

## A Prática Clínica Baseada em Evidências. Parte III – Avaliação Crítica das Informações de Pesquisas Clínicas<sup>(\*)</sup>

### *Evidence Based Clinical Practice. Part III – Critical Appraisal of Clinical Research*

Moacyr Roberto Cuce Nobre, Wanderley Marques Bernardo, Fábio Biscegli Jatene

#### RESUMO

No presente artigo discutiremos como avaliar criticamente as informações obtidas. Para tanto, serão revistos conceitos que a epidemiologia clínica tem colocado à disposição da prática clínica baseada em evidências. A pesquisa clínica busca desenvolver meios diagnósticos e terapêuticos medindo associações, ou relações de causa/efeito, entre um fator em estudo e um desfecho clínico. Onde o fator em estudo é o sintoma, sinal propedêutico, teste laboratorial, exame de imagem, ou tratamento. E o desfecho clínico é o reconhecimento da doença, cura, morte ou limitação. É um erro básico de interpretação tomar os marcadores intermediários como desfechos clínicos. Os desenhos de estudo na área clínica apresentam quatro enfoques principais: diagnóstico, prognóstico, terapêutico, dano/etiológico. Experimentos animais, estudos anatômicos, fisiológicos, genéticos, farmacológicos, análises econômicas são necessários à formação básica do médico, no entanto, não colaboram diretamente para a tomada de decisão clínica. Os desenhos de estudo são classificados de acordo com a presença de grupo controle, seguimento dos pacientes ao longo do tempo e presença de intervenção dos pesquisadores no sentido de modificar a evolução dos pacientes. A hierarquia da força de evidência científica está fundamentada nestas características e na susceptibilidade aos vícios decorrentes do tipo desenho de estudo. Revisões sistemáticas são consideradas de maior força do que os estudos primários que lhe deram origem, e quando envolvem ensaios clínicos randomizados são colocadas no topo da hierarquia. Desde 1998 é crescente a proporção de “diretrizes clínicas baseadas em evidências” quando comparada às revisões sistemáticas, ou às publicações sobre diretrizes em geral, embora as primeiras sejam ainda em número reduzido. A avaliação crítica do artigo deve responder uma questão clínica, e ser consistente quanto à adequação do desenho de estudo e ao controle dos vícios. Concluímos que se deve oferecer atualização metodológica aos profissionais interessados, e colocar a informação já avaliada à disposição de todos, elaborando e divulgando diretrizes baseadas em evidências.

**Palavras-chave:** Avaliação crítica. Desfecho clínico. Desenho de estudo. Medicina baseada em evidências. Diretrizes. Pesquisa clínica.

#### ABSTRACT

*Evidence based health care begins with a clinical question and the search on data bases to retrieve the relevant information, that was the issue of two preceding articles of this series. At present it will be discussed how to critically appraise the medical literature using the clinical epidemiological methodology. Clinical research aims to develop diagnostic and therapeutic procedures measuring association and causality between the exposure and outcome. In this case the exposures are signs, symptoms, laboratorial or image exam, and therapy intervention. It is a mistake to take surrogate end-points instead of clinical outcomes. The main types of clinical study design are diagnostic, prognostic, therapeutic and harm/etiology. Experimental, physiologic and animal studies are useful for the medical undergraduate education, but do not contribute with clinical decisions. The study designs are classified according with the presence of a control group, patient's follow-up, and therapy interventions. The evidence hierarchy was done by the previous characteristics and the presence of systematic bias. Systematic reviews are stronger than the primary observational studies and are on the top when they revised randomized clinical trial. Since 1998 the proportion of evidence based practice guidelines was increasing compared with systematic reviews or other types of practice guidelines, although the former still are in a few numbers. The article critical appraisal must answer the clinical question, and need to have consistent study design and bias under control. In conclusion we ought to offer methodological actualization to interested physicians and put the information already critically assessed on evidence-based practice guidelines. [Rev Assoc Med Bras 2004; 50(2): 221-8].*

**Keywords:** Critical appraisal. Outcome. Study design. Evidence based medicine. Guidelines. Clinical research.

\* Trabalho realizado por representantes da Associação Médica Brasileira, São Paulo, SP, publicado na Rev Assoc Med Bras 50(2): 221-8, 2004. Recebido em 08/02/04. Aprovado, após revisão, em 28/02/04.

## INTRODUÇÃO

A prática da medicina de boa qualidade começa pelo reconhecimento da dúvida sobre qual o procedimento mais eficiente para o atendimento do paciente. Qual a estratégia diagnóstica de maior acurácia, considerando as limitações de acesso e custo? Qual o procedimento terapêutico mais indicado para o controle da doença? Quais as medidas preventivas para se evitar o aparecimento da doença ou de suas complicações?

No primeiro artigo desta série sobre “Prática Clínica Baseada em Evidências”, vimos como a partir de um cenário clínico do paciente alvo do atendimento podemos construir uma pergunta bem estruturada na forma de P.I.C.O, incorporando as características do paciente (P), a intervenção ou indicador (I) da qual se quer a evidência, o controle ou condição habitual a ser comparada (C), e o desfecho ou *outcome* (O) que se espera<sup>(1)</sup>. No segundo artigo da série vimos como podemos fazer a busca da resposta em bases de dados, onde se encontram essas bases, e como a internet pode ser útil na busca da resposta<sup>(2)</sup>. No presente artigo discutiremos como avaliar criticamente as informações obtidas. Para tanto serão revistos conceitos que a epidemiologia clínica tem colocado à disposição da prática clínica baseada em evidências. Para os interessados em se aprofundar nessa área do conhecimento recomenda-se a literatura básica desenvolvida nas décadas de 80 e 90 por Fletcher<sup>(3)</sup>, Sackett<sup>(4)</sup>, Feinstein<sup>(5)</sup>, introduzida no nosso meio por Schmidt e Duncan<sup>(6)</sup>.

A primeira questão que se coloca é se o assunto ligado à metodologia científica é de interesse de quem pratica a medicina longe dos centros universitários. Qual a necessidade deste profissional? Será que ele precisa reciclar seus conhecimentos e dominar essa metodologia? Ou será melhor ele receber a informação previamente avaliada? Ou ainda, será que ele tem condições de praticar a medicina de acordo com a informação mais gabaritada? Quem precisa fazer a avaliação crítica da literatura médica? O que é prioritário: ensinar o profissional a avaliar criticamente a informação ou colocar à sua disposição a informação já avaliada?

Ao se terminar uma aula em que diversos aspectos da medicina baseada em evidências e dos fundamentos da epidemiologia clínica são apresentados, costuma vir a pergunta fatal e inevitável: Como é que essas recomendações podem ser praticadas no mundo real? Como, na rotina assistencial, se pode solicitar exames com a racionalidade preconizada pela probabilidade condicional de Bayes?

Como fazer isso, se no mundo real quem determina os exames que podem ser pedidos é a restrição econômica da fonte pagadora? Não bastasse a pressão que limita a autonomia profissional, há a pressão no sentido oposto da sociedade, que estimula o consumo desenfreado de procedimentos diagnósticos e terapêuticos. Embora muitas vezes seja expressa pelo paciente, a pressão tem origem numa poderosa dupla composta pela mídia fantástica e sua parceira, a lucrativa indústria da saúde. Sem falar na influência dos interesses securitários, a mão de ferro corretora dos desvios que se contrapõem aos interesses do mercado. Convenhamos, é muita pressão sobre quem tem como objetivo auferir remuneração digna em decorrência da prática qualificada. Mas será que esses dois mundos são tão distantes como parecem? Qual o tipo de ponte que precisa ser construída para aproximá-los? Quais os buracos que precisam ser tapados?

Tudo começa na boa relação médico-paciente, na atenção que o profissional dispensa ao paciente quanto aos seus valores, crenças e preferências. Daí a necessidade de uma forte formação calcada em valores humanitários, que atenda as necessidades de comunicação interpessoal. Numa sociedade que privilegia as responsabilidades individuais em detrimento das causas estruturais do adoecimento, a comunicação, o diálogo e as questões educativas desempenham papel central no atendimento. Isto reconhecido, como melhorar a efetividade dos procedimentos médicos?

## PROCEDIMENTOS EFETIVOS

Os procedimentos clínicos são de duas naturezas principais: aqueles que visam reconhecer, diagnosticar e prever a evolução de agravos na condição de saúde e as intervenções que visam tratá-los ou preveni-los.

A medida da associação existente entre o fator em estudo e o desfecho clínico pode ser feita em diferentes contextos de pesquisa, como os que lidam com as pessoas em seu mundo real, outros que elegem situações artificiais de controle sobre o objeto de pesquisa, ou atribuem medidas de custo financeiro. A epidemiologia clínica coloca esses estudos dentro dos conceitos de eficácia, efetividade e eficiência<sup>(3)</sup>. A eficácia diz que uma intervenção qualquer, por exemplo, a prescrição de um medicamento, funciona no mundo artificial, onde não só as questões relacionadas com a administração do remédio e aderência ao protocolo de pesquisa estão sob controle, como também as funções orgânicas dos participantes da pesquisa. É o que acontece nos ensaios fase I ou II realizados com pacientes hospi-

talizados, voluntários sadios, e experimentos com modelos animais. Diferentemente da artificialidade que cerca os estudos de eficácia, a efetividade diz que uma intervenção funciona no mundo real, já que o paciente se encontra no ambiente livre do seu dia-a-dia, como os tratamos no consultório. Por último, o conceito de eficiência diz que determinado procedimento além de efetivo é economicamente vantajoso.

O custo do atendimento acaba interferindo com os limites da prática médica. Por maior que seja a disponibilidade de recursos econômicos, eles são finitos. Ao contrário, as alternativas que a tecnologia diagnóstica e terapêutica colocam à disposição do atendimento à saúde crescem continuada e indefinidamente. No entanto, as análises econômicas dos estudos de custo/eficiência dizem mais respeito aos planejadores de serviços, como os de eficácia interessam mais de perto aos pesquisadores envolvidos nas fases iniciais de desenvolvimento de novas drogas. A todos nós que clinicamos, os estudos de efetividade são os prioritários.

Tanto o planejamento do estudo, o cenário da pesquisa, como a forma de se medir os resultados interferem quando vamos aplicar a informação da pesquisa no paciente o qual estamos cuidando. A maior possibilidade de aplicação dos resultados depende das semelhanças entre o nosso paciente e os pacientes que foram pesquisados, em sua dimensão biológica, clínica, psíquica, social e cultural. Por razões óbvias, as informações geradas em animais de laboratório, voluntários e pacientes hospitalizados com a finalidade de garantir a aderência ao protocolo de pesquisa têm menor possibilidade de serem aplicadas na prática médica habitual. Não podemos esquecer que mesmo os ensaios de efetividade gerados em populações e países com características muito diversas das nossas nem sempre podem ser aplicados ao nosso paciente.

## DESFECHOS CLÍNICOS

A pesquisa clínica busca desenvolver meios diagnósticos e terapêuticos medindo associações, ou relações de causa/efeito, entre um fator em estudo e um desfecho clínico. Onde o fator em estudo é o sintoma, sinal propedêutico, teste laboratorial, exame de imagem, ou tratamento. E o desfecho clínico, *outcome* na língua inglesa, é o reconhecimento da doença, cura, morte, limitação funcional, complicação evolutiva, ou qualquer outro desfecho clínico que interfira com o tempo ou a qualidade de vida. Não são incluídos entre esses últimos os marcadores biológicos

ou de imagem porque a presença do fator reumatóide não significa a presença da doença reumatóide, como a elevação do colesterol sérico não representa a doença coronária, como a contagem de CD4 não significa infecção oportunista no portador do vírus HIV, como a hemoglobina glicosilada não representa a evolução da diabete, e o abaulamento do disco intervertebral não quer dizer que existe uma hérnia discal. A definição da doença pressupõe obrigatoriamente a presença de outras manifestações além da presença destes marcadores intermediários, da mesma forma que essas condições clínicas acontecem na ausência deles. Sempre haverá um conjunto de manifestações que representa o padrão-ouro para definição da doença, ou ela se dará por meio de sua evolução clínica.

Outras evidências mostram que o marcador intermediário pode enganar o raciocínio médico<sup>(7)</sup>. A redução endoscópica de ulceração gástrica observada com os anti-inflamatórios inibidores da COX-2 não é acompanhada por redução de perfurações e sangramentos. O aumento da densidade óssea ao exame radiológico devido à terapia com flúor na osteoporose pós-menopausa não reduz a ocorrência de fraturas. O equívoco do marcador intermediário é capaz de induzir mortes evitáveis, como o uso profilático de lidocaína pós-infarto do miocárdio que reduz o risco de arritmia ventricular, mas aumenta a mortalidade. O uso de quinidina após o tratamento cardioversivo na fibrilação atrial, apesar de favorecer a manutenção do sinal eletrocardiográfico de ritmo sinusal, triplica o número de mortes.

Portanto, é um erro básico de interpretação de pesquisa tomar os marcadores intermediários como desfechos clínicos, por mais que essa confusão atenda aos interesses de mercado da indústria farmacêutica, de equipamentos e marcadores biológicos. Eles devem ser considerados exclusivamente como fatores preditivos que auxiliam na tomada de decisão clínica que exige muitas outras competências do profissional qualificado. Fontes secundárias de informação têm colocado restrições na divulgação de pesquisas clínicas que utilizam desfecho intermediário, como os *InfoPoems*<sup>(7)</sup>, que só divulgam pesquisas que tenham desfechos clínicos.

## TIPOS DE ESTUDO

A pesquisa clínica estuda a influência de determinados fatores sobre a saúde das pessoas, na forma de um desfecho clínico.

Se esse fator favorece o aparecimento de doença, o enfoque da pesquisa é chamado etiológico. No caso específico do agente ser um medicamento ou outra intervenção

médica, é chamado de enfoque de dano. Se o fator em estudo visa o reconhecimento da doença, o enfoque é diagnóstico. Se o fator em estudo influencia uma determinada evolução da doença, o enfoque é prognóstico. Se o fator em estudo é o benefício de uma cirurgia, medicamento ou uma outra intervenção experimental sobre a doença, o enfoque é dito terapêutico. Experimentos animais, estudos anatômicos, fisiológicos, genéticos, farmacológicos, análises econômicas e outros estudos biológicos são necessários à formação básica do médico ou para gerar indagações clínicas, no entanto, não colaboram diretamente para a tomada de decisão clínica, e por esse motivo devem ser considerados como de apoio ou complementares às pesquisas clínicas (Figura 1).

A relação temporal entre o fator em estudo e o desfecho clínico determina os três tipos básicos de desenho de estudo. Se o registro dos dados de pesquisa se inicia com o fator estudado para posterior aferição do desfecho clínico, o desenho observa no tempo a história natural de vida das pessoas, por exemplo, o estudo de coorte que segue indivíduos com obesidade com a finalidade de estudar sua influência sobre o surgimento da diabete.

As informações sobre o fator em estudo como os do desfecho clínico podem ser colhidas ao mesmo tempo ou no mesmo intervalo de tempo. Por exemplo, quando o

índice de massa corporal como fator em estudo é avaliado em conjunto com a aferição dos níveis pressóricos para se estimar a associação existente entre obesidade e hipertensão arterial, esse desenho é chamado transversal, ou *survey* na língua inglesa. Se no mesmo intervalo de tempo se mede o fator em estudo de forma agregada como a taxa de infecção hospitalar e o desfecho clínico registrado como óbitos hospitalares na unidade de terapia intensiva, para se avaliar a associação entre os dois, o desenho é chamado estudo agregado ou ecológico. A diferença básica entre eles é que no transversal o pesquisador afere os dados avaliando individualmente as pessoas que participam do estudo e no agregado as informações provêm de registros coletivos, e não individuais.

O terceiro e último desenho básico de pesquisa clínica colhe as informações de maneira invertida com relação à evolução natural das doenças. Parte do desfecho clínico já está instalada e se pergunta ao participante do estudo sobre o fator em estudo ocorrido no passado. Por exemplo, o desenho do tipo caso-controle com pessoas que têm câncer de próstata como desfecho clínico, que são indagadas se foram submetidas à vasectomia no passado, comparadas com os controles que não têm o câncer. A indagação clínica, neste exemplo, é se a vasectomia, como fator em estudo, está associada ao câncer de próstata como desfecho clínico.

Esses são os únicos três desenhos básicos de pesquisa clínica: coorte, transversal e caso/controlado. Todos os demais são variações destes três, como a coorte histórica em que os dados sobre o fator em estudo foram registrados no passado. A diferença para o caso/controlado é que a coorte se vale de registro confiável, realizado no momento que o fator em estudo realmente aconteceu. Por exemplo, as coortes de nascimento que avaliam a correlação negativa entre baixo peso ao nascer, como fator em estudo, e a hipertensão arterial na idade adulta, como desfecho clínico. No desenho do tipo coorte histórica, como na coorte em que os eventos são registrados de forma contemporânea e prospectiva, é garantida a precedência do fator em estudo com relação ao desfecho clínico, evitando-se o vício que depende da memória das pessoas estudadas, comum nos estudos do tipo caso-controlado.

Há também o ensaio clínico que apresenta o mesmo desenho da coorte, com uma única diferença: enquanto na coorte as informações são provenientes da observação evolutiva, no ensaio clínico elas decorrem de uma intervenção terapêutica dos pesquisadores. No planejamento de um ensaio clínico, a aleatorização dos grupos que são comparados é a melhor forma de torná-los semelhantes

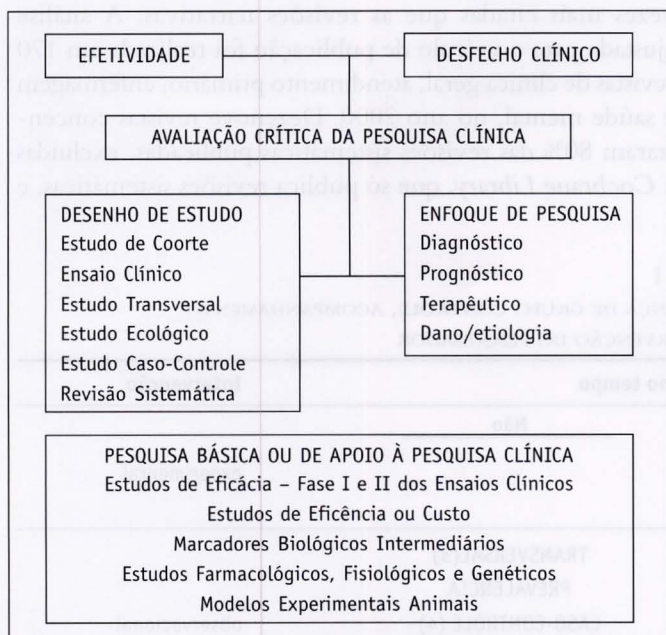


FIGURA 1 – Características a serem observadas na avaliação crítica da pesquisa clínica.

em todas as suas características, menos quanto ao tratamento que se estuda. Estes estudos, conhecidos como ensaios clínicos randomizados, são considerados os de maior força de evidência científica entre as pesquisas de enfoque terapêutico.

Os desenhos de estudo são classificados de acordo com a presença de grupo controle, seguimento dos pacientes ao longo do tempo e presença de intervenção dos pesquisadores no sentido de modificar a evolução dos pacientes (Tabela 1). A hierarquia da força de evidência científica fundamentada na susceptibilidade aos vícios decorrentes do tipo desenho de estudo é consensual entre as agências preventivas canadense e norte-americana, e as sociedades de especialidades médicas. A presença das três características acima citadas no desenvolvimento da pesquisa confere a maior força de evidência científica ao ensaio clínico randomizado. Na perda da condição experimental, os estudos observacionais, por serem mais suscetíveis aos vícios de confusão, ficam no segundo patamar da hierarquia, como os estudos de coorte. A ausência de seguimento no tempo confere potencial maior de susceptibilidade aos vícios de confusão, além de não garantir a precedência da exposição com relação ao desfecho, colocando o transversal no terceiro patamar. Os potenciais vícios de memória nas informações dadas pelos pacientes de ocorrências do passado, e os que se devem ao fato dos casos e controles se originarem de populações diferentes, colocam o caso-controle no quarto patamar. E, por fim, a ausência do grupo controle coloca o relato de casos no quinto patamar da hierarquia. No entanto, a lógica desta classificação deve ser considerada como mais um instrumento auxiliar na avaliação crítica. Não pode ser confundida com uma regra inflexível. Existem inúmeros exemplos de trabalhos cientí-

ficos não controlados, como relatos de casos, com força suficiente para comprovar o benefício clínico da penicilina, ou retirar do mercado os anoréticos à base de fenfluramina e fentermina, em função do malefício para a válvula cardíaca.

O artigo de revisão clássico, por vezes na forma de capítulo de livro, é considerado uma revisão narrativa. Não apresenta descrição de como as referências foram obtidas, nem como foram avaliadas criticamente, como também não inclui métodos de controle sobre os vícios que podem comprometer suas conclusões. Não preenche, portanto, a premissa básica para ser considerado um trabalho científico, no qual a metodologia deve garantir as informações que permitam que outros autores possam chegar às mesmas conclusões, caso queiram reproduzir o mesmo trabalho. Algumas revisões são compilações de resumos de artigos, com a finalidade precípua de refletir a opinião do autor, por vezes contrária às melhores evidências científicas, ou discordante da opinião de publicações feitas por outros autores da mesma especialidade.

Estas observações não pretendem desqualificar a revisão narrativa que, por sua excelência didática, tem um papel fundamental para a educação médica, mas argumentar a favor da superioridade da força de evidência científica da revisão sistemática sobre a revisão narrativa. Trabalho de boa qualidade científica com controle de vícios sistemáticos e aleatórios mostrou que as revisões sistemáticas são duas vezes mais citadas que as revisões narrativas. A análise ajustada para o veículo de publicação foi realizada em 170 revistas de clínica geral, atendimento primário, enfermagem e saúde mental, no ano 2000. Dezenove revistas concentraram 80% das revisões sistemáticas publicadas, excluídas a *Cochrane Library*, que só publica revisões sistemáticas, e

TABELA 1  
TIPOS DE DESENHO DE PESQUISA EM FUNÇÃO DA PRESENÇA DE GRUPO CONTROLE, ACOMPANHAMENTO DOS PACIENTES NO TEMPO E INTERVENÇÃO DO PESQUISADOR

Grupo controle	Acompanhamento no tempo		Intervenção
	Sim	Não	
sim	ENSAIO CLÍNICO (1)		experimental
não	ENSAIO NÃO CONTROLADO		
sim	COORTE (2)	TRANSVERSAL(3)	observacional
não	INCIDÊNCIA	PREVALÊNCIA	
sim		CASO-CONTROLE (4)	
não		RELATO DE CASOS (5)	

Os números entre parênteses correspondem à hierarquia decrescente na força de evidência científica.

o *New England Journal of Medicine* que veta este tipo de publicação<sup>(8)</sup>. Revisões sistemáticas são consideradas de maior força de evidência científica do que os estudos primários que lhes deram origem, tanto o caso-controle como a coorte. A revisão sistemática de ensaios clínicos randomizados é colocada no topo da hierarquia da evidência científica<sup>(9)</sup>.

Tomando o banco de dados do Medline como indicador da incorporação de conceitos da epidemiologia clínica na literatura médica, constatamos que em 1990 é introduzida a “revisão sistemática” como tipo de estudo. Em 1992, é publicado no JAMA o artigo que lança o movimento da “medicina baseada em evidências” propondo um novo paradigma para o ensino da medicina. Em 1993, é acrescida a “diretriz clínica” como tipo de estudo no Medline. Em 1995, são publicados no JAMA os três primeiros artigos sobre diretrizes clínicas baseadas em evidências da série *Users’ guides to the medical literature*. Em 1997, é a vez do termo “medicina baseada em evidências” ser acrescido como descritor no Medline. Desde 1998 é crescente a proporção de “diretrizes clínicas baseadas em evidências” (Figura 2) quando comparada às revisões sistemáticas, ou às publicações sobre diretrizes em geral, embora em número absoluto, os 254 trabalhos estejam bem aquém da quantidade observada dos outros tipos de publicação (Tabela 2). No período de seis anos, que vai de 1998 até 2003, foram classificadas, no Medline, 19.759 revisões narrativas, 12.216 artigos de medicina baseada em evidências, 5.435 revisões sistemáticas e 4.455 diretrizes clínicas.

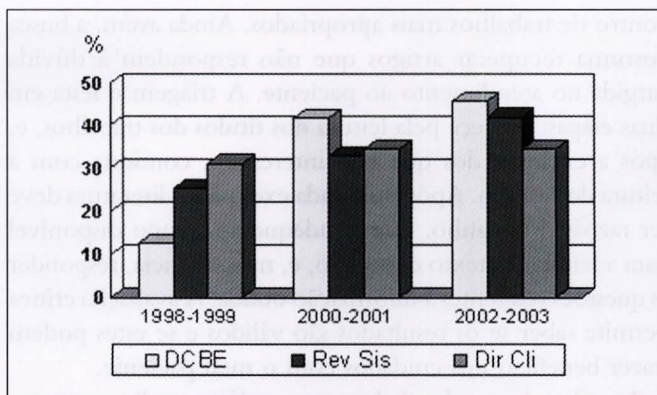


FIGURA 2 – Distribuição percentual dos artigos recuperados sobre diretrizes clínicas baseadas em evidências (DCBE), revisões sistemáticas (Rev Sis), e diretrizes clínicas (Dir Cli), entre os anos de 1998 e 2003 no banco de dados Medline.

### CONSISTÊNCIA DO ARTIGO<sup>(10)</sup>

A primeira questão a ser aplicada a todo artigo obtido na busca para responder a uma questão clínica está relacionada à consistência da informação, se os componentes adequados que devem compor o desenho de estudo estão presentes.

Se a busca na base de dados primária resultar em grande quantidade de trabalhos, é necessário reduzi-los através da combinação adequada das palavras-chave por meio dos booleanos, iniciando com “OR” e concluindo com “AND” E “NOT”. A aplicação dos filtros de faixa etária, sexo e exclusão dos trabalhos com animais também auxilia o en-

TABELA 2

REVISÕES SISTEMÁTICAS, DIRETRIZES CLÍNICAS E ARTIGOS CLASSIFICADOS COMO SENDO DE MEDICINA BASEADA EM EVIDÊNCIAS NO BANCO DE DADOS MEDLINE, ANO DE INTRODUÇÃO NO BANCO DE DADOS E NÚMERO DE ARTIGOS RECUPERADOS ENTRE OS ANOS DE 1998 E 2003

Forma de busca no Medline	Ano de introdução	Artigos entre 1998 e 2003
Revisões sistemáticas ("Meta-Analysis"[Publication Type])	1990	5.435
Diretrizes clínicas ("Practice Guideline"[Publication Type])	1993	4.455
Medicina Baseada em Evidências ("Evidence-Based Medicine"[MeSH])	1997	12.216
Diretrizes clínicas baseadas em Evidências ("Evidence-Based Medicine"[MeSH] AND "Practice Guideline"[Publication Type])	1997	254

contro de trabalhos mais apropriados. Ainda assim, a busca costuma recuperar artigos que não respondem à dúvida surgida no atendimento ao paciente. A triagem é feita em duas etapas: começa pela leitura dos títulos dos trabalhos, e, após a exclusão dos que não interessam, continua com a leitura do resumo. Após a segunda exclusão, a literatura deve ter razoável tamanho, que se adeque ao tempo disponível para a leitura do texto completo, e, na seqüência, responder às questões referentes à informação obtida. A avaliação crítica permite saber se os resultados são válidos e se estes podem trazer benefício nos cuidados com o meu paciente.

A análise de estudos de bases secundárias avalia se a questão foi adequadamente formulada, se a evidência foi coletada, avaliada e sumarizada apropriadamente.

A avaliação crítica de pesquisas clínicas depende do enfoque do estudo realizado, se diagnóstico, prognóstico, terapêutico ou dano.

## ESTUDOS DIAGNÓSTICOS

Os elementos necessários, para que um artigo de estudo do tipo diagnóstico ou de rastreamento seja considerado consistente, podem ser agrupados de forma a responder a três diferentes questões:

a) Existe uma comparação cegada e independente do teste diagnóstico com o “padrão ouro” usado no reconhecimento da doença?

Os pacientes em estudo devem ser submetidos aos dois procedimentos diagnósticos, aquele sendo testado e o “padrão ouro” de referência, reconhecidos como provas de que o paciente tem ou não tem uma determinada doença. Além disso, os resultados de um procedimento não podem ser conhecidos por quem aplica ou interpreta o teste que está sendo comparado. Esses dois elementos são fundamentais para que sejam afastados os vícios de interpretação, conscientes ou inconscientes, dos investigadores. O padrão-ouro de referência deve ser aplicado em todos os pacientes testados.

b) O teste diagnóstico foi avaliado em uma amostra apropriada de pacientes semelhante àquela encontrada na prática clínica diária?

Os pacientes em estudo devem apresentar os achados clínicos comumente presentes na doença em questão, bem como as características clínicas presentes em doenças que entram no diagnóstico diferencial. Estudos que aplicam testes diagnósticos comparando casos clínicos atípicos, muito específicos, ou utilizando pacientes assintomáticos não são relevantes.

c) O teste diagnóstico foi validado em um segundo grupo independente de pacientes?

O primeiro estudo (com o teste diagnóstico) serve para se fazer uma predição da sua acurácia. A confirmação dos níveis semelhantes de desempenho com a aplicação em uma segunda amostra, independente, de pacientes permite a estimativa real da acurácia do teste diagnóstico em estudo.

## ESTUDOS TERAPÊUTICOS

A avaliação de estudos individuais de terapêutica, identificando a presença dos elementos que os caracterizam como consistentes, não define esses estudos como a melhor evidência, uma vez que as revisões sistemáticas desses estudos individuais fornecem resultados de melhor consistência. No entanto, ao obtermos a resposta para uma dúvida clínica de tratamento em um estudo individual, devemos analisar os seguintes itens:

a) A seleção dos pacientes foi aleatorizada e cegada?

Aleatorizar os grupos de pacientes a serem comparados faz com que esses grupos, o tratado e o controle, tenham a mesma chance de ocorrência do evento que se espera prevenir com o tratamento. Além disso, promove um equilíbrio nos preditores de bom e mau prognóstico, evitando exagerar, reduzir ou contrariar os efeitos da terapia.

O cegamento busca impedir o pesquisador de saber qual o tratamento que o próximo paciente receberá, e, assim, evitar a influência consciente ou inconsciente na resposta dos grupos em comparação.

O objetivo principal da aleatorização e do cegamento é se obter o maior número possível de verdadeiros positivos e negativos na resposta terapêutica. Outros recursos como a análise por intenção de tratamento, o cegamento de outros elementos do estudo como o paciente, profissionais de saúde, analisador, e do próprio sorteio têm como objetivo garantir a aleatorização inicial.

b) O tempo de seguimento dos pacientes foi suficientemente longo e completo?

Perdas de seguimento dos pacientes ao longo do estudo podem afetar as conclusões deste, uma vez que a resposta desconhecida desses pacientes ao tratamento poderia mudar os resultados da comparação. Trabalhos com perda de pacientes acima de 20% não devem ser aceitos.

O efeito do tratamento, na dependência da história natural da doença, só pode ser avaliado após um período adequado de tempo de seguimento. Esse período, quando encurtado, pode levar, por meio de desfechos intermediários,

a conclusões inadequadas, que exageram ou reduzem os efeitos do tratamento.

c) Todos os pacientes foram analisados dentro do grupo no qual foram inicialmente aleatoriamente alocados?

A fim de preservar a aleatorização, a análise por intenção de tratamento deve ser utilizada. Todos os pacientes que completam o estudo são analisados nos grupos em que foram inicialmente alocados, independentemente do tipo de tratamento que esses pacientes tenham recebido.

## ESTUDOS PROGNÓSTICOS

As questões sobre prognóstico têm feito parte da prática clínica diária, seja quando originadas em dúvidas do paciente, por exemplo, quando estes questionam quanto tempo de vida ainda possuem, seja quando originadas em dúvidas de decisão médica. Essas questões possuem três componentes principais relacionados ao desfecho: um componente qualitativo relacionado ao tipo de desfecho, um componente quantitativo relacionado à probabilidade de ocorrência e um componente relacionado à duração de tempo.

Assim como nos estudos diagnósticos, a estrutura de avaliação crítica da evidência em prognóstico, visando validar esses estudos, é baseada na resposta a quatro questões principais:

a) Foi reunido um grupo representativo de pacientes no mesmo estágio da doença, preferencialmente inicial?

A alocação ideal, mas impossível de pacientes seria aquela que incluísse toda a população de pacientes com uma determinada doença, estudada a partir do seu início. Então, ao avaliarmos um determinado estudo prognóstico, o que devemos e podemos fazer frente ao ideal é observarmos, com atenção, como a doença foi definida e como os pacientes foram alocados.

A amostra precisa representar as características da doença, da maneira que a encontramos em nossa prática diária, mas principalmente, a doença deve estar em seu curso inicial, e no mesmo ponto em todos os pacientes alocados, o que podemos definir como uma coorte a partir da fase inicial da doença, na língua inglesa *inception cohort*. É claro que, se nosso objetivo é avaliar o prognóstico tardio de uma doença, e só nessa situação, podemos alocar pacientes no mesmo ponto avançado de doença.

b) Os pacientes foram acompanhados durante um tempo suficientemente longo e completo?

O desfecho de interesse do estudo nem sempre é obtido através de um longo tempo de seguimento. No entanto,

desfechos prognósticos devem ser avaliados de maneira que todos os pacientes tenham seguimento suficiente e completo para que possam manifestar esse desfecho. O bom prognóstico é constatado quando, após um tempo suficientemente longo, o evento adverso não está presente nos pacientes alocados. O estudo pode ser invalidado quando o seguimento dos doentes é interrompido antes que o evento adverso possa se manifestar. Não podemos estabelecer conclusões prognósticas baseadas em estudos cujo desfecho é analisado como um marcador intermediário, e nessa situação devemos buscar outra evidência.

Seguir pacientes ao longo do tempo tem como problema central o número de perda de seguimento dos pacientes que iniciaram o estudo. O motivo da perda pode não estar relacionado ao desfecho, como por exemplo, mudança de cidade ou de país, mas as perdas podem estar relacionadas diretamente ao evento adverso, como no caso de morte ou agravamento da doença, o que produz prejuízo às conclusões do estudo prognóstico.

Uma perda inferior a 5% da amostra não deve produzir efeitos negativos nas conclusões do estudo, no entanto, uma perda de 20%, certamente alterará as conclusões, uma vez que é mais provável que somente os pacientes de melhor condição clínica terminem o estudo. Ao avaliarmos o número de perdas do estudo, devemos sempre levar em consideração que o evento morte pode estar presente em todos os pacientes com seguimento perdido.

c) O critério de aferição do desfecho é objetivo e foi aplicado de maneira cegada?

Os desfechos extremos relacionados à morte e cura total não são difíceis de serem avaliados: nesta situação dificilmente ocorrerão vícios que dependam do pesquisador. O mesmo não pode ser afirmado quando imaginamos que o desfecho pode estar relacionado às causas de mortalidade ou à capacidade de retorno às atividades de trabalho. Por isso, os critérios de aferição do desfecho devem ser objetivos e claramente definidos no início do estudo. É preferível que o investigador que afere o desfecho seja vendado, ou seja, não conheça previamente as características clínicas e os fatores prognósticos dos pacientes, uma vez que poderá interferir na observação do desfecho, por mais que os critérios de aferição sejam objetivos.

d) Se diferentes subgrupos prognósticos foram identificados: Os mesmos foram ajustados com relação aos fatores prognósticos principais? Houve teste piloto em grupo independente de pacientes para validar os fatores prognósticos?

Sabe-se que em grupo de pacientes estudados pode haver subgrupos com diferentes prognósticos. Muitas vezes esses



subgrupos são claramente definidos, mas em outras situações clínicas não. Por exemplo, o risco de acidente vascular cerebral em pacientes com fibrilação atrial pode variar em função do tamanho do átrio esquerdo.

É necessário que se leve em consideração possível interação com outros fatores prognósticos relevantes, que nesse exemplo poderiam ser a hipertensão arterial e a função ventricular. A ocorrência de distribuição heterogênea desses fatores nos subgrupos pode provocar interação e influências mútuas que levam a um desvio das conclusões. Deve se avaliar a sessão do método e resultados em busca dessas características antes de se aceitar um prognóstico diferente para o subgrupo de interesse.

A análise estatística de subgrupos deve ser preditora e não explanatória. Não deve levar em consideração se o fator prognóstico é plausível no ponto de vista fisiológico, como no exemplo do átrio esquerdo, ou biologicamente sem sentido, ou ainda, quando é parte dos dados do paciente, mas semnexo causal com o desfecho, como o

signo e o time do coração. Fatores prognósticos supostos podem ser sexo, idade, extensão da doença e comorbidades.

Por essas razões, não há garantia de que a identificação de um fator prognóstico possa afirmar que ele identifica subgrupo de pacientes de prognóstico diferente. Por isso a necessidade de que o estudo tenha avaliado e confirmado previamente o poder de predição desses fatores prognósticos, em grupos independentes de pacientes, denominados grupos de teste ou grupos de validação.

Respondendo às questões colocadas no início do artigo sobre qual deve ser o papel das entidades médicas: ensinar o profissional a avaliar criticamente a informação ou colocar à sua disposição a informação já avaliada. Concluímos que as duas devem ser contempladas. Ensinar os profissionais interessados, por meio de curso de atualização metodológica, e colocar a informação já avaliada à disposição de todos, elaborando e divulgando diretrizes baseadas em evidências.

**Conflito de interesse:** não há.

## REFERÊNCIAS

1. Nobre MR, Bernardo WM, Jatene FB. Evidence based clinical practice. Part 1 — well structured clinical questions. *Rev Assoc Med Bras* 49(4):445-9, 2003 Out-Dec.
2. Bernardo WM, Nobre MR, Jatene FB. A prática clínica baseada em evidências: parte II - buscando as evidências em fontes de informação. *Rev Assoc Med Bras* 50(1):104-8, 2004 Jan-Fev.
3. Fletcher RH, Fletcher SW, Wagner EH. *Epidemiologia Clínica: Elementos Essenciais*, 3a. ed. Porto Alegre: Artes Médicas, 1996.
4. Sackett DL, Haynes RB, Tugwell P, Guyatt GH. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. 2a.ed. Boston: Little Brown, 1992.
5. Feinstein AR. *Clinical Epidemiology: The Architecture of Clinical Research*. Philadelphia: Saunders, 1985, 812 p.
6. Schmidt MI, Duncan BB. O método epi-demiológico na conduta e na pesquisa clínica. In: Rouquayrol MZ. *Epidemiologia e Saúde*. 4a. ed. Rio de Janeiro: Medsi, 1993. p. 183-207.
7. Shaughnessy AF, Slawson DC. What happened to the valid POEMs? A survey of review articles on the treatment of type 2 diabetes. *BMJ* 327 (7409):266, 2003 Aug 2.
8. Montori VM, Wilczynski NL, Morgan D and Haynes RB. Systematic reviews: a cross-sectional study of location and citation counts. *BMC Medicine* 1:2, 2003.
9. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA* 274(22):1800-4, 1995 Dec 13.
10. Sackett DL, Straus S, Richardson S, Rosenberg W, Haynes RB. *Evidence-based medicine: how to practice and teach EBM*. 2ª ed. Londres: Churchill Livingstone; 2000.