

Avaliação da qualidade de estudos clínicos e seu impacto nas metanálises

Assessment of clinical trial quality and its impact on meta-analyses

Carlos Rodrigues da Silva Filho^a, Humberto Saconato^b, Lucieni Oliveira Conterno^a, Iara Marques^b e Álvaro Nagib Atallah^c

^aFaculdade Estadual de Medicina de Marília. Marília, SP, Brasil. ^bUniversidade Federal do Rio Grande do Norte. Natal, RN, Brasil. ^cEscola Paulista de Medicina. Universidade Federal de São Paulo. São Paulo, SP, Brasil

Descritores

Ensaio controlado aleatório. Garantia de qualidade dos cuidados de saúde. Questionários. Avaliação de resultados, cuidados de saúde. Controle de qualidade.

Resumo

Objetivo

Analisar se diferentes instrumentos de avaliação de qualidade, aplicados a um grupo de estudos clínicos que se correlacionam e qual seu impacto no resultado na metanálise.

Métodos

Foram analisados 38 estudos clínicos randomizados e controlados, selecionados para a revisão sistemática sobre a eficácia terapêutica do Interferon Alfa no tratamento da hepatite crônica pelo vírus B. Utilizaram-se os seguintes instrumentos: Maastricht (M), Delphi (D) e Jadad (J) e o método da Colaboração Cochrane (CC), considerado padrão-ouro. Os resultados definidos pelos três instrumentos foram comparados pelo teste de Correlação de Spearman. O teste de Kappa (K) avaliou a concordância entre os revisores na aplicação dos instrumentos e o teste de Kappa ponderado analisou o ordenamento de qualidade definido pelos instrumentos. O clareamento do HBV-DNA e HbeAg foi o desfecho avaliado na metanálise.

Resultados

Os estudos foram de regular e baixa qualidade. A concordância entre os revisores foi, de acordo com o instrumento: D=0,12, J=0,29 e M=0,33 e CC= 0,53. A correlação foi moderada e homogênea (D/J=0,51; D/M=0,53 e J/M=0,52). Os resultados da metanálise (HBV-DNA), variaram de RR=0,71; IC 95%: 0,66-0,77 a RR=0,67; IC 95%: 0,58-0,79 e (HbeAg) de RR=0,85; IC 95%: 0,80-0,90 a RR=0,85; IC 95%: 0,77-0,93, dependendo da qualidade dos estudos incluídos.

Conclusões

Os instrumentos de avaliação de qualidade têm boa correlação. Nas revisões sistemáticas que apontem à mesma direção do efeito, a avaliação pode não alterar significativamente seu resultado. O método da Colaboração Cochrane é o mais reprodutível e de simples aplicação.

Keywords

Randomized controlled trials. Quality assurance, healthcare. Questionnaires. Outcome assessment. Quality control.

Abstract

Objective

To evaluate whether different quality assessment tools applied to a group of clinical trials could be correlated, and what would be their impact on meta-analysis results.

Methods

Thirty-eight randomized controlled clinical trials were analyzed. These had been selected for a systematic review of the therapeutic efficacy of alpha interferon for

Correspondência/ Correspondence:

Carlos Rodrigues da Silva Filho
Departamento de Medicina
Faculdade Estadual de Medicina de Marília
Av. Monte Carmelo, 800
17519-030 Marília, SP, Brasil
E-mail: silvacr@famema.br / crodrig2@uottawa.ca

Baseado na tese de doutorado apresentada à Universidade Federal de São Paulo, em 2003. Estudo realizado em colaboração entre a Disciplina de Medicina Interna da Faculdade Estadual de Medicina de Marília, Pós-Graduação em Medicina Interna e Terapêutica e Centro Cochrane do Brasil (afiliado à Colaboração Cochrane) da Escola Paulista de Medicina e Departamento de Medicina da Universidade Federal do Rio Grande do Norte. Apresentado no International Clinical Epidemiology Network (INCLIN) Global Meeting XIX e XX em Kunming, China (2003) e em Agra - Índia (2004), respectivamente: no 12th Cochrane Colloquium, em Ottawa, Canadá (2004). Recebido em 13/12/2004. Reapresentado em 11/7/2005. Aprovado em 14/7/2005.

treating chronic hepatitis B. The following tools were utilized: Maastricht (M), Delphi (D), Jadad (J) and the Cochrane Collaboration (CC) method (gold standard). The Spearman correlation coefficient was used to compare the results from the three methods. The Kappa test was used to assess the concordance between the reviewers in applying the tools, and the weighted Kappa test was applied to compare the quality ranking determined by the tools. The outcomes assessed in the meta-analyses were clearance of HBV-DNA and HBeAg.

Results

The studies presented regular to low quality. The concordance between reviewers varied according to the instrument utilized: D=0.12; J=0.29; M=0.33; and CC=0.53. The correlation was moderate and homogeneous (D/J=0.51; D/M=0.53; and J/M=0.52). The meta-analysis result relating to HBV-DNA ranged from RR=0.71 (95% CI: 0.66-0.77) to RR=0.67 (95% CI: 0.58-0.79). For HBeAg, the results ranged from RR=0.85 (95% CI: 0.80-0.90) to RR=0.85 (95% CI: 0.77-0.93). These results depended on the quality of the studies included.

Conclusions

The quality assessment tools presented good correlation. In systematic reviews with the same direction of effect, the quality assessment may not significantly change the results. The Cochrane Collaboration method was the most reproducible method and easiest to apply.

INTRODUÇÃO

A assunção do paradigma do embasamento científico para a prática médica deu-se num movimento lento e gradual no correr do último século, acelerando no seu final. Isso proporcionou aumento da quantidade de artigos publicados na literatura médica. Para permitir que tais informações fossem progressivamente adotadas, utilizou-se intensamente da produção de revisões, chamadas narrativas, calcada nessa experiência concentrada.¹³

Esse processo estava sujeito à subjetividade dos autores, com relação à seleção e interpretação das informações dos tópicos clínicos relacionados, publicados e não publicados.¹⁵ Em sua maior parte, nenhum critério explícito de análise crítica da informação era empregado, fazendo com que, as conclusões e sugestões daí advindas, estivessem sujeitas a vieses e erros de várias fontes. No seu extremo, esse modelo de prática médica levou à perda de efetividade e eficiência dos serviços de saúde.¹⁸

Na tentativa de equacionar o problema da qualidade da informação e dos caminhos adotados, a aplicação do método epidemiológico à prática clínica tem se mostrado o mais consistente. Tal método é conhecido por epidemiologia clínica.⁶ Sob a égide deste paradigma, no que tange a intervenções profiláticas ou terapêuticas, a mais forte evidência deriva de estudos, nos quais vários ensaios clínicos casualizados e controlados são desenvolvidos, focando um mesmo tratamento para uma mesma condição. Segue-se não mais os preceitos de uma revisão narrativa, mas sim, os de uma revisão sistemática.¹⁸

Tal revisão se constitui num método moderno para a avaliação de um conjunto de dados simultâneos¹ realizando um sumário da literatura médica (pesquisa primária). Ela se vale de métodos explícitos e reprodutíveis para a busca sistemática, análise crítica e síntese dos estudos individuais, numa, agora, chamada revisão sistemática (pesquisa secundária).⁴

Em relação à intervenção terapêutica, o nível de evidência 1a, ou seja, a melhor evidência existente em relação a um dado efeito terapêutico, advém de revisões sistemáticas de estudos clínicos casualizados e controlados, acompanhada ou não de uma síntese estatística chamada metanálise.¹

Naturalmente, a qualidade dos estudos individuais é de óbvia relevância para a revisão sistemática e influencia a magnitude dos resultados. Se a qualidade do material primário não for adequada, isso pode falsear e tornar não verdadeiras as conclusões da revisão.¹¹

Estudos clínicos podem ser corrompidos fortuita ou deliberadamente, de várias formas: na casualização, no mascaramento do tratamento alocado, na geração aleatória da seqüência de números randômicos, na análise ou ainda quando não se implementa o mascaramento tipo duplo cego.

A necessidade de uma análise da qualidade desses estudos torna-se óbvia e diferenças no ordenamento por qualidade, resultante do uso de diferentes listas de critérios, podem causar sérios problemas na condução de uma revisão sistemática. Portanto, as evidências empíricas são necessárias para estabelecer se

os métodos de avaliação de qualidade são válidos e reprodutíveis.²⁰

O objetivo do presente estudo foi comparar os resultados obtidos pela aplicação de diferentes instrumentos de avaliação de qualidade de estudos clínicos aplicados em uma revisão sistemática sobre a eficácia do Interferon Alfa no tratamento da hepatite crônica tipo B.

MÉTODOS

Foi realizado um estudo observacional e analítico desenvolvido no contexto de uma revisão sistemática sobre a eficácia do tratamento da hepatite crônica do tipo B com o Interferon Alfa. Foram utilizadas estratégias de busca com sensibilidade e especificidade diferentes, como o sugerido pelo grupo de pesquisa hepato-biliares da Colaboração Cochrane, incluídos todos os ensaios clínicos identificados nos bancos de dados MedLine, EmBase, Lilacs e da Biblioteca da Colaboração Cochrane. O período abrangido foi de 1966 até 2001.

Os estudos clínicos relevantes foram identificados, em todos os bancos citados, usando-se os termos específicos *hepatitis B* e *Interferon* e a seguinte estratégia: *randomized controlled trial [publication Type] OR controlled clinical trial [Publication Type] OR randomized controlled trial [Mesh Terms] OR random allocation [Mesh Terms] OR double blind methods [Mesh Terms] or single blind methods [Mesh Terms]*.

Foram identificados ao todo, 2.838 estudos, sendo 46 clínicos randomizados e controlados. Desses foram retirados aqueles que se referiam apenas à comparação de doses distintas do Interferon Alfa (8), totalizando 38 estudos selecionados.

Quatro distintos métodos de avaliação de qualidade de estudos clínicos foram utilizados: de Maastricht, Delphi, Jadad e da Colaboração Cochrane.

Participaram do estudo dois pares de revisores, que trabalharam de forma independente. A aplicação do método da Colaboração Cochrane foi realizada por duas infectologistas, treinadas em epidemiologia clínica, no mesmo ano o local, e serviu de parâmetro de comparação para os outros três instrumentos. Esse método classifica os estudos em A, B ou C, de acordo com o baixa, moderada ou alta chance de viés dos estudos primários, respectivamente. Tal classificação se detém principalmente na validade interna do estudo, sua forma de aleatorização e de como contornou ou minimizou os vieses, não sendo como as outras, um instrumento estruturado e sim um escrutínio.¹⁷

O Sistema Delphi, assim chamado porque dos 206 itens associados à qualidade dos estudos inicialmente relacionados, foram reduzidos a nove por meio da técnica de consenso Delphi. Buscam avaliar três dimensões da qualidade: validade interna, validade externa e análise estatística.¹⁹

O Sistema de Jadad consta de três tópicos, diretamente relacionados com a redução de vieses (centrados na validade interna). Todas as questões têm a opção sim/não. Seu escore de qualidade possui cinco pontos: três vezes um ponto para as respostas sim, e dois pontos adicionais para métodos apropriados de randomização e sigilo de alocação.⁹

O Sistema de Maastricht consiste de 15 itens principais baseados em critérios metodológicos de avaliação de qualidade. Estes são divididos em 47 subitens, que totalizam 100 pontos, que avaliam três dimensões da qualidade de um ensaio clínico: validade interna, validade externa e método estatístico. A lista comporta quatro opções de resposta e pesos são atribuídos aos itens que reflitam importância relativa.²⁰

Optou-se pelo uso desses três métodos, por serem instrumentos diversos de avaliação, com número de itens bastante diferentes, com forma ponderada ou não de pontuação, por vezes aparentemente complementares. Assim, busca-se incluir a avaliação de um maior número possível de características, que compõem a metodologia dos estudos clínicos.

Foram realizados os cálculos do escore de qualidade para cada sistema, de acordo com os seus valores ponderados originais (Jadad e Maastricht). Para a lista Delphi, usou-se peso igual para cada item.

Para comparar os diferentes sistemas entre si, os escores de qualidade obtidos foram transformados em porcentagem, em relação ao escore máximo de cada um deles. Os percentuais obtidos em cada sistema foram categorizados em cinco classes: de 0 a 20% =1, de 21 a 40% =2, de 41 a 60% =3, de 61 a 80% =4 e de 81 a 100% =5.

Calculou-se o Coeficiente de Correlação de Spearman, entre os percentuais de qualidade obtidos pelos revisores.

O teste de Kappa foi realizado para se avaliar a concordância entre os revisores na classificação dos estudos em categorias, ao utilizarem os diferentes métodos. Posteriormente, foram realizadas reuniões de consenso para se chegar à pontuação definitiva de cada Estudo Clínico Randomizado Controlado (ECRC).

Após obterem as pontuações consensuais, os estudos foram ordenados de forma decrescente em termos da qualidade, tendo como referência a obtida após a aplicação do sistema Maastricht.

Utilizou-se o Teste de Kappa ponderado para se comparar a ordenação de qualidade definida pela aplicação de cada um dos sistemas de avaliação.

Para a análise do efeito da intervenção contra a não intervenção, determinou-se que seriam considerados os melhores, aqueles estudos que obtivessem um escore igual ou melhor que 50% do possível, em cada instrumento, em pelo menos dois dos instrumentos de qualidade utilizados. Para as epidemiologistas clínicas, foram considerados os melhores aqueles classificados como A ou B.

O pacote estatístico *Review Manager 4.1* foi utilizado para a realização da metanálise. Como só foram utilizadas variáveis dicotômicas, foram calculados os riscos relativos com intervalo de confiança de 95%.

Para avaliar a estimativa da magnitude e/ou direção do efeito do uso do Interferon Alfa no tratamento da hepatite crônica pelo vírus B, foram incluídos na metanálise os estudos que relatavam como desfechos primários, o clareamento do DNA do vírus B (HBV-DNA) e/ou do antígeno e do vírus B (HbeAg), após seis meses do final do tratamento.

Dada a heterogeneidade identificada entre os estudos clínicos incluídos, foi utilizado o modelo de efeito randômico para a realização da metanálise.

RESULTADOS

Em relação aos instrumentos utilizados, o Delphi e Jadad foram considerados de fácil aplicação. O de Maastricht ofereceu maiores problemas e requereu adaptações. O da Colaboração Cochrane, por não ser estruturado, dependem de prévia familiaridade com o método epidemiológico clínico.

Dos 38 estudos selecionados, apenas seis (15,7%) foram divulgados após 1996, ano de publicação do

Tabela 1 - Ordenamento de qualidade dos estudos sobre tratamento da hepatite B crônica, com Interferon Alfa, após consenso, dispostos de forma decrescente.

| N | Autor* | Ano | Maastricht (rank) | Delphi (rank) | Jadad (rank) | EC (rank) |
|----|-----------------------|------|-------------------|---------------|--------------|-----------|
| 1 | Sokal / Conjeevaran | 1998 | 64 (1) | 5 (1) | 3 (1) | A (1) |
| 2 | Jansen / Gerken | 1999 | 62 (2) | 4 (2) | 3 (1) | A (1) |
| 3 | Brook / Macdonald | 1989 | 60 (3) | 3 (3) | 3 (1) | A (1) |
| 4 | Fattovich / Farci | 1992 | 58 (4) | 4 (2) | 3 (2) | B (2) |
| 5 | Porres / Carreño | 1988 | 57 (5) | 4 (2) | 2 (2) | B (2) |
| 6 | Di-Bisceglia / Fong | 1993 | 56 (6) | 4 (2) | 3 (2) | B (2) |
| 7 | Hoofnagle / Peters | 1988 | 56 (6) | 4 (2) | 3 (2) | B (2) |
| 8 | Sarin / Guptan | 1996 | 56 (6) | 4 (2) | 3 (1) | A (1) |
| 9 | Lok / Wu | 1988 | 55 (7) | 4 (2) | 3 (3) | C (3) |
| 10 | Pastore / Santantonio | 1988 | 55 (7) | 4 (2) | 2 (2) | B (2) |
| 11 | Saracco / Mazzela | 1989 | 55 (7) | 4 (2) | 3 (2) | B (2) |
| 12 | Carreño / Marcelin | 1999 | 54 (8) | 4 (2) | 2 (2) | B (2) |
| 13 | Muller / Baungarten | 1990 | 54 (8) | 3 (3) | 2 (2) | B (2) |
| 14 | Ruiz-Moreno / Rua | 1991 | 54 (8) | 3 (3) | 2 (2) | B (2) |
| 15 | William / Graig | 1990 | 54 (8) | 4 (2) | 2 (3) | C (3) |
| 16 | Fattovich / Broilo | 1989 | 52 (9) | 3 (3) | 3 (1) | A (1) |
| 17 | Ruiz-Moreno / Jimen | 1990 | 52 (9) | 3 (3) | 3 (2) | B (2) |
| 18 | Wong / Yim | 1995 | 52 (9) | 4 (2) | 2 (1) | A (1) |
| 19 | Carreño / Porres | 1987 | 51 (10) | 3 (3) | 3 (2) | B (2) |
| 20 | Carreño / Porres | 1991 | 51 (10) | 3 (3) | 3 (3) | B (2) |
| 21 | Pastore / Millella | 1992 | 51 (10) | 4 (2) | 2 (2) | B (2) |
| 22 | Realdi / Fattovich | 1990 | 51 (10) | 2 (4) | 4 (2) | B (2) |
| 23 | Alexander / Brahn | 1986 | 50 (11) | 3 (3) | 3 (2) | C (3) |
| 24 | Barbera / Bortolotti | 1994 | 50 (11) | 3 (3) | 3 (2) | A (1) |
| 25 | Brunetto / Olivieri | 1989 | 50 (11) | 3 (3) | 3 (2) | A (1) |
| 26 | Dusheiko / Paterson | 1986 | 50 (11) | 3 (3) | 3 (1) | C (3) |
| 27 | Perrilo / Schiff | 1990 | 50 (11) | 4 (2) | 2 (1) | A (1) |
| 28 | Brook / Chan / Yap | 1989 | 48 (12) | 5 (1) | 1 (2) | C (3) |
| 29 | Rumi / Romeu | 1993 | 47 (13) | 3 (3) | 3 (2) | B (2) |
| 30 | Tchrveniakova / Radev | 1999 | 47 (13) | 3 (3) | 3 (3) | C (3) |
| 31 | Utilli / Sagnelli | 1991 | 45 (14) | 2 (4) | 4 (3) | C (3) |
| 32 | Thomas / Lok | 1994 | 44 (15) | 4 (2) | 2 (2) | C (3) |
| 33 | Lampertico / Delninn | 1997 | 43 (16) | 4 (2) | 2 (1) | C (3) |
| 34 | Waked / Amin | 1990 | 39 (17) | 3 (3) | 3 (2) | C (3) |
| 35 | Hadzyianes / Bramou | 1990 | 36 (18) | 3 (3) | 3 (2) | C (3) |
| 36 | Mazzela / Villanova | 1988 | 34 (19) | 3 (3) | 3 (2) | C (3) |
| 37 | Lok / Lai | 1986 | 32 (20) | 2 (4) | 4 (2) | C (3) |
| 38 | Barbara / Mazzela | 1986 | 30 (21) | 2 (4) | 4 (2) | C (3) |
| | Mínimo/ Máximo | | 30 64 | 2 5 | 1 3 | |
| | Escore Médio (DP) | | 50,13 (7,7) | 3,42 (0,75) | 2,28 (0,61) | |

*A lista dos artigos analisados pode ser obtida mediante solicitação ao primeiro autor do presente artigo
DP: Desvio-padrão; EC: Epidemiologistas clínicas

CONSORT,³ consenso que buscou regular a realização e publicação de estudos clínicos. Todos se definiam como randomizados e, de fato, foram submetidos a algum método de aleatorização. Doze (31,5%) deles descreveram sua seqüência e metodologia. Em apenas um dos artigos, foi descrito o controle por um placebo, um complexo de vitaminas, o de Tchreniakova/ Radev de 1999, porém sem a descrição se foi ou não utilizado de forma análoga ao placebo.

Em apenas sete (18,4%) dos artigos, houve o relato do cálculo amostral.

Em 22 (57,8%) dos artigos, ocorreu a identificação de perdas e retiradas e de suas causas. Na análise detalhada de todos eles, em muitos dos quais não foram relatadas perdas, estas de fato foram constatadas, e por vezes, numa porcentagem maior do que 10%, introduzindo aí um importante potencial viés na conclusão dos mesmos.

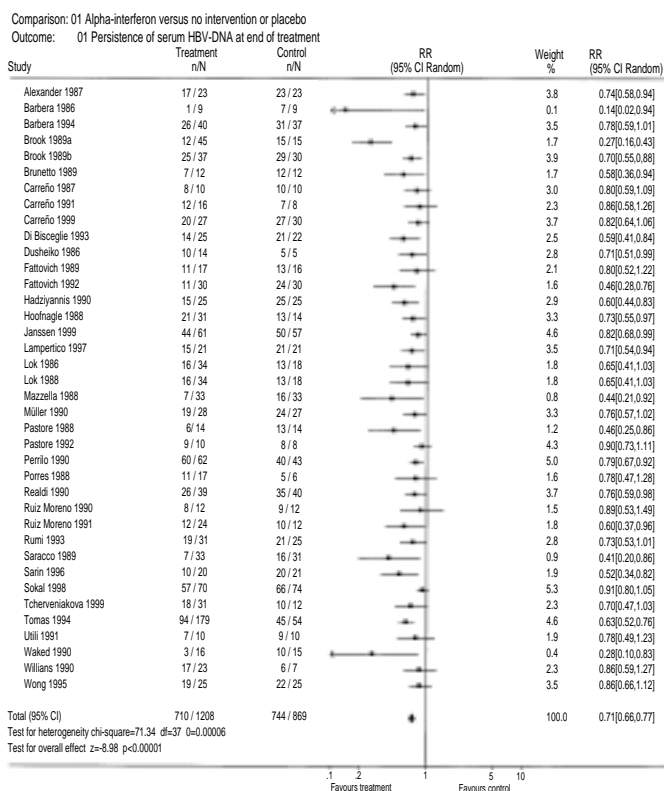
Em 23 (60,5%) dos artigos, a análise se deu pela “intenção de tratar”, ou seja, foram considerados na análise todos os pacientes que, de fato, iniciaram os estudos no grupo de intervenção ou de controle, e não apenas os que terminaram.

Em 27 (71%) dos artigos, as perdas relatadas foram iguais ou menores do que 10% dos pacientes randomizados, número geralmente considerado aceitável para se evitar conclusões equivocadas.

A Tabela 1 apresenta a disposição decrescente dos escores de qualidade dos instrumentos utilizados, após consenso entre revisores.

A Tabela 2 mostra o coeficiente de concordância (Kappa), na aplicação dos instrumentos.

O coeficiente de correlação de Spearman entre os instrumentos utilizados mostrou que a correlação obtida entre os métodos Maastricht e Jadad, foi de 0,52 (p<0,0005), entre Maastricht e o instrumento Delphi de 0,53 (p<0,0007), e a correlação entre os métodos Delphi e Jadad foi de 0,51 (p=0,0001). Ainda que



Obs.: Figura apresentada em inglês, pois pelo pacote estatístico utilizado, os gráficos não puderam ser re-editados.

Figura 1 - Clareamento do HBV-DNA, ao final do tratamento com Interferon Alfa, incluindo todos os estudos (38).

seja uma correlação apenas moderada, foi bastante homogênea e muito maior do que a esperada apenas pelo acaso, portanto, estatisticamente significativa.

Quando comparado o ordenamento de qualidade dos estudos (rank), obtidos pela aplicação dos sistemas estruturados, por meio da aplicação do teste de Kappa ponderado, obteve-se concordância combinada de 0,16 considerada estatisticamente significativa (p=0,02).

A Figura 1 mostra a metanálise do efeito da intervenção versus a não intervenção em relação ao desfecho clareamento do HBV-DNA, incluindo todos os estudos, sem a avaliação de qualidade. A Figura 2 apresenta a análise agrupada quando só os estudos de melhor qualidade foram incluídos. O risco relativo (RR) e o intervalo de confiança de 95% (IC 95%) quando se incluíram só os piores (n=25) foi respectivamente, RR=0,73; IC 95%: 0,68-0,78.

Tabela 2 - Coeficiente de concordância (Kappa) de estudos sobre o tratamento da hepatite crônica B com Interferon Alfa.

| Método | Concordância observada | Concordância esperada | Kappa | DP | p |
|------------|------------------------|-----------------------|-------|------|--------|
| Delphi | 52,63 | 45,98 | 0,12 | 0,12 | 0,15 |
| Jadad | 55,26 | 36,91 | 0,29 | 0,10 | 0,002 |
| Maastricht | 78,95 | 68,35 | 0,33 | 0,12 | 0,004 |
| Cochrane | 68,42 | 31,72 | 0,53 | 0,10 | 0,0001 |

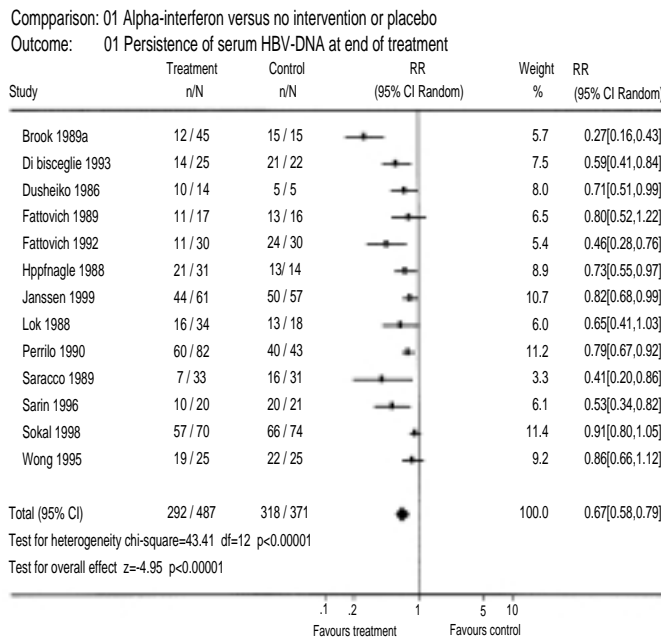


Figura 2 - Clareamento do HBV-DNA ao final do tratamento com Interferon Alfa, incluindo os estudos que obtiveram $\geq 50\%$ do escore nos instrumentos de avaliação de qualidade (13).

A metanálise realizada com os estudos considerados A e B pelas infectologistas utilizando-se instrumento sugerido pela Colaboração Cochrane determinaram um RR=0,71; IC95%: 0,65-0,78, pouco diferente dos anteriores.

A Figura 3 apresenta o efeito do tratamento quando se considerou o clareamento do HbsAg quando todos os estudos são incluídos na metanálise (n=23). Quando só os melhores são incluídos (n=11) obteve-se RR=0,85; IC 95%: 0,77-0,93, só diferindo portanto no intervalo de confiança.

DISCUSSÃO

Os melhores artigos foram os que relataram na sua publicação, os procedimentos de randomização, seu cálculo de amostra, o relato das perdas e suas causas e a análise pela intenção de tratar. Principalmente, aqueles que relataram de forma bastante uniforme, uma perda de menos de 10% do total de pacientes aleatorizados, diminuindo a chance de viés de análise e conclusões equivocadas.

Naturalmente, os piores tiveram essas características em percentagem muito menor ou não as apresentaram.

Observou-se uma correlação moderada e

semelhante entre 0,51 e 0,53, estatisticamente significativa entre os instrumentos utilizados para avaliar a qualidade dos estudos. Isso, sem dúvida, é consequência da presença de um núcleo comum de itens que buscam analisar a validade interna dos estudos. Todas as três listas contêm itens que averiguam os procedimentos de aleatorização e a presença/ausência do sigilo de alocação, por exemplo.

Ao comparar-se a reprodutibilidade entre os revisores na aplicação desses instrumentos, medida pelo Kappa, ainda que tenha sido maior que a esperada, não atingiu significância estatística para o sistema Delphi. De alguma forma é um dado inesperado, pois é um sistema que contempla a avaliação das dimensões envolvidas com a qualidade dos estudos clínicos, sendo de fácil compreensão e rápida aplicação. Três de seus itens, os referentes a presença ou não de controle com placebo, simplesmente não foram aplicáveis, visto um só dos estudos incorporados à revisão foi controlado por placebo. Isso, por si só, diminuiria a chance de discordância entre os revisores.

Os números mostram que entre as epidemiologistas e o método por elas adotado para a avaliação da qualidade dos estudos, o coeficiente de concordância

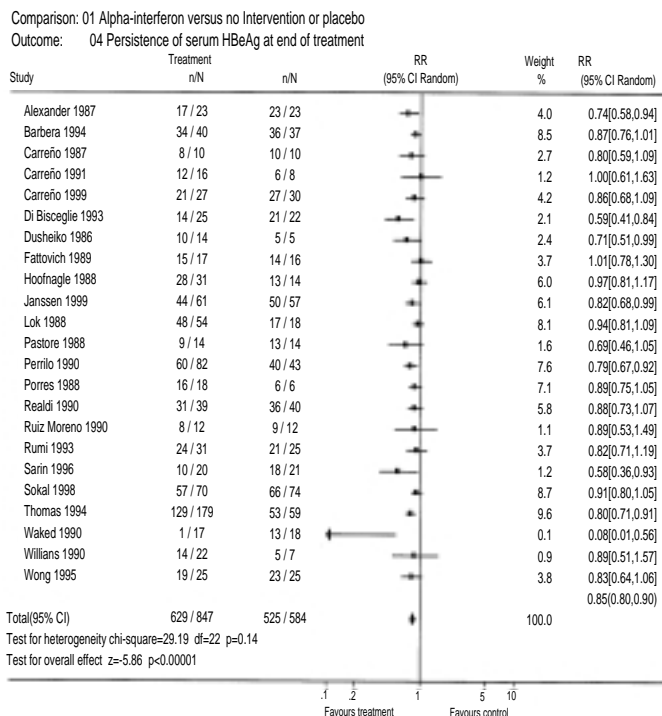


Figura 3 - Clareamento do HBeAg ao final do tratamento com Interferon Alfa, incluindo todos os estudos que avaliaram este desfecho (23).

cia (Kappa) foi moderado (0,53) e estatisticamente significativa ($p=0,0001$).

A diferença observada entre os revisores com a utilização dos instrumentos de avaliação de qualidade e as epidemiologistas, com o método sugerido pela Colaboração Cochrane, pode ser explicada pelo menor número de itens e praticidade do segundo. Uma outra seria que a avaliação de escore é aplicada baseando em se uma característica foi reportada (como os pacientes foram alocados, por exemplo), mais do que se de fato foi realizado adequadamente no estudo.¹⁷

O presente estudo mostrou que, dependendo do tipo de escala utilizada para avaliar a qualidade das publicações, se definirão diferentes *ranks* de qualidade entre os estudos incorporados à revisão sistemática. Isso já havia sido constatado por outros investigadores.¹⁰

O resultado do teste de Kappa ponderado, estatisticamente significativo ($p=0,02$), permite concluir que apesar da pequena concordância obtida, (Kappa=0,16), o ordenamento de qualidade dos estudos foi semelhante.

Ao se incluir todos os estudos e se medir o efeito no clareamento do HBV-DNA, obteve-se RR=0,71; IC 95%: 0,66-0,77. Ao se incluir apenas os postulados como melhores o RR=0,67; IC 95%: 0,58-0,79, portanto uma diferença de risco de 33%. Ao se testar somente os que não atingiram o limiar de qualidade, o RR foi de 0,73; IC 95%: 0,68-0,76, muito próximo dos anteriores e na direção de um menor efeito da intervenção (27%).

A literatura, de forma mais regular, aponta para uma superestimação do efeito ao se introduzir os estudos menores na metanálise, pois são habitualmente desenvolvidos com menor rigor metodológico.⁴

O mesmo ocorreria ao se introduzir estudos com menor escore de qualidade.¹² Por outro lado, quando se buscou comparar o efeito do tratamento nos estudos de baixa e alta qualidade, não se identificou correlação entre medida da qualidade e magnitude do efeito do tratamento, em particular com relação às variáveis duplo-cego e sigilo de alocação. No entanto não descarta a possibilidade de que certas medidas de qualidade possam se associar com o efeito do tratamento, em situações clínicas ou questões de interesse específico.²

O resultado da metanálise, utilizando-se os melhores artigos (A+B) definidos pelas epidemiologistas, aferindo o mesmo desfecho (RR=0,72; IC 95%: 0,64-0,80), foi muito semelhante ao obtido pela escrutinização, pelos outros três instrumentos de ava-

liação de qualidade. Isso reforça a eficácia do método proposto pela Colaboração Cochrane, para a avaliação de qualidade dos estudos clínicos.

Analisando a magnitude do efeito, em relação ao segundo desfecho examinado, o clareamento do HBeAg, os resultados mostram que os mesmos só diferem com relação ao intervalo de confiança, sendo a magnitude do efeito aferido menor do que do desfecho anterior.

Todos os artigos analisados determinaram um desfecho a favor da intervenção. Assim, a comparação dos melhores contra todos ou contra os piores, não alterou a direção do efeito, mantendo-o com diferentes magnitudes, mas sempre a favor da intervenção.

Pode-se levantar a hipótese de que, ao se realizar uma revisão sistemática e metanálise, com estudos discordantes em relação à direção do efeito, existirá maior probabilidade de se observar variação do efeito da intervenção, na dependência da qualidade dos estudos incluídos.

A inclusão dos artigos, tanto os bons quanto os de má qualidade, nas revisões sistemáticas, pode aumentar a sua heterogeneidade (a extensão na qual os resultados de ECRC “similares” divergem entre si). Isso pode não ser necessariamente um mal, dado que a heterogeneidade é uma realidade na prática clínica.

Tal aspecto já havia sido levantado anteriormente, destacando que, se todos os ECRC se ocuparem de pacientes com idade, sexo, local e intervenção terapêutica uniformes, então a análise agrupada estimada deveria ser muito precisa, porém sua conclusão só se aplicaria a grupo e locais similares. Sua validade externa seria reduzida.¹⁶

Ao se desenvolver estudos secundários, como uma revisão sistemática, e na avaliação de qualidade de estudos primários, deve-se levar em conta algumas características que trazem conseqüências para as modernas revisões. Uma delas é de que os resultados de estudos clínicos mais antigos podem ser menos confiáveis, pois foram conduzidos num momento em que a metodologia era menos rigorosa e reprodutível. Portanto, podem não ser comparáveis aos estudos mais modernos, no que se refere, por exemplo, aos desfechos e forma de medi-los.⁷

Por outro lado, estudos que oferecem uma descrição inadequada dos testes realizados ou da população estudada, ainda que seja plausível acreditar que revelem deficiências de desenho e escondam vieses, nem sempre é o que de fato acontece. Muitas

vezes não é possível distinguir entre má qualidade do desenho e condução da pesquisa, e má qualidade da descrição, que pode se dar, por exemplo, pelo exíguo espaço disponível para publicação nas revistas de referência.⁴

Os procedimentos necessários para a realização desse ensaio, documentam que não é simples a avaliação de qualidade dos ECRC a serem incluídos numa revisão sistemática e metanálise. A menos que um consenso, na definição de medida da qualidade, seja construído de forma robusta e validado, a concordância na sua aplicação pode chegar a ser inaceitavelmente baixa.

Considerem que epidemiologistas treinadas, valendo-se de uma metodologia mais simples, da Colaboração Cochrane, para a classificação de qualidade dos estudos primários, chegaram praticamente ao mesmo resultado.

Está claro que, até que os consensos para o desenvolvimento e publicação de estudos clínicos randomizados e controlados, sejam assumidos de forma irrestrita³ assim como a forma de relatá-los, que a avaliação da qualidade metodológica dos ECRC deva ser considerada como procedimento de rotina, na realização de revisão sistemática e metanálise.¹⁴

A avaliação da qualidade dos ensaios clínicos

aleatórios oferece uma estimativa da probabilidade de que seus resultados expressem a verdade. Isso não exclui que se devam realizar procedimentos de avaliação individuais e metuculosos, principalmente com relação à geração da aleatorização e condução dos procedimentos de alocação para a intervenção, o sigilo da distribuição e avaliação do efeito da intervenção, assim como a adesão ou não com o tratamento, e de como as perdas são manejadas na análise.

Pesquisadores pouco éticos ou despreparados e, mesmo parte das indústrias farmacêuticas, utilizando-se das mais variadas escalas de qualidade, podem fazer com que seus ECRC atinjam o máximo dos escores possíveis. Assim, podem burlar premissas, objetivos, condutas, análises e o momento adequado para disseminação da informação, necessária à boa prática clínica.⁸

Talvez a melhor maneira de se superar o problema seja a presença cada vez maior da explicitação de potenciais conflitos de interesses no desenvolvimento e publicação dos estudos. Adicionalmente, de um progressivo fomento público a pesquisadores bem treinados, não atrelados a outros objetivos que não a busca do bem estar da maioria da população, e que pacientemente desenvolvam detalhada e criteriosa análise do desenho e condução dos estudos clínicos, na construção e fortalecimento da Medicina Baseada em Evidências.

REFERÊNCIAS

1. Atallah AN. Revisão sistemática da literatura médica e metanálise. *Diagn Tratamento* 1997;2(2):12-5.
2. Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JPA, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002; 287(22):2973-82.
3. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 1996;276(8):637-9.
4. Berlin JA, Rennie D. Measuring the quality of trials: the quality of quality scales. *JAMA* 1999;282(11):1083-5.
5. Chalmers I, Haynes B. Reporting, updating, and correcting systematic reviews of the effects of health care. *BMJ* 1994;309(6958):862-5.
6. Fletcher RH, Fletcher SW, Wagner EH. Clinical epidemiology: the essentials. 3rd ed. Baltimore, MD: Williams & Wilkins; 1996.
7. Hopayian K. The need for caution in interpreting high quality systematic reviews. *BMJ* 2001;323(7314):681-4.
8. Ioannidis JPA, Lau J. Can quality of clinical trials and meta-analyses be quantified? *Lancet* 1998(9128):352:590-1.
9. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17(1):1-12.
10. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282(11):1054-60.
11. Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001;323(7303):42-6.
12. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analysis. *Ann Intern Med* 2001;135(11):982-9.

13. Leite AJM. Medicina baseada em evidências: um exemplo no campo da pediatria. *J Pediatr* (Rio de J) 1999;75(4):215-26.
14. Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001;134(8):657-62.
15. Mulrow CD. The medical review article: state of science. *Ann Intern Med* 1987;106(3):485-8.
16. Naylor CD. Two cheers for meta-analysis: problems and opportunities in aggregating results of clinical trials. *CMAJ* 1988;138(10):891-5.
17. Oxman A, Clarke M, editors. Cochrane reviewers' handbook 4.1.1: updated december 2000. In: Assessment of study quality. Oxford: The Cochrane Library; 2001. p. 39-50.
18. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. 2nd ed. London: Churchill Livingstone; 2000.
19. Verhagen AP, Vet HC, Bie RA, Kessels AG, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol* 1998;51(12):1235-41.
20. Verhagen AP. Quality assessment of randomized clinical trials. In: Quality assessment of trials. Maastricht: Database; 1999. p. 9-11.
21. Verhagen AP, Bie RA, Lenssen AF, Vet HC, Kessels AGH, Boers M, et al. Quality assessment of trial: a comparison of three criteria lists. *Phys Ther Rev* 2000;5:49-58.