peso ou escores, mais fáceis de serem entendidos e utilizados. A fórmula proposta por Rembold e Watson seria a seguinte:

$$\text{In chance pós-teste (+)} = \text{In chance pré-teste} + \text{In} \frac{S}{1-E}$$

$$\text{In chance pós-teste (-)} = \text{In chance pré-teste} + \text{In} \frac{1-S}{E}$$

(Onde In = logaritmo natural, S = sensibilidade e E = especificidade) chance doença = $\text{prob } D / \text{pro } (1-D)$ D =

Mesmo sem entrar no mérito matemático desta proposição ela se torna atraente por sugerir que ao atribuímos escores positivos para testes ou achados clínicos em função principalmente da especificidade (se presentes ou positivos) e escores negativos em função da sensibilidade (se ausentes ou negativos) estar-se-ia, de certa forma, aplicando o teorema de Bayes.

Outra razão para ver com simpatia propostas deste tipo é que a idéia de fazer diagnósticos através de escores é de há muito familiar aos clínicos, sendo enorme a literatura sobre o assunto. Na cardiologia, talvez tenha sido Jones³⁵ o primeiro a propor, com aceitação, uma forma de diagnóstico através de alguma coisa análoga a escores. Os chamados critérios de Jones (dois sinais maiores ou um maior e dois menores como condição para o diagnóstico de febre reumática) são facilmente transformáveis em escores. Basta no caso atribuir o valor 5 para os sinais maiores e 2 para os sinais menores e considerar 9 como a somatória de pontos capaz de tornar provável o diagnóstico de febre reumática. Os sinais maiores são obviamente os de alta especificidade; os menores, os menos específicos. Ainda dentro da cardiologia, além da própria febre reumática, doenças tão diversas quanto insuficiência coronariana²⁹, insuficiência cardíaca³⁶, prolapso de válvula mitral³⁷ e derrame pericárdico já tiveram propostos critérios de diagnósticos através da somatória de escores, ou algo semelhante (combinação de sinais maiores e menores). Na reumatologia são bem conhecidos os critérios do Colégio Americano de Reumatologia para diagnóstico de artrite reumatóide³⁹ e LES⁴⁰. Na endocrinologia é interessante lembrar o critério de pontos de Wayne para o diagnóstico de hipertireoidismo⁴¹. A lista potencial é enorme, sugerindo que o diagnóstico através de escores apresenta uma base lógica muito forte (no fundo de natureza bayesiana, ainda que sem o rigor matemático do emprego da fórmula), servindo de instrumento de utilização extremamente simples em softwares de diagnóstico diferencial.

Por todas estas razões programas como o Internist-I e sua variante QMR tentam fazer diagnósticos através de escores que são atribuídos à relação entre os achados clínicos e as doenças. Um destes escores refere-se à evoking strength, que, segundo os autores, representaria grosseiramente o conceito de valor preditivo positivo, sendo fortemente in-

fluenciado pela especificidade e pela prevalência. O segundo refere-se à frequência do achado na doença, a qual guardaria relação com a sensibilidade do achado. O programa Internist foi testado em artigo clássico¹², mostrando desempenho inferior ao dos responsáveis pela discussão dos casos em sessões anatomoclínicas do New England Journal of Medicine, mas comparável ao dos responsáveis pelos pacientes (diagnóstico clínico). Em avaliação posterior, a sua variante QMR⁴² teve desempenho considerado como satisfatório em termos de auxílio ao diagnóstico em casos reais em hospitais universitários. Além desses o Dx-Plain⁴³, da Universidade de Harvard e o Iliad⁴⁴ inspiram-se em princípios semelhantes, sendo este último rigorosamente bayesiano. O programa Consultor, desenvolvido inicialmente na área de cardiologia⁴⁵ e depois estendido a toda medicina interna⁴⁶ apresenta importantes analogias com os referidos sistemas. Consultor foi avaliado inicialmente através da submissão ao mesmo de provas de residência médica⁴⁶ e, mais recentemente, através de sessões anatomoclínicas do Hospital Universitário Walter Cantídio e do New England Journal of Medicine (neste caso apenas casos cardiológicos). O sistema teve, à luz destas avaliações, um desempenho bastante aceitável, em medicina interna de maneira geral e em cardiologia em particular^{1,47}. Os quatro sistemas norte-americanos avaliados comparativamente apresentaram desempenho também aceitável e comparativamente similar^{48,49}. Algumas das suas limitações, no entanto, levam os especialistas na área a concluir que deveriam ser utilizados apenas por médicos capazes de identificar e utilizar as informações relevantes e ignorar as irrelevantes por eles fornecidas.

Uma série de questionamentos poderia ser levantada em relação aos sistemas especialistas de apoio ao diagnóstico médico. Poderão tais programas embotar o desenvolvimento do raciocínio clínico? Deverá ou não ser permitida a sua utilização por leigos ou paramédicos? Poderão sistemas como este, no futuro, substituir os clínicos? Em relação à 1ª pergunta a nossa experiência é de que a utilização de sistemas especialistas de apoio ao diagnóstico por estudantes

de medicina leva-os a dominar mais intensamente conceitos como sensibilidade, especificidade e prevalência e suas relações com o teorema de Bayes, melhorando, em vez de embotar o raciocínio clínico. Obviamente, uma avaliação mais definitiva exigiria que se comparasse o desempenho diagnóstico de estudantes treinados no uso de tais sistemas com o de estudantes não expostos aos referidos programas. A 2ª pergunta é mais difícil de responder. Embora nenhum conhecimento deva, em princípio, ser privilégio de uma corporação, a tendência atual dos responsáveis por sistemas especialistas é não liberá-los para a população leiga. Isso se deve fundamentalmente a duas razões: a mais importante é que não se pode confiar ainda plenamente no valor preditivo dos diagnósticos desses sistemas, o que poderia levar prejuízo a pacientes se usado por indivíduos não adequadamente treinados; a outra é que o desempenho de sistemas especialistas exige obtenção e interpretação acuradas dos achados clínicos, que não são encontradas no momento fora da profissão médica.

Parece, portanto, evidente que a liberação de sistemas especialistas como o Consultor, o QMR, o Iliad e o Dxpain para uso por pessoal leigo poderá trazer mais riscos que benefícios para os pacientes, uma vez que o algoritmo desses sistemas necessita ser aprimorado, pois um longo caminho continua a separá-los de um comportamento realmente similar ao raciocínio médico⁴⁹⁻⁵². Tal caminho passa, particularmente em relação ao Consultor²⁸, pela incorporação de um grau maior de "raciocínio" fisiopatológico à sua lógica e pelo estabelecimento de relações de causa e efeito entre doenças. Possivelmente, isso só será atingido através da substituição das atuais linguagens por outras mais adequadas à inteligência artificial. Mas, ainda que o desempenho destes sistemas melhore, tornando-os realmente inteligentes e independentes, dificilmente poder-se-á aceitar a possibilidade de virem a substituir os médicos, devido ao lado humano da profissão, da necessidade de colheita minuciosa dos dados e da capacidade de compaixão pelos semelhantes, que manterão sempre um lugar para o médico no cuidado dos pacientes.

Referências

1. Ribeiro JC. Aula da Saudade. Rev Med Univ Fed Ceará 1976; 15: 67-70.
2. Genaro S. Sistemas Especialistas: O Conhecimento Especializado. Rio de Janeiro - São Paulo: Livros Técnicos e Científicos Editora, 1986.
3. Goldstein J, Papert S. Cogn Sci. 1977; 1: 84. Apud Duda RO, Sotliffe EH. Expert systems research. Science 1983; 220: 261-20.
4. Baxt WB. Use of an artificial neural network for the diagnosis of myocardial infarction. Ann Intern Med 1991; 115: 843-8.
5. Ortiz J, Sabbatini RM, Gheffer CGM, Silva CES. Uso de redes neurais artificiais na avaliação da sobrevida na insuficiência cardíaca. Arq Bras Cardiol 1995; 64: 87-90.
6. Duda RO, Sotliffe EH. Expert system research. Science 1983; 220: 261-8.
7. Slozovitz P, Patil R, Schwartz WB. Artificial intelligence in medical diagnosis. Ann Intern Med 1988; 108: 80-7.
8. Pauker SG. Clinical decision making. In: Wyngarden JB, Smith Jr LH, Bennett JC (Ed). Cecil's Textbook of Internal Medicine. Philadelphia: WB Saunders, 1992: 68-73.
9. Braunwald E. Introduction to clinical medicine. In: Braunwald et al (Ed). Harrison's Principles and practice of Internal Medicine. New York: McGraw-Hill, 1987: 1-5.
10. Johns RJ, Hazzard WR. Clinical information and problem solving. In: Harvey AM, et al (Ed). The Principles and Practice of Medicine. Norwalk-Connecticut: Appleton Century Crafts, 1984: 5-27.
11. Harvey AM, Bordley J. Differential Diagnosis - Abridgement of the 2ª Ed. Philadelphia: WB Saunders, 1972: 1-18.
12. Miller RA, Pople Jr, HE, Myers JD. Internist I, an experimental computer based diagnostic consultant in general internal medicine. N Eng J Med 1982; 307: 468-75.
13. Miller RA, McNeil MA, Chaliner SM, Masarie FE, Myers JDP. The internist I/quick medical reference project- Status report. West J Med 1986; 145: 816-22.
14. Rodrigues PMM. Lógica diagnóstica. Ceará Médico 1981; 3: 71-2.
15. Voytovich AE, Ryppey RM, Sufreddem AS. Premature conclusions in diagnostic reasoning. J Med Educ 1985; 60: 302-7.
16. Sox Jr HC. Medical decision making. In: Baronness JA, Carpenter G, Harvey AM. Differential Diagnosis. Philadelphia: Lea & Febiger, 1994: 9-22.
17. Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary artery disease. N Eng J Med 1979; 300: 1350-8.
18. Wasson JH, Sox HC, Neff RK. Clinical prediction rules: applications and methodological standards. N Eng J Med 1980; 392: 1109-17.
24. Warner HR, Toronto F, Veasy LG. Experience with Bayes's theorem for computer diagnosis of congenital heart disease. Ann NY AcadSci 1964; 115: 558-67.
25. Warner HR, Toronto AK, Veasy LG, Stephenson RA. Mathematical approach to me-

- dical diagnosis : application to congenital heart disease. JAMA 1964; 177: 177-83.
26. Rifkin RD, Hood Jr WB. Bayesian analysis of the electrocardiogram exercise testing. N Eng J Med 1977; 297: 681-5.
 27. Wegner DA, Ryan TJ, McCabe CH. Exercise stress testing: correlations between history of angina, ST response and prevalence of coronary artery disease in the coronary artery surgery study (CASS). N Eng J Med 1979; 30: 230-5.
 28. Andrade PJN. Apresentação e avaliação de um programa de computador de auxílio ao diagnóstico médico. Tese de Livre-Docência-UFC, 1995.
 29. Gibbons RJ, Balady GJ, Beasley JW, et al. Guidelines for exercise testing. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Exercise Testing) J Am Coll Cardiol 1997; 30: 260-315.
 30. Cheilín MD, Solokox M, McIlroy MB. Clinical Cardiology. Lange Medical Publishers 1994: 280.
 31. Fagen TJ. Nomogram for Bayes's theorem. N Eng J Med 1975; 243: 57.
 32. Brito AHX. Análise bayesiana do teste ergométrico. Arq Bras Cardiol 1991; 56: 97-103.
 33. Sox HL. Probability theory in the use of diagnostic tests. Ann Intern Med 1986; 104: 60-6.
 34. Rembold C, Watson DW. Pos-test probability by weights: a simple form of Bayes's theorem. Ann Intern Med 1988; 108: 115-20.
 34. Jones TD. The diagnosis of rheumatic fever. JAMA 1944; 126: 481-2.
 35. McKee PA, Castelli WB, McNamara PM. The natural history of congestive heart failure: the Framingham study. N Eng J Med 1971; 285: 1441-6.
 37. Perloff JK, Child JS, Edward JE. New guidelines for the clinical diagnosis of mitral valve prolapse. Am J Cardiol 1986; 57: 1124-8.
 38. Von Reyv CF, Levy DS, Arbeit RD, Friedlan G, Crompacks CS. Infective endocarditis: an analysis based on strict case definition. Ann Intern Med 1981; 94: 505-18.
 39. Arnett FC, Edworthy SM, Chlock DA. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1987; 31: 315-20.
 40. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982; 25: 1271-8.
 41. Wayne EJ. Clinical and metabolic studies in thyroid disease. Brit Med J 1960; 1: 78-82.
 42. Bankowitz RA, McNeil MA, Chaliner SM, Parker RC, Kappoor WN, Miller RA. A computer-assisted medical diagnostic consultation service. Ann Intern Med 1989; 110: 824-32.
 43. Batnett GO, Cimino JJ, Hupp JA, Hoffer EP. Dxpain: an evolving diagnostic decision system. JAMA 1987; 87: 67-74.
 44. Warner HR Jr. Iliad: moving medical decision-making into new frontiers. Method Inf Med 1989; 28: 370-2.
 45. Andrade PJN, Menezes HVB, Teixeira CAC, Colares FAN, Lima JMC. Projeto Hipócrates - Sistema de auxílio ao diagnóstico diferencial em Cardiologia. Sessões de Temas Livres: VIII Congresso Norte-Nordeste de Cardiologia, 1988.
 46. Andrade PJN, Menezes HVB, Teixeira CAC, et al. - Avaliação de um software de diagnóstico diferencial em Medicina Interna e Cardiologia. Arq Bras Cardiol 1993; 60: 285-8.
 47. Andrade PJN, Menezes HVAB, Rocha ELA, Magalhães HA. Avaliação com ênfase em Cardiologia de um software de diagnóstico diferencial em Medicina Interna. Arq Bras Cardiol 1996; 67(supl 1): 110.
 48. Berner ES, Webster GD, Shugerman AA, et al. Performance of four computer-based diagnostic systems. N Eng J Med 1994; 330: 1792-6.
 49. Kassirer JP. A report on computer-assisted diagnosis - the grade. CN Eng J Med 1994; 330: 1884-5.
 50. Barnett GO. The computer and clinical judgement. N Eng J Med 1982; 307: 493-4.
 51. Sabatini RME. O diagnóstico médico por computador. Informédica 1993; 1: 5-10.
 52. Schwartz P, Papil RM, Slozovits P. Artificial intelligence in medical diagnosis: where do we stand? N Eng J Med 1987; 316: 685-7.