

O significado do grau de recomendação e força de evidência A da classificação da Associação Médica Brasileira

Aroldo Fernando Camargos
Editor científico da revista Femina

A revista Femina vem utilizando os graus de recomendações e força de evidência da Associação Médica Brasileira (AMB)¹ para classificar as referências bibliográficas contidas no texto de seus artigos com o objetivo de dar ao leitor o verdadeiro valor atribuído a cada intervenção utilizada nas investigações (Quadro 1).

Quadro 1 - Grau de recomendação e força de evidência da AMB

- | |
|--|
| <p>A: Estudos experimentais ou observacionais de maior consistência (metanálises ou ensaios clínicos randomizados).</p> <p>B: Estudos experimentais ou observacionais de menos consistência (outros ensaios clínicos não-randomizados ou estudos observacionais ou estudos caso-controle).</p> <p>C: Relatos ou séries de casos (estudos não-controlados).</p> <p>D: Opinião desprovida de avaliação crítica, baseada em consensos, estudos fisiológicos ou modelos animais.</p> |
|--|

A proposta do meu editorial é oferecer ao leitor da Femina a fundamentação de como é constituído o nível A de evidência da AMB; ocupo-me apenas de analisar a qualidade do desenho, a interpretação dos seus resultados e sua aplicabilidade na prática clínica; para um conhecimento mais apurado sobre revisão sistematizada e metanálise, sugiro a leitura de duas publicações da Femina.^{2,3}

O grau de recomendação e o nível de evidência A são obtidos por pelo menos um estudo adequadamente desenhado, randomizado e controlado.^{1,4}

As investigações clínicas são estudos experimentais que comparam uma intervenção específica, por exemplo, uma nova medicação para tratar uma determinada doença, com uma intervenção alternativa, placebo ou nenhum tratamento.

Os objetivos da pesquisa têm de ser bem definidos e estar claramente indicados; a seleção dos investigados deve respeitar critérios bem determinados e a alocação de pacientes deve ser realizada de maneira randomizada e preferencialmente cega, ou seja, os investigadores e os sujeitos da pesquisa não devem saber a que tipo de tratamento cada indivíduo será submetido, e essa é uma grande marca da qualidade do desenho do estudo; devem também ser descritos os critérios de inclusão e exclusão, gênero, idade e origem étnica dos sujeitos da pesquisa, estágio da doença, dose e via de administração do medicamento, períodos do tratamento e do seguimento dos indivíduos pesquisados. Todas essas variáveis mostram se a intervenção é relevante para a prática clínica.⁴

O estudo tem que ser adequadamente avaliado por meio das seguintes perguntas:^{5,6}

1. O estudo tem um bom desenho metodológico? Essa pergunta deve ser respondida com outras perguntas como as exemplificadas abaixo:
 - Como as pacientes foram randomizadas? As randomizações mais seguras são feitas através da internet por um serviço denominado *Randomization Service for Multi-Center Clinical Trials*,⁷

por telefone⁸ e por meio do emprego de programas de computador. Deve-se evitar a randomização de pacientes com empregos de moedas (cara ou coroa), urnas, listas, número final do prontuário (par ou ímpar), já que esses métodos estão sujeitos a viés. A randomização não garante o balanço de distribuição de variáveis de confusão, mas é essencial para prevenir o viés. Esse termo é usado para expressar o erro sistemático ou a tendenciosidade e, por extensão de sentido, usa-se a palavra “viés” para designar qualquer comentário ou análise tendenciosa, isto é, que não respeite os princípios da imparcialidade;

- o seguimento foi adequado em relação ao tempo, completo e suficientemente longo? A perda do seguimento não pode ser maior do que 20% dos sujeitos sob a pena de comprometer seriamente a validade dos resultados. Valores inferiores a 5% de perda são ótimos. O número de perdas entre os grupos deve ser balanceado, e é preciso checar se os resultados dos sujeitos que se retiraram dos estudos causaram significativo impacto nos resultados finais. A apropriada duração de um seguimento é importante, mas esta variável tem um caráter subjetivo;
2. Todos os pacientes foram analisados no grupo em que eles foram alocados? Uma importante consideração é se os sujeitos randomizados nos grupos controle e tratamento foram incluídos na análise chamada “intenção de tratar”;
 3. O tamanho amostral foi adequado? A investigação deve possuir um número de indivíduos suficiente para demonstrar diferenças estatisticamente significativas entre os grupos intervenção (GI) e controle (GC). Para que uma intervenção seja efetiva, tem que ser produzida por chance e não por evento acidental ou viés, que é o desvio sistemático da causa da verdade produzido por outros fatos estranhos à intervenção. A probabilidade de que a diferença entre os grupos se deva à chance (erro tipo I) é comumente estabelecida em 1/20 ou 5%. As medidas estatísticas dos testes do erro tipo I têm ocorrência expressa em valores de p (significativo quando menor que 0,05) e ou de intervalo de confiança (significativo quando não passa pelo número 1). O intervalo de confiança (IC) estima a possível variação de valores dentro de uma população e indica que tais medidas são esperadas se o estudo for repetido, com probabilidade aproximada de 95%;
 4. O estudo tem poder adequado? A probabilidade por chance de o estudo não ser capaz de detectar uma diferença estatisticamente significativa é estabelecida em 0,1 a 0,2 (erro tipo II), ou seja, aceita-se que há uma chance entre 10 a 20% de que o real efeito do tratamento exista, mas não foi encontrado;
 5. Como os resultados são expressos? Por exemplo: o estudo randomizado e placebo-controlado utilizando o medicamento X para o tratamento da doença Y mostrou os seguintes resultados:
 - Grupo Intervenção (GI): cura de 70 indivíduos em uma amostra de 120 pessoas tratadas;
 - Grupo Placebo (GC): cura de 20 indivíduos em uma amostra de 110 pessoas tratadas;
 - taxa de cura do GI = $70/120 = 0,58$;
 - taxa de cura do GC = $20/110 = 0,17$;
 - risco relativo (RR): é a razão da probabilidade de sucesso do tratamento experimental sobre a probabilidade de sucesso no Grupo Controle;
 - $RR = \text{taxa de cura do GI} / \text{taxa de cura GC} = 0,58/0,17 = 3,4$;

- razão da diferença (RD): é a diferença absoluta entre a probabilidade de sucesso do tratamento experimental e a probabilidade de sucesso no Grupo Controle;
- $RD=0,58-0,17=0,41$;
- número que precisa ser tratado (NPT): o número a mais de sujeitos que devem ser tratados com medicação em relação ao controle para atingir o objetivo (no exemplo em questão, o objetivo é a cura da doença Y);
- $NPT=1/\text{risco da diferença } -1/0,41=2,4$;
- *Odds Ratio* (OR): a razão da chance do sucesso no GI sob a chance de sucesso do GC;
- $OR=70 \times 120/20 \times 110=3,4$;

Os valores acima somente são válidos quando seguidos do cálculo do intervalo de confiança (IC). Exemplos de intervalo de confiança:

- a administração de citrato de clomifeno (CC) aumenta a taxa de gravidez quando comparada ao emprego de placebo em portadoras de ovários policísticos (OR: 5,8; IC95%: 1,6 a 21,5; NPT: 5,9; IC95%: 3,6 a 16,7);⁹ a interpretação desses achados é: as pacientes do grupo que usou CC têm 5,8 vezes mais chances de engravidarem do que as pacientes do GC; cada 5,9 pacientes tratadas com CC produzem uma gravidez a mais em relação ao GC, e esses números são estatisticamente significativos, pois os valores do IC não passam pelo número de 1;
 - a administração de CC ou tamoxifeno em portadoras de ovários policísticos produz a mesma taxa de gravidez (OR: 1,0; IC95%: 0,5 a 2,1); o IC passa pelo número 1, o que indica que o sucesso do tratamento entre os grupos não é estatisticamente diferente;⁹
6. O resultado do estudo é clinicamente importante? Os achados são suficientemente importantes para que sejam recomendados ao tratamento de pacientes? O efeito do tratamento foi longo o suficiente para ser clinicamente relevante? O efeito do tratamento é preciso? As conclusões foram baseadas em questões colocadas e resultados obtidos? Os resultados são relevantes para a sua prática diária? A população estudada é similar às pacientes da prática diária? As intervenções são reproduzíveis e passíveis de serem aplicadas na rotina clínica? Quais são os potenciais riscos em benefícios para as pacientes do tratamento? Existe algum tratamento alternativo disponível?

Uma vez validado o estudo, é hora de medir o peso das suas conclusões, verificar a sua importância clínica. As investigações devem produzir conclusões baseadas nos objetivos primários e não nos secundários; os trabalhos de pesquisas produzem evidências diretas quando se baseiam em resultados exatamente dirigidos a uma questão (objetivo primário) e indiretas quando a evidência se baseia na extrapolação de outros resultados relevantes; os achados de uma investigação têm que refletir o verdadeiro efeito produzido pela intervenção, e não pode ser um artefato. A intervenção também deve ser clinicamente importante e útil para os pacientes na prática diária, e o seu efeito primário deve ser suficientemente diferente do alternativo para que seja recomendado à paciente.

Quando se consideram testes de diagnóstico, as seguintes perguntas devem ser feitas: o teste é capaz de identificar a doença no grupo-alvo? Existe alguma evidência de que o teste diminui a mortalidade ou morbidade dos pacientes? Qual a prevalência da doença no grupo-alvo? O risco é identificado com segurança em outros grupos? Qual foi a sensibilidade e a especificidade do teste? O teste foi realizado da mesma

maneira pelos diversos investigadores? Quantas pessoas foram identificadas e tratadas com a aplicação do teste? Quais foram os efeitos adversos do rastreamento?

Por fim, o grau de recomendação e a força de evidência de uma intervenção só podem ser feitos após as avaliações criteriosas do desenho metodológico, resultados e conclusões do estudo no qual foi investigada; portanto, a classificação de um artigo pelos critérios da AMB não pode ser baseada em seu título ou resumo.

Leituras suplementares

1. Associação Médica Brasileira. [Homepage on the internet] [cited 30 nov. 2009]. Available from: <http://www.amb.org.br>
2. Bernardo WM. A Revisão sistemática na prática clínica baseada em evidência. *Femina*. 2008;36(6):335-44.
3. Parente RCM, Coutinho ESF, Oliveira MAP, Celeste RK, Mendes PH, Lasmar RB, *et al.* Revisão sistemática e metanálise de intervenções terapêuticas: como melhor entendê-las? *Femina*. 2008;36(5):265-74.
4. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D. Methods Work Group, Third US Preventive Services Task Force. Current methods of the US Preventive Services Task Force: a review of the process. *American Journal of Preventive Medicine*. 2001;20(Suppl 3):21-35.
5. World Health Organization. Medical eligibility criteria for contraceptive use. 3rd ed. WHO: Geneva; 2004.
6. The Practice Committee of the American Society for Reproductive Medicine. Interpretation of clinical trial results. *Fertil Steril*. 2008;90(Suppl 3):S114-20.
7. Web-Based Patient Randomization Service for Multi-Center Clinical Trials. [Homepage on the Internet] [cited 30 nov. 2009]. Available from: <http://www.randomizer.at/demo.php>
8. IVR Clinical Concepts. [Homepage on the internet] [cited 30 nov, 2009] Available from: <http://www.ivrc.com/recruitment.htm>
9. Brown J, Farquhar C, Beck J, Boothroyd C, Proctor M, Hughes E. Oral anti-oestrogens and medical adjuncts for subfertility associated with anovulation. *Cochrane Database of Systematic Reviews*. The Cochrane Library, Issue 3, Art. No CD002249. DOI: 10.1002/14651858.CD002249.pub1