

**Comissão de Avaliação de Tecnologias de Saúde
(CATS)**

***Metodologia para Avaliação
Farmacoterapêutica***

Versão 2.0 de 23 novembro de 2016

Índice

| | |
|---|-----------|
| CAPÍTULO I | 3 |
| I – Operacionalização da Avaliação | 3 |
| CAPÍTULO II | 4 |
| II – Metodologia geral | 4 |
| 1. Medidas de eficácia terapêutica e segurança | 4 |
| 2. A relevância da certeza dos resultados | 4 |
| 3. A ligação entre a certeza dos resultados e a proximidade às condições do dia-a-dia | 5 |
| 4. Avaliação de benefício da intervenção em análise | 5 |
| CAPÍTULO III | 15 |
| III – Valor terapêutico acrescentado | 15 |
| 1. Determinação da existência de valor terapêutico acrescentado | 15 |
| 2. Classificação da extensão do valor terapêutico acrescentado | 15 |
| CAPÍTULO IV | 19 |
| IV - Estudos de superioridade, não inferioridade, e equivalência: definições e critérios para mudança de objetivos | 19 |
| 1. Demonstração de equivalência | 19 |
| 2. Demonstração de não inferioridade | 20 |
| 3. Estudos de superioridade | 20 |
| 4. Relevância da pré-definição do Δ nos estudos de não inferioridade e de equivalência | 21 |
| 5. Relevância da pré-definição do estudo como de superioridade, não inferioridade ou equivalência | 22 |
| 6. É possível mudar o objetivo de uma comparação? | 22 |
| CAPÍTULO V | 24 |
| V – Comparações diretas e indiretas | 24 |
| 1. Definições | 24 |
| 2. Comparações diretas | 24 |
| 3. Comparações indiretas | 25 |
| CAPÍTULO VI | 27 |
| VI – Análises de sub-grupos | 27 |
| CAPÍTULO VII | 28 |
| VII – Elaboração e Emissão do Relatório de Avaliação Farmacoterapêutica pelo Grupo de Avaliação da Evidência | 28 |
| 1. Notas sobre a elaboração do relatório farmacoterapêutico | 28 |
| REFERÊNCIAS | 30 |

CAPÍTULO I

I – Operacionalização da Avaliação

1. O Grupo de Avaliação deve, no prazo de uma semana após aceitar participar na avaliação, enviar à Comissão Executiva da CATS uma proposta para essa avaliação contendo os seguintes elementos:
 - 1.1. Medidas de eficácia terapêutica e segurança;
 - 1.2. Subpopulações/sub-indicações a analisar (se aplicável);
 - 1.3. Medicamento comparador ou medicamentos comparadores a utilizar. O comparador ou alternativa de referência selecionado deve cumprir cumulativamente os seguintes requisitos:
 - 1.3.1.1. Ser utilizado habitualmente na prática clínica;
 - 1.3.1.2. Estar validado para a respectiva indicação/população por evidência científica em relação à sua eficácia e segurança;
 - 1.3.1.3. Caso existam mais do que uma terapêutica que cumpra os anteriores pontos 1.3.1.1. e 1.3.1.2. deve ser selecionada a com menores custos, preferivelmente uma terapêutica para a qual exista um preço de referência ou preço máximo;
 - 1.3.1.4. Quando o comparador é um fármaco, a dosagem e o intervalo de administração devem, obrigatoriamente, estar de acordo com o RCM.
 - 1.3.1.5. Ter já decisão de financiamento, comercializados e com consumos reportados;
2. A Comissão Executiva da CATS decide, no prazo de 1 semana, se aceita ou modifica esta proposta na sequência da interação com o Grupo Avaliador.
3. Após definição destes parâmetros de avaliação (medidas de eficácia terapêutica e segurança, sub-populações/sub-indicações a analisar e comparador ou comparadores a utilizar), o Grupo Avaliador recebe o *dossier* submetido pelo Titular de AIM para análise. Na análise do *dossier* recomenda-se que o Grupo de Avaliação siga a metodologia descrita neste documento.

CAPÍTULO II

II – Metodologia geral

1. Medidas de eficácia terapêutica e segurança

Deve existir uma definição prévia de quais as medidas de eficácia terapêutica e segurança que vão ser utilizadas na avaliação. As medidas de eficácia terapêutica e segurança utilizadas devem ser relevantes para o doente. Para este fim recomenda-se o uso das seguintes medidas:

- 1.1. Medidas de eficácia (ex.: mortalidade, morbidade, efeito clínico);
- 1.2. Expectativa de vida
- 1.3. Qualidade de vida relacionada com a saúde (HRQoL)
- 1.4. Duração média da doença
- 1.5. Medidas de segurança (ex.: Efeitos adversos)
- 1.6. Outras medidas (ex.: utilização de recursos de saúde como hospitalizações)

2. A relevância da certeza dos resultados

2.1. A principal tarefa do Grupo de Avaliação é dar informação ao Conselho Diretivo do INFARMED, I.P., com a maior confiança possível, sobre se existe evidência disponível que prove os benefícios ou danos de uma intervenção específica. Sempre que apropriado, a avaliação farmacoterapêutica deve ser feita em sub-populações/sub-indicações de doentes.

2.2. Para a avaliação do valor terapêutico acrescentado o Grupo de Avaliação utiliza a metodologia da Medicina-Baseada-na-Evidência. De salientar, que em Medicina o benefício de intervenções é avaliado em termos de probabilidade: o benefício é demonstrado quando a intervenção aumenta a probabilidade de um *outcome* benéfico ou reduz a probabilidade de um *outcome* não-benéfico.

2.3. A Medicina-Baseada-na-Evidência permite avaliar até que ponto a evidência disponível é confiável. Com este objetivo, utiliza um conjunto de regras e de instrumentos internacionalmente aceites e que constituem a base das avaliações de benefício efetuadas pelo Grupo de Avaliação. Das avaliações faz parte analisar um conjunto de

detalhes sobre a forma como os estudos foram planeados, conduzidos, analisados e publicados.

2.4.O Grupo de Avaliação baseia as suas avaliações do valor terapêutico adicional de intervenções utilizando apenas estudos com suficiente certeza de resultados. É **responsabilidade da empresa titular da Autorização de Introdução no Mercado (AIM) a submissão dos processos para avaliação pelo Infarmed incluindo toda a evidência que considerar relevante**. Se da análise dessa evidência resultar que os estudos incluídos no processo de submissão não apresentam a qualidade necessárias, o grupo de avaliação concluirá que, com a documentação submetida, não existe evidência disponível que prove os benefícios ou danos de uma intervenção específica.

3. A ligação entre a certeza dos resultados e a proximidade às condições do dia-a-dia

3.1.É frequentemente referido que os estudos com uma elevada certeza de resultados apresentam uma elevada validade interna mas muitas vezes não representam a prática corrente, que são prestados a uma população muito mais heterogénea. Ou seja, os resultados têm uma baixa validade externa e, por conseguinte, não são generalizáveis.

3.2.Contudo, esta crítica não resulta da metodologia utilizada (por exemplo, estudos aleatorizados) mas do facto de os critérios de elegibilidade serem em geral muito restritivos, excluindo frequentemente doentes idosos ou com múltiplas comorbilidades. Assim, aumentar a validade externa não implica reduzir o grau de certeza dos resultados mas antes fazer estudos adicionais que incluam os grupos de doentes considerados relevantes.

3.3.Assim, elevada certeza de resultados e proximidade às condições do dia-a-dia não se excluem mutuamente. A resposta às questões postas pela prestação de cuidados no dia-a-dia requer apenas que se implementem os estudos necessários para responder de forma apropriada a essas mesmas questões, sem abdicar de uma elevada certeza de resultados.

4. Avaliação de benefício da intervenção em análise

4.1.Benefício e dano relevantes para o doente: definições

4.1.1.O termo **“benefício”** refere-se a efeitos causais positivos, e o termo **“dano”** refere-se a efeitos causais negativos resultantes de uma intervenção médica, em medidas de eficácia terapêutica e segurança relevantes para os doentes. O termo

“causal” é aqui utilizado significando que existe certeza suficiente de que os efeitos observados podem ser atribuídos exclusivamente à intervenção que está a ser testada.

4.1.2.No caso de comparação entre duas intervenções, os seguintes termos serão usados na avaliação comparativa de benefício ou dano:

4.1.2.1. **Benefício:** no caso de maior benefício, o termo “maior benefício” é usado. No caso de menor benefício ou benefício comparável, são usados os termos “menor benefício” ou “benefício comparável”.

4.1.2.2. **Dano:** São usados os termos “maior dano”, “dano comparável” e “menor dano”. A descrição do dano é um componente essencial na avaliação de benefício de uma intervenção. Os estudos são geralmente desenhados para medir a efetividade da intervenção, enquanto os efeitos adversos são registados concomitantemente como eventos adversos. Assim, os resultados relativos a eventos adversos são frequentemente de fraca qualidade, estando altamente dependentes dos métodos utilizados para a colheita de dados, podendo estes reduzir a certeza dos resultados. **A descrição detalhada dos métodos utilizados para colheita e análise dos efeitos adversos é um componente crítico do processo de avaliação de benefício de uma intervenção.**

4.1.2.3. **Da conjugação da avaliação dos benefícios e dos danos da intervenção, conclui-se sobre “Valor Terapêutico Acrescentado”, “Não inferioridade”, “Comparabilidade”, “Equivalência Terapêutica” ou “Inferioridade Terapêutica”.**

4.1.3.O Grupo de Avaliação avalia o efeito de intervenções sobre medidas de eficácia terapêutica e segurança relevantes. Tanto os efeitos intencionais como os não-intencionais de uma intervenção devem ser considerados. Assim, recomenda-se que seja considerado o impacto, determinado pela doença ou pelo seu tratamento, nas seguintes medidas de eficácia terapêutica e segurança:

4.1.3.1. Mortalidade

4.1.3.2. Morbilidade (sintomas e complicações)

4.1.3.3. Duração da doença

4.1.3.4. Qualidade de Vida Relacionada com a Saúde (HRQoL).

4.1.4.Na avaliação da HRQoL devem ser considerados instrumentos validados.

- 4.1.5. Na avaliação de benefício podem ser considerados como substitutos de medidas de eficácia terapêutica relevantes-para-o-doente medidas de eficácia terapêutica sub-rogada desde que estas tenham sido previamente validados.
- 4.1.6. Para que uma medida de eficácia terapêutica sub-rogada seja validada é necessário responder afirmativamente às 3 seguintes questões:
- 4.1.6.1. Existe uma associação forte, consistente e independente entre o *outcome* sub-rogado e o *outcome* relevante? Este critério é necessário mas não é, por si só, suficiente.
 - 4.1.6.2. Existe evidência de estudos aleatorizados utilizando a mesma classe de medicamentos de que melhorias no *outcome* sub-rogado conduziram de forma consistente a melhorias no *outcome* relevante?
 - 4.1.6.3. Existe evidência de estudos aleatorizados utilizando outras classes de medicamentos de que melhorias no *outcome* sub-rogado conduziram de forma consistente a melhorias no *outcome* relevante-para-o-doente? Para ser validado o *outcome* sub-rogado deve cumprir os critérios 1 a 3.
 - 4.1.6.4. A verificação dos critérios anteriores requer geralmente a realização de meta-análises de estudos aleatorizados.
 - 4.1.6.5. A utilização de *outcomes* sub-rogados não validados não é apropriada para emitir conclusões sobre a intervenção.
- 4.1.7. É importante focar a avaliação de eventos adversos nos efeitos adversos relevantes. São considerados “relevantes” os efeitos adversos que:
- 4.1.7.1. Contrabalançam, completa ou quase completamente, o benefício de uma intervenção.
 - 4.1.7.2. Incluem morbidade grave, aumento da mortalidade, ou diminuição marcada da HRQoL.
 - 4.1.7.3. Ocorrem predominantemente com opções terapêuticas que são particularmente efetivas.
 - 4.1.7.4. Variam substancialmente entre duas opções de tratamento que seriam equivalentes.
 - 4.1.7.5. Apresentam uma relação dose-efeito.
 - 4.1.7.6. Considerados pelos doentes como especialmente importantes.

4.2. Avaliação da qualidade da evidência por cada outcome (por cada medida de eficácia ou de segurança)

4.2.1. O Grupo de Avaliação avalia a qualidade da evidência disponível utilizando a metodologia do GRADE - “Grading of Recommendations, Assessment, Development and Evaluation”. Para o GRADE, e no contexto de uma revisão sistemática, a “qualidade” reflete a nossa confiança de que as estimativas de efeito estão corretas. Na avaliação de um *dossier*, o Grupo de Avaliação utiliza esta mesma definição mas no contexto da evidência (conjunto de estudos) incluída no dossier submetido pelo Titular.

4.2.2. A qualidade da evidência deve ser avaliada para cada *outcome* (e não para cada estudo), mas a classificação final deve referir-se ao conjunto de estudos avaliados.

4.2.3. O GRADE classifica a qualidade da evidência em 4 níveis: alta, moderada, baixa e muito baixa. Esta classificação aplica-se não a estudos individuais mas a um corpo de evidência (um conjunto de estudos). Na avaliação inicial os estudos aleatorizados começam como evidência de alta qualidade, mas esta classificação inicial pode ser reduzida por 5 fatores (risco de viés, imprecisão, inconsistência, evidência indireta e viés de publicação). Detalhes sobre a metodologia a utilizar para avaliação da qualidade da evidência serão descritos nos pontos 4.4. a 4.7. A classificação da qualidade da evidência permite hierarquizar a certeza de resultados:

4.2.3.1. Qualidade alta significa elevada certeza de resultados (**Significado:** Estamos muito confiantes de que o verdadeiro efeito está muito próximo das estimativas de efeito)

4.2.3.2. Qualidade moderada significa moderada certeza de resultados (**Significado:** Estamos moderadamente confiantes na estimativa de efeito: o verdadeiro efeito é provável que esteja próximo da estimativa de efeito, mas existe a possibilidade que possa ser substancialmente diferente).

4.2.3.3. Qualidade baixa significa baixa certeza de resultados (**Significado:** A nossa confiança nas estimativas de efeito é limitada. O verdadeiro efeito pode ser substancialmente diferente da estimativa de efeito).

4.2.3.4. Qualidade muito baixa significa muito baixa certeza de resultados (**Significado:** A nossa confiança nas estimativas de efeito é muito

limitada. O verdadeiro efeito pode ser muito diferente da estimativa de efeito).

- 4.2.3.5. **As conclusões sobre a existência de “Valor terapêutico acrescentado” ou de “Equivalência terapêutica” ou “Inferioridade terapêutica” só podem ser emitidas se estiverem disponíveis estudos apropriados, ou seja, se a evidência for de alta qualidade.**

4.3. **Classificação global (para todos os *outcomes*) da confiança nas estimativas de efeito (“qualidade”) de uma intervenção**

4.3.1. As conclusões anteriores obtidas separadamente para cada *outcome*, são depois sumarizadas numa avaliação final única sobre os benefícios e os danos da intervenção.

4.3.2. Nesta avaliação final, em geral, a classificação atribuída à qualidade da evidência global é a mesma que foi atribuída ao *outcome* “crítico” que fornece a confiança mais baixa. Para ver definição de *outcome* crítico (consultar bibliografia).

4.4. **Classificação da qualidade da evidência: risco de viés**

4.4.1. A qualidade dos estudos aleatorizados é inicialmente classificada como alta, mas pode ser reduzida pelos 5 fatores referidos no anterior ponto 9.2.3. De salientar que a classificação da qualidade não se refere a estudos individuais **mas a cada *outcome* (avaliação para cada *outcome*)**, sendo esta a base para uma posterior avaliação global (conjunto de estudos avaliados) da qualidade.

4.4.2. O primeiro fator (risco de viés) resulta de problemas metodológicos no desenho ou condução do estudo e inclui um conjunto de 5 problemas metodológicos:

4.4.2.1. **Ausência de alocação oculta:** os investigadores têm conhecimento prévio do grupo a que vai ser alocado o próximo doente incluído.

4.4.2.2. **Ausência de ocultação:** os doentes, os investigadores, os que registam os *outcomes*, os que adjudicam os *outcomes*, os que analisam os dados têm conhecimento do braço a que os doentes estão alocados (ou da medicação que estão atualmente a receber no caso de um ensaio *crossover*).

4.4.2.3. **Inclusão incompleta de doentes:** perda de doentes para *follow up* e não aderência ao princípio intenção-de-tratar nos estudos de superioridade.

Historicamente, os metodologistas têm sugerido limiares arbitrários para uma perda de *follow up* aceitável (por exemplo, menos de 20%). Contudo, o significado de uma perda para *follow up* depende da relação entre perda para *follow up* e número de eventos. Como regra geral, **quanto maior for a diferença entre a percentagem de perda para *follow up* e a percentagem de eventos nos grupos de intervenção e controlo, maior o risco de viés**. Por exemplo, se os eventos forem 2% e 4% nos grupos de intervenção e controlo, uma perda para *follow up* de 5% é preocupante.

- 4.4.2.4. **Reporte seletivo de *outcomes***: reporte incompleto ou ausente de alguns *outcomes* influenciado pelos resultados. Os membros do Grupo de Avaliação **devem analisar os protocolos registados em bases de dados de ensaios clínicos** (por exemplo, <https://clinicaltrials.gov>) e avaliar se todos os *outcomes* registados foram analisados.
- 4.4.2.5. **Outras limitações**: Interrupção precoce do estudo por benefício, uso de medidas de *outcome* não validadas, etc.). Evidência recente sugere que **os estudos interrompidos precocemente por benefício sobrestimam os efeitos do tratamento**. Este facto deve ser tido em conta pelo Grupo de Avaliação.
- 4.4.2.6. Na classificação inicial da qualidade da evidência, o Grupo de Avaliação deve ter em conta que se trata de uma avaliação específica para cada *outcome*, pelo que o impacto destes problemas metodológicos pode variar substancialmente. O risco de viés por ausência de ocultação ou por ausência de alocação oculta é maior em estudos com *outcomes* subjetivos. Por exemplo, a ausência de ocultação não é um problema grave se o *outcome* a avaliar for a mortalidade global, pelo que neste caso a classificação da qualidade não deverá ser reduzida.
- 4.4.2.7. O Grupo de Avaliação deve reduzir a qualidade da evidência em um nível (para moderada) se considerar que o conjunto da evidência apresenta limitações (risco de viés) graves, e deve reduzir a qualidade da evidência em dois níveis (para baixa) se considerar que o conjunto da evidência apresenta limitações (risco de viés) muito graves.

- 4.4.2.8. Contudo, se as limitações não forem ao nível do *outcome* mas ao nível de um estudo individual, o Grupo de Avaliação deve excluir esse estudo do processo de avaliação, incluindo apenas os estudos de alta qualidade.
- 4.4.2.9. Por questões de transparência, cada elemento do Grupo de Avaliação deve fazer individualmente esta avaliação do risco de viés para cada *outcome* preenchendo o Formulário FORM 005-QE-RV. Caso existam discrepâncias entre as diferentes avaliações, estas devem ser ultrapassadas por consenso entre os elementos do Grupo de Avaliação. No caso de não ser possível o consenso, prevalecerá a avaliação dominante. O processo deverá incluir os registos da avaliação individual inicial, assim como o registo final (Formulário FORM 006-QE-RV). No caso de não haver consenso, devem ficar registadas as divergências e respetivos motivos.

4.5. Classificação da qualidade da evidência: Imprecisão

- 4.5.1. Os critérios utilizados pelo GRADE para referir “imprecisão” aplicam-se no contexto de uma meta-análise.
- 4.5.2. O principal critério para avaliar a precisão é o intervalo de confiança a 95% à volta da diferença de efeito entre a intervenção e o controlo para cada *outcome*. Em geral, deve ser considerado o intervalo de confiança à volta do efeito absoluto, e não do efeito relativo. Conceptualmente, o intervalo de confiança a 95% pode ser interpretado como o intervalo dentro do qual, em 95% dos casos, o verdadeiro valor se encontra. Quando avaliando a qualidade da evidência, a questão é saber **se o intervalo de confiança à volta da estimativa do efeito da intervenção é suficientemente estreito**.
- 4.5.3. Assim, para avaliar se os resultados são suficientemente precisos o Grupo de Avaliação deverá utilizar cumulativamente os 2 seguintes critérios:
- 4.5.3.1. O intervalo de confiança a 95% é suficientemente estreito e exclui o efeito nulo.
- 4.5.3.2. O número de participantes incluídos nos estudos em análise é igual ou superior ao “tamanho ótimo de informação (TOI)” (o TOI é obtido calculando o número de doentes necessários a incluir num estudo com poder estatístico suficiente. Trata-se no fundo de fazer uma estimativa

do tamanho da amostra de um estudo com suficiente poder estatístico. Existem múltiplas calculadoras online para calcular o tamanho da amostra disponível. Uma calculadora simples de utilizar encontra-se em <http://www.stat.ubc.ca/~rollin/stats/ssize/b2.html> e requer apenas que se introduza a percentagem de eventos nos grupos comparador e de intervenção, podendo deixar-se por defeito um α de 0,05, e um β de 0,2.

- 4.5.3.3. Se os critérios definidos nos 2 parágrafos anteriores não forem, cumulativamente, cumpridos, o Grupo de Avaliação deverá reduzir a classificação da qualidade da evidência por imprecisão.
- 4.5.3.4. Se o Grupo de Avaliação não tiver disponível uma meta-análise, ou entender que os dados disponíveis não são adequados para este tipo de análise, deverá colocar na classificação da qualidade, no item “imprecisão”, a seguinte avaliação: Não determinado.
- 4.5.3.5. Por questões de transparência, cada elemento do Grupo de Avaliação deve fazer individualmente esta avaliação da precisão para cada *outcome* preenchendo o Formulário FORM 007-QE-IMP. Caso existam discrepâncias entre as diferentes avaliações, estas devem ser ultrapassadas por consenso entre os elementos do Grupo de Avaliação. No caso de não ser possível o consenso, prevalecerá a avaliação dominante. O processo deverá incluir os registos da avaliação individual inicial, assim como o registo final (Formulário FORM 008-QE-IMP). No caso de não haver consenso, devem ficar registadas as divergências e respetivos motivos.

4.6. Classificação da qualidade da evidência: Inconsistência

- 4.6.1. Os critérios utilizados pelo GRADE para referir “inconsistência” referem-se a medidas relativas (*risk ratio, hazard ratio ou odds ratio*), não a medidas absolutas, e aplicam-se no contexto de uma meta-análise.
- 4.6.2. A classificação da qualidade da evidência pode ser reduzida por inconsistência, mas não é aumentada por consistência.
- 4.6.3. O Grupo de Avaliação deve utilizar um conjunto de 4 critérios para avaliar a inconsistência nos resultados no contexto de uma meta-análise:
 - 4.6.3.1. As estimativas de efeito variam substancialmente entre estudos.

- 4.6.3.2. Os intervalos de confiança não mostram qualquer sobreposição ou mostram apenas uma sobreposição mínima.
- 4.6.3.3. Os testes estatísticos de heterogeneidade (geralmente o teste Q) – que testam a hipótese nula de que todos os estudos incluídos numa meta-análise apresentam a mesma magnitude de efeito – mostram um valor de P significativo.
- 4.6.3.4. O I^2 apresenta um valor elevado. O valor do I^2 pode variar entre 0 e 100%. I^2 informa-nos de qual é a proporção (%) da variância dos efeitos observados que reflete a variância dos verdadeiros efeitos (e, por conseguinte, que não resulta apenas de erro na amostragem [sampling error] situação em que o I^2 seria 0%).
- 4.6.3.5. Se utilizando os critérios anteriores o Grupo de Avaliação chegar à conclusão de que os resultados apresentam problemas de consistência deverá reduzir a classificação da qualidade da evidência por inconsistência.
- 4.6.3.6. Se o Grupo de Avaliação não tiver disponível uma meta-análise, ou entender que os dados disponíveis não são adequados para este tipo de análise, deverá colocar na classificação da qualidade, no item inconsistência, a seguinte avaliação: Não determinado.
- 4.6.3.7. Por questões de transparência, cada elemento do Grupo de Avaliação deve fazer individualmente esta avaliação da consistência/inconsistência de resultados para cada *outcome* preenchendo o Formulário FORM 009-QE-INC. Caso existam discrepâncias entre as diferentes avaliações, estas devem ser ultrapassadas por consenso entre os elementos do Grupo de Avaliação. No caso de não ser possível o consenso, prevalecerá a avaliação dominante. O processo deverá incluir os registos da avaliação individual inicial, assim como o registo final (Formulário FORM 010-QE-INC). No caso de não haver consenso, devem ficar registadas as divergências e respetivos motivos.

4.7. Classificação da qualidade da evidência: Evidência indireta

A evidência pode ser indireta de 4 modos:

- 4.7.1.A população de estudo é diferente da população de interesse.

- 4.7.2.A intervenção testada é diferente da intervenção de interesse.
- 4.7.3.Os *outcomes* são diferentes dos *outcomes* de interesse, por exemplo, o uso de *outcomes* sub-rogados.
- 4.7.4.A comparação do efeito da intervenção de interesse e do comparador não é feita diretamente (*head-to-head*) mas indiretamente. Neste caso a classificação da qualidade da evidência deve ser reduzida um ou dois níveis dependendo da amplitude das diferenças entre as populações de doentes, co-intervenções, e medidas de *outcome*.
- 4.7.5.O Grupo de Avaliação deve avaliar o impacto possível da evidência indireta (diferentes populações ou intervenções e uso de sub-rogados) nos resultados e decidir se deve reduzir a classificação da qualidade da evidência.
- 4.7.6.Por questões de transparência, cada elemento do Grupo de Avaliação deve fazer individualmente esta avaliação da importância da evidência indireta nos resultados para cada *outcome* preenchendo o Formulário FORM 011-QE-IND. Caso existam discrepâncias entre as diferentes avaliações, estas devem ser ultrapassadas por consenso entre os elementos do Grupo de Avaliação. No caso de não ser possível o consenso, prevalecerá a avaliação dominante. O processo deverá incluir os registos da avaliação individual inicial, assim como o registo final (Formulário FORM 012-QE-IND). No caso de não haver consenso, devem ficar registadas as divergências e respetivos motivos.

CAPÍTULO III

III – Valor terapêutico acrescentado

1. Determinação da existência de valor terapêutico acrescentado

Com o objetivo de determinar o valor terapêutico acrescentado, e com base na análise científica dos dados disponíveis, **deve concluir-se separadamente para cada *outcome* relevante-para-o-doente do valor terapêutico acrescentado ou do dano classificado em 4 níveis em função do grau de certeza dos resultados: “prova” (elevada certeza de resultados quando a qualidade da evidência é alta), “indicativo” (moderada certeza de resultados quando a qualidade da evidência é moderada), “sugestivo” (baixa certeza de resultados quando a qualidade da evidência é baixa), ou nenhum dos anteriores quando não existem dados disponíveis ou a qualidade da evidência é muito baixa.** O resultado da avaliação da existência de valor terapêutico acrescentado deve ser expressa numa das seguintes formas:

- 1.1. Existe prova, indicação ou sugestão de valor terapêutico acrescentado de uma intervenção;
- 1.2. Existe prova, indicação ou sugestão de equivalência terapêutica de uma intervenção;
- 1.3. Existe prova, indicação ou sugestão de inferioridade terapêutica de uma intervenção.

2. Classificação da extensão do valor terapêutico acrescentado

2.1. A extensão do valor terapêutico acrescentado deve ser classificada numa das seguintes formas:

- 2.1.1. Valor terapêutico acrescentado maior.
- 2.1.2. Valor terapêutico acrescentado moderado.
- 2.1.3. Valor terapêutico acrescentado menor.
- 2.1.4. Valor terapêutico acrescentado não quantificável.

2.2. Outcomes binários

A determinação da extensão do valor terapêutico acrescentado deve ter em conta a qualidade do *outcome*, e ser baseada no risco relativo tendo em consideração as Tabelas em anexo. Se não existir este dado, deve ser calculado o risco relativo e respetivo intervalo de confiança a 95% utilizando uma das múltiplas calculadoras

disponíveis na internet (por exemplo em https://www.medcalc.org/calc/relative_risk.php).

2.3. Tempo até ao evento

O intervalo de confiança a 95% do *hazard ratio* é necessário para determinar a extensão do efeito no caso de *outcomes* avaliados por “tempo até ao evento”. Se existe uma meta-análise de vários estudos deverá ser usado o *hazard ratio*. Para determinar a extensão do valor terapêutico acrescentado utilizam-se os mesmos limiares da tabela abaixo.

Se não existe um *hazard ratio* ou este não for calculável, deve considerar-se a possibilidade de se calcular um risco relativo. Se apropriado, deve calcular-se o risco relativo numa determinada data.

2.4. Extensão do Valor Terapêutico Acrescentado

Nem sempre é possível quantificar a extensão do valor terapêutico acrescentado ao nível do *outcome*. Por exemplo, se se observar um efeito estatisticamente significativo num sub-rogado considerado como validado, mas não existirem dados de confiança sobre o efeito sobre um *outcome* relevante para o doente, não será possível quantificar a extensão do efeito. Neste caso, o valor terapêutico acrescentado será considerado como “não quantificável”.

| | | Categoria de <i>outcome</i> | | | |
|------------------------|--|--|---|---|---|
| | | Mortalidade global | Sintomas (ou complicações tardias) e eventos adversos graves (ou severos) | Qualidade de vida-relacionada com a saúde | Sintomas (ou complicações tardias) e eventos adversos não graves (ou não severos) |
| Categorias de extensão | Maior Grande melhoria de forma sustentada na medida de avaliação de benefício, que não foi anteriormente atingida em relação ao comparador apropriado | Aumento importante (maior) no tempo de sobrevida | Supressão ou evicção prolongada | Melhoria maior | Não aplicável |
| | Moderado Melhoria marcada na medida de avaliação de benefício, que não foi anteriormente atingida em relação ao comparador apropriado | Aumento moderado no tempo de sobrevida | Supressão ou evicção relevante | Melhoria importante | Evicção importante |
| | Menor Melhoria moderada e não apenas marginal na medida de avaliação de benefício, que não foi anteriormente atingida em relação ao comparador apropriado | Qualquer aumento no tempo sobrevida | Qualquer redução | Melhoria relevante | Evicção relevante |

| | | Categoria de <i>outcome</i> | | |
|------------------------|--|-----------------------------|---|---|
| | | Mortalidade global | Sintomas (ou complicações tardias) e eventos adversos graves (ou severos) | Sintomas (ou complicações tardias) e eventos adversos não graves (ou não severos) |
| Categorias de extensão | Maior Grande melhoria de forma sustentada na medida de avaliação de benefício, que não foi anteriormente atingida em relação ao comparador apropriado | 0,85 | 0,75 e risco $\geq 5\%$ | Não aplicável |
| | Moderado Melhoria marcada na medida de avaliação de benefício, que não foi anteriormente atingida em relação ao comparador apropriado | 0,95 | 0,90 | 0,80 |
| | Menor Melhoria moderada e não apenas marginal na medida de avaliação de benefício, que não foi anteriormente atingida em relação ao comparador apropriado | 1,00 | 1,00 | 0,90 |

Os valores na Tabela referem-se a um limiar superior abaixo do qual deve estar o intervalo de confiança a 95% do risco relativo. O intervalo de confiança a 95% estimado do risco relativo deve ser inferior (mais afastado de 1) ao limiar definido, ou seja, o limite superior do intervalo de confiança deve ser inferior ao limiar referido. Por exemplo, em relação à mortalidade global, para que o valor terapêutico acrescentado seja considerado disruptivo, o limite superior do intervalo de confiança a 95% do risco relativo deve ser pelo menos de 0,84, ou seja, a redução do risco relativo de morte em relação ao comparador apropriado deve estar incluído num intervalo de confiança a 95% cujo valor superior é igual ou inferior a 16% (0,84)

CAPÍTULO IV

IV - Estudos de superioridade, não inferioridade, e equivalência: definições e critérios para mudança de objetivos

1. Demonstração de equivalência

1.1. Um dos erros graves mais frequentes na interpretação de dados médicos é classificar um resultado não significativo de um teste de significância como evidência de que a hipótese nula é verdade.

1.2. Para demonstrar “equivalência” é necessário utilizar métodos que permitam testar a hipótese de equivalência, ou seja, o estudo deve **ter um desenho de equivalência** que permita confirmar a ausência de uma diferença significativa entre tratamentos (por exemplo, que os valores médios da diferença entre 2 grupos é exatamente zero).

1.3. **Este objetivo obtém-se através do cálculo e observação dos intervalos de confiança, uma vez que não é possível o uso de testes estatísticos.** Na altura da elaboração do protocolo é necessário definir uma margem de equivalência clínica (Δ), através da definição da maior diferença que é clinicamente aceitável, de tal modo que uma diferença maior seria relevante na prática clínica. Os dois tratamentos são considerados equivalentes se o intervalo de confiança a 95% (*two-sided*), que define o intervalo das diferenças plausíveis entre os dois tratamentos, está dentro do intervalo $-\Delta$ a $+\Delta$. Na prática, o que se demonstra não é a existência de uma equivalência exata (que a diferença entre os valores médios dos 2 grupos é exatamente zero), mas que a diferença entre os dois grupos é irrelevante. Como nos estudos de superioridade, é necessário estimar o tamanho da amostra nos estudos de equivalência na altura da elaboração do protocolo.

1.4. No caso de estudos de bioequivalência tem sido aceite como o padrão a utilização de intervalos de confiança a 90% da diferença entre tratamentos, na avaliação dos valores médios de parâmetros farmacocinéticos. No caso de um genérico inalado ou de um produto aplicado topicamente, em que estudos de bioequivalência são impossíveis, aceita-se a realização de estudos de bioequivalência clínica, utilizando intervalos de confiança a 95%.

1.5. Quando o intervalo de confiança a 95% que define o intervalo das diferenças plausíveis entre os dois tratamentos é *one-sided*, o estudo é referido como de não inferioridade.

2. Demonstração de não inferioridade

2.1. Nos estudos de não inferioridade pretende-se demonstrar que o novo tratamento não é menos eficaz (significando que pode ter a mesma eficácia ou ter mais eficácia) do que um tratamento já existente. Como referido, também aqui se aplica a abordagem dos intervalos de confiança, mas agora apenas estamos interessados em avaliar a possível diferença numa única direção. Assim, o intervalo de confiança a 95% (*two-sided*) da diferença entre tratamentos deve estar completamente para a direita do valor $-\Delta$.

2.2. Como já referido, **os estudos de não inferioridade são por vezes designados, erroneamente, como estudos de equivalência**, representando uma fonte de confusão.

2.3. Intervalos de confiança one-sided e two-sided

De acordo com o ICH E9 Note for Guidance, deverão sempre ser usados intervalos de confiança a 95% *two-sided* em todos os estudos clínicos independentemente do seu objetivo. No caso de serem utilizados intervalos de confiança *one-sided* devem se usados para cobrir uma probabilidade de 97,5%.

No caso especial de estudos de bioequivalência, tem sido recomendada a utilização de intervalos de confiança a 90% *two-sided*.

3. Estudos de superioridade

3.1. Os estudos de superioridade são desenhados para detetar uma diferença entre tratamentos, e a demonstração de superioridade faz-se através da utilização de um teste de significância estatística. O teste de significância estatística testa o efeito nulo, ou seja, coloca a hipótese de que não existem diferenças nos efeitos clínicos entre 2 tratamentos. O grau de significado estatístico (valor de p) indica a probabilidade de que a diferença observada tenha ocorrido por acaso na hipótese de que, na realidade, não existe diferença entre os tratamentos.

3.2. Uma vez que se considere que a hipótese de não-diferença entre tratamentos é insustentável por pouco provável (um $p < 0,05$ indica que essa probabilidade é inferior a 5%), é importante estimar a magnitude da diferença para avaliar se essa magnitude é clinicamente relevante. Com esse objetivo, é necessário calcular a melhor estimativa da magnitude da diferença entre tratamentos, habitualmente designada por estimativa pontual (*point estimate*), que, em dados com distribuição normal, corresponde à diferença nos valores médios da medida de eficácia utilizada. Deve-se também calcular o intervalo de confiança, que corresponde ao intervalo de valores plausíveis da verdadeira diferença. Este intervalo não deve incluir o efeito nulo, que é o **valor zero** no caso do *effect size* ser uma variável contínua (por exemplo, *standardized mean difference*), ou que é o **valor um** no caso dos rácios (*risk ratio*, *odds ratio* ou *hazard ratio*), uma vez que a hipótese nula (diferença zero entre tratamentos) já foi rejeitada. Assim, as seguintes afirmações são consideradas equivalentes: o intervalo de confiança a 95% da diferença entre tratamentos exclui o valor nulo e os dois valores são estatisticamente diferentes no nível de significância de 5% *two-sided* ($p < 0,05$).

3.3. Nos estudos de superioridade, avaliar se uma diferença entre tratamentos com significado estatístico é clinicamente relevante requer um juízo de valor *a posteriori*. Pelo contrário, nos estudos de equivalência e de não inferioridade a relevância clínica é definida previamente, na altura da elaboração do protocolo de estudo, através da definição do Δ (margem de não inferioridade $[-\Delta]$ ou margem de equivalência $[\pm\Delta]$).

4. Relevância da pré-definição do Δ nos estudos de não inferioridade e de equivalência

4.1. A demonstração de “equivalência” ou de “não inferioridade” depende do valor Δ selecionado que deve representar a diferença máxima aceitável para o objetivo de interesse. É importante salientar que, após inspeção dos dados, é sempre possível selecionar um valor de Δ que conduz à conclusão de “equivalência” ou de “não inferioridade”.

4.2. Uma vez que a escolha do Δ requer sempre julgamento clínico, existe nesta escolha sempre um risco de viés, mas esse risco aumenta exponencialmente se o Δ for selecionado após inspeção dos dados. Assim, a seleção do Δ deve ser **sempre feita na**

altura da elaboração do protocolo de estudo, e justificada baseada em argumentos plausíveis.

5. Relevância da pré-definição do estudo como de superioridade, não inferioridade ou equivalência

5.1. De acordo com a *Guideline Points to consider on switching between superiority and non-inferiority* do *Committee for Proprietary Medicinal Products (CPMP)*¹ da European Medicines Agency, é fundamental a **pré-definição** do estudo como de superioridade, de não inferioridade, ou de equivalência pelos seguintes motivos:

- Assegura que os comparadores, doses dos medicamentos, populações a incluir, e *outcomes* são apropriados;
- Permite que a estimativa do tamanho da amostra seja baseada em cálculos apropriados do poder estatístico;
- Assegura que os critérios de equivalência ou não inferioridade são pré-definidos;
- Permite que o protocolo descreva em detalhe a análise estatística apropriada;
- Assegura que o estudo tem sensibilidade suficiente para atingir os seus objetivos.

6. É possível mudar o objetivo de uma comparação?

6.1. A mudança entre superioridade e não inferioridade é a única mudança com relevância prática. **Os estudos de equivalência são tão específicos que não existe a possibilidade de mudança**, seja entre equivalência e superioridade, seja entre equivalência e não inferioridade.

Interpretando um estudo de não inferioridade como um estudo de superioridade

6.2. Nos casos em que **todo** o intervalo de confiança a 95% da diferença entre tratamentos está não apenas para a direita de $-\Delta$ mas também para a direita (acima de) do valor nulo (zero no caso das variáveis contínuas, e um no caso dos rácios), existe evidência de superioridade em termos de significado estatístico ao nível de 5% ($p < 0,05$). Neste caso é aceitável calcular o valor de p associado a um teste de superioridade (teste

estatístico de significância) e avaliar se tem significado estatístico. A interpretação deste teste não é afetada pelo problema de multiplicidade (erro de tipo I), uma vez que corresponde a um único teste estatístico de significância.

6.3. Assim, de acordo com o CPMP¹ é possível mudar o objetivo de um estudo de não inferioridade para um estudo de superioridade desde que:

6.3.1. O estudo tenha sido desenhado e conduzido de acordo com os requisitos de um estudo de não inferioridade;

6.3.2. Sejam apresentados os valores de p para a superioridade;

6.3.3. O estudo tenha sido analisado de acordo com o princípio *intention-to-treat*.

Interpretando um estudo de superioridade como um estudo de não inferioridade

6.4. Se um estudo de superioridade não mostra uma diferença estatisticamente significativa entre tratamentos, poderia haver interesse num objetivo menor de estabelecer não inferioridade. Se a diferença entre tratamentos num estudo de superioridade é apresentada sob a forma de intervalo de confiança, o limite inferior do intervalo de confiança fornece uma estimativa do efeito mínimo do novo tratamento em relação ao comparador. **Se o protocolo de estudo define uma margem de não inferioridade considerada como aceitável**, mudar o objetivo do estudo de superioridade para não inferioridade é possível e aceitável.

6.5. Nos estudos de superioridade em que a margem de não inferioridade não foi definida na altura da elaboração do protocolo, não é aceitável mudar o objetivo do estudo, uma vez que, após inspeção dos dados, é sempre possível selecionar um valor de Δ que conduz à conclusão de “equivalência” ou de “não inferioridade”. Assim, uma definição *post hoc* do Δ está associado a um risco de viés elevado pelo que não é aceitável.

6.6. Assim, **um estudo de superioridade em que o teste de significância para a diferença entre tratamentos não mostrou significado estatístico**, caso não tenha definido a margem de não inferioridade na altura de elaboração do protocolo, **deve ser considerado apenas como um estudo de superioridade negativo**, não sendo aceitável mudar o objetivo de superioridade para não inferioridade.

CAPÍTULO V

V – Comparações diretas e indiretas

1. Definições

- 1.1. Entende-se por **comparação direta** a comparação *head-to-head* de dois tratamentos ativos, ou a combinação de múltiplos estudos *head-to-head* para gerar uma estimativa combinada (meta-análise) da eficácia relativa de dois tratamentos.
- 1.2. **Comparação indireta** é a estimativa da eficácia relativa de dois ou mais tratamentos na ausência de estudos *head-to-head*.
- 1.3. Por **comparação de tratamentos múltiplos** entende-se a estimativa da eficácia relativa de três ou mais tratamentos.
- 1.4. **Comparação mista de tratamentos** define-se como a estimativa da eficácia relativa de 3 ou mais tratamentos usando simultaneamente comparações diretas e indiretas.

2. Comparações diretas

- 2.1. As comparações diretas utilizam os resultados de estudos comparando *head-to-head* dois tratamentos ativos para estimar a eficácia relativa dos tratamentos em comparação. É a comparação que fornece a melhor evidência da eficácia relativa dos dois tratamentos pelo deve ser o método utilizado preferencialmente.
- 2.2. Quando a evidência disponível inclui vários estudos *head-to-head*, por vezes combina-se os resultados desses estudos utilizando técnicas meta-analíticas, para gerar uma estimativa combinada (*pooled estimate*) da eficácia relativa dos 2 tratamentos.

3. Comparações indiretas

- 3.1. Contudo, por vezes existem dados insuficientes para estimar de forma confiável a eficácia relativa de 2 tratamentos ou pode haver necessidade de comparar mais de 2 tratamentos simultaneamente, situação em que é necessário utilizar métodos de comparação de múltiplos tratamentos.
- 3.2. Assim, os métodos de comparação de múltiplos tratamentos podem ser usados para inferir a eficácia relativa de 2 tratamentos na ausência de estudos *head-to-head* (por comparação indireta) ou através da combinação de comparações diretas e indiretas.
- 3.3. **É importante salientar que, nas comparações indiretas, apenas deverão ser utilizados métodos ajustados (meta-analíticos) que mantenham a randomização.** As comparações indiretas não ajustadas, quando comparadas com as comparações diretas, resultam num elevado número de discrepâncias em relação ao significado e direção da eficácia relativa⁴. Comparações indiretas não ajustadas têm o mesmo valor das comparações utilizando estudos observacionais e não são recomendadas.
- 3.4. Só devem ser realizadas comparações indiretas se os estudos disponíveis forem comparáveis, homogêneos e consistentes, de modo que os resultados obtidos possam ser confiáveis.
- 3.5. Existem múltiplos métodos de comparações múltiplas, genericamente denominadas Network Meta-Analysis, e incluem (entre outros) o método de Bucher de comparações indiretas ajustadas, o método de Lumley de network meta-analysis, comparações mistas de tratamento (MTC, mixed treatment comparison) Bayesianas. A escolha do método pode ser individualizada para cada caso. Descrição detalhada destas metodologias está para além do âmbito deste documento.
- 3.6. Durante a avaliação da evidência submetida pela empresa, e em caso de necessidade, os Grupos de Avaliação da Evidência (GAE) deverão solicitar à empresa a apresentação das comparações consideradas apropriadas à avaliação de interesse.
- 3.7. O GAE deverá ter em conta que os métodos de Network Meta-Analysis são particularmente relevantes nos casos em que na prática clínica corrente se

utiliza vários (mais do que um) tratamentos para a mesma indicação, e em que, por conseguinte, podem existir vários comparadores selecionados para essa avaliação.

CAPÍTULO VI

VI – Análises de sub-grupos

1. O GAE durante a definição dos critérios de avaliação pode considerar a possibilidade de analisar a eficácia e/ou segurança em sub-populações de doentes.
2. Recomenda-se que a avaliação da eficácia e/ou segurança de tratamentos em sub-populações seja reservada para os casos em que sejam considerados comparadores distintos para as diferentes sub-populações.

CAPÍTULO VII

VII – Elaboração e Emissão do Relatório de Avaliação Farmacoterapêutica pelo Grupo de Avaliação da Evidência

1. Notas sobre a elaboração do relatório farmacoterapêutico

- 1.1. Por razões de harmonização e consistência, e sem prejuízo de situações de preenchimento casuístico, são identificados para cada secção do relatório o conteúdo e formato da informação a constar. Os campos para os quais não seja necessário preenchimento deverão ser completados com a referência “Não aplicável”;
- 1.2. Toda a discussão e alteração aos campos deste relatório, em particular quando haja lugar a pedido de elementos, deverá ser sempre incluída nos próprios campos e não sob a forma de adendas no final do relatório. Cada versão do relatório deve ser identificada e datada;
- 1.3. A secção “Conclusão” deverá ser sempre atualizado de forma a refletir, em cada momento, a conclusão da avaliação face a toda a informação submetida (incluindo avaliação da resposta aos pedidos de elementos, não devendo por isso contemplar conclusões anteriores do relatório, que serão rastreadas pelas datas de atualização dos relatórios). A secção “Pedido de elementos” deverá ser atualizado em cada versão, no sentido de referir apenas os pedidos que permaneçam em aberto, se existirem;
- 1.4. Apenas devem ser considerados os aspetos clínicos do medicamento, não devendo ser feita referência à avaliação económica nem a custos com a terapêutica ao longo deste relatório;
- 1.5. Não obstante o ponto anterior, deverão ser tidas em consideração as Orientações Metodológicas de Avaliação de Estudos Económicos de medicamentos, na elaboração deste relatório (ex. horizonte temporal, medidas de eficácia, etc.);

1.6. O conteúdo das várias secções do relatório alimentará a versão pública do relatório final de avaliação (que inclui também a avaliação económica e a decisão) que será disponibilizado na página eletrónica do Infarmed (sem prejuízo de eliminação de informação confidencial), pelo que a redação deve ser clara e concisa, para evitar diferentes interpretações e estar adequada à sua divulgação aos profissionais de saúde.

REFERÊNCIAS

1. GRADE guidelines: A new series of articles in the *Journal of Clinical Epidemiology*

Gordon H. Guyatt, Andrew D. Oxman, Holger J. Schünemann, Peter Tugwell, Andre Knottnerus

Vol. 64, Issue 4, p380–382

Published online: December 27 2010

2. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables

Gordon Guyatt, Andrew D. Oxman, Elie A. Akl, Regina Kunz, Gunn Vist, Jan Brozek, Susan Norris, Yngve Falck-Ytter, Paul Glasziou, Hans deBeer, Roman Jaeschke, David Rind, Joerg Meerpohl, Philipp Dahm, Holger J. Schünemann

Vol. 64, Issue 4, p383–394

Published online: January 3 2011

3. GRADE guidelines: 2. Framing the question and deciding on important outcomes

Gordon H. Guyatt, Andrew D. Oxman, Regina Kunz, David Atkins, Jan Brozek, Gunn Vist, Philip Alderson, Paul Glasziou, Yngve Falck-Ytter, Holger J. Schünemann

Vol. 64, Issue 4, p395–400

Published online: January 3 2011

4. GRADE guidelines: 3. Rating the quality of evidence

Howard Balshem, Mark Helfand, Holger J. Schünemann, Andrew D. Oxman, Regina Kunz, Jan Brozek, Gunn E. Vist, Yngve Falck-Ytter, Joerg Meerpohl, Susan Norris, Gordon H. Guyatt

Vol. 64, Issue 4, p401–406

Published online: January 6 2011

5. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias)

Gordon H. Guyatt, Andrew D. Oxman, Gunn Vist, Regina Kunz, Jan Brozek, Pablo Alonso-Coello, Victor Montori, Elie A. Akl, Ben Djulbegovic, Yngve Falck-Ytter, Susan L. Norris, John W. Williams Jr., David Atkins, Joerg Meerpohl, Holger J. Schünemann

Vol. 64, Issue 4, p407–415

Published online: January 20 2011

6. GRADE guidelines: 5. Rating the quality of evidence—publication bias

Gordon H. Guyatt, Andrew D. Oxman, Victor Montori, Gunn Vist, Regina Kunz, Jan Brozek, Pablo Alonso-Coello, Ben Djulbegovic, David Atkins, Yngve Falck-Ytter, John W. Williams Jr., Joerg Meerpohl, Susan L. Norris, Elie A. Akl, Holger J. Schünemann

Vol. 64, Issue 12, p1277–1282

Published online: August 1 2011

7. GRADE guidelines 6. Rating the quality of evidence—imprecision

Gordon H. Guyatt, Andrew D. Oxman, Regina Kunz, Jan Brozek, Pablo Alonso-Coello, David Rind, P.J. Devereaux, Victor M. Montori, Bo Freyschuss, Gunn Vist, Roman Jaeschke, John W. Williams Jr., Mohammad Hassan Murad, David Sinclair, Yngve Falck-Ytter, Joerg Meerpohl, Craig Whittington, Kristian Thorlund, Jeff Andrews, Holger J. Schünemann

Vol. 64, Issue 12, p1283–1293

Published online: August 12 2011

8. GRADE guidelines: 7. Rating the quality of evidence—inconsistency

Gordon H. Guyatt, Andrew D. Oxman, Regina Kunz, James Woodcock, Jan Brozek, Mark Helfand, Pablo Alonso-Coello, Paul Glasziou, Roman Jaeschke, Elie A. Akl, Susan Norris, Gunn Vist, Philipp Dahm, Vijay K. Shukla, Julian Higgins, Yngve Falck-Ytter, Holger J. Schünemann, The GRADE Working Group

Vol. 64, Issue 12, p1294–1302

Published online: August 2 2011

9. GRADE guidelines: 8. Rating the quality of evidence—indirectness

Gordon H. Guyatt, Andrew D. Oxman, Regina Kunz, James Woodcock, Jan Brozek, Mark Helfand, Pablo Alonso-Coello, Yngve Falck-Ytter, Roman Jaeschke, Gunn Vist, Elie A. Akl, Piet N. Post, Susan Norris, Joerg Meerpohl, Vijay K. Shukla, Mona Nasser, Holger J. Schünemann, The GRADE Working Group

Vol. 64, Issue 12, p1303–1310

Published online: August 1 2011

10. GRADE guidelines: 9. Rating up the quality of evidence

Gordon H. Guyatt, Andrew D. Oxman, Shahnaz Sultan, Paul Glasziou, Elie A. Akl, Pablo Alonso-Coello, David Atkins, Regina Kunz, Jan Brozek, Victor Montori, Roman Jaeschke, David Rind, Philipp Dahm, Joerg Meerpohl, Gunn Vist, Elise Berliner, Susan Norris, Yngve Falck-Ytter, M. Hassan Murad, Holger J. Schünemann, The GRADE Working Group

Vol. 64, Issue 12, p1311–1316

Published online: August 1 2011

11. GRADE guidelines: 10. Considering resource use and rating the quality of economic evidence

Massimo Brunetti, Ian Shemilt, Silvia Pregno, Luke Vale, Andrew D. Oxman, Joanne Lord, Jane Sisk, Francis Ruiz, Suzanne Hill, Gordon H. Guyatt, Roman Jaeschke, Mark Helfand, Robin Harbour, Marina Davoli, Laura Amato, Alessandro Liberati, Holger J. Schünemann

Vol. 66, Issue 2, p140–150

Published online: August 6 2012

12. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes

Gordon Guyatt, Andrew D. Oxman, Shahnaz Sultan, Jan Brozek, Paul Glasziou, Pablo Alonso-Coello, David Atkins, Regina Kunz, Victor Montori, Roman Jaeschke, David Rind, Philipp Dahm, Elie A. Akl, Joerg Meerpohl, Gunn Vist, Elise Berliner, Susan Norris, Yngve Falck-Ytter, Holger J. Schünemann

Vol. 66, Issue 2, p151–157

Published online: April 30 2012

13. GRADE guidelines: 12. Preparing Summary of Findings tables—binary outcomes

Gordon H. Guyatt, Andrew D. Oxman, Nancy Santesso, Mark Helfand, Gunn Vist, Regina Kunz, Jan Brozek, Susan Norris, Joerg Meerpohl, Ben Djulbegovic, Pablo Alonso-Coello, Piet N. Post, Jason W. Busse, Paul Glasziou, Robin Christensen, Holger J. Schünemann

Vol. 66, Issue 2, p158–172

Published online: May 21 2012

14. GRADE guidelines: 13. Preparing Summary of Findings tables and evidence profiles—continuous outcomes

Gordon H. Guyatt, Kristian Thorlund, Andrew D. Oxman, Stephen D. Walter, Donald Patrick, Toshi A. Furukawa, Bradley C. Johnston, Paul Karanicolas, Elie A. Akl, Gunn Vist, Regina Kunz, Jan Brozek, Lawrence L. Kupper, Sandra L. Martin, Joerg J. Meerpohl, Pablo Alonso-Coello, Robin Christensen, Holger J. Schunemann

Vol. 66, Issue 2, p173–183

Published online: October 31 2012

15. GRADE guidelines—an introduction to the 10th–13th articles in the series

Gordon H. Guyatt, Andrew D. Oxman, Holger J. Schünemann

Vol. 66, Issue 2, p121–123

Published online: September 10 2012

16. IQWiG – Institute for Quality and Efficiency in Health Care. General Methods. Version 4.2 of 22 April 2015. Acessível em

https://www.iqwig.de/download/IQWiG_General_Methods_Version_%204-2.pdf

17. Guideline Points to consider on switching between superiority and non-inferiority. Committee for Proprietary Medicinal Products (CPMP/EWP/482/99). European Medicines Agency

18. Note for guidance on statistical principles for clinical trials (CPMP/ICH/363/96). ICH Topic E 9. European Medicines Agency

19. Guideline Comparators & Comparisons: direct and indirect comparisons. EunetHTA. Adapted version (2015)

20. Glenny AM, Altman DG, Song F, Sakarovitch C et al. Indirect comparisons of competing interventions. Health Technology Assessment 2005; 9(26)

Anexo - Formulários

FORMULÁRIO DE AVALIAÇÃO INDIVIDUAL (POR ELEMENTO DO GRUPO DE AVALIAÇÃO) DA QUALIDADE DA EVIDÊNCIA SUBMETIDA (FORM 005-QE-RV)

Avaliação de benefício adicional da intervenção:

| Risco de Viés | | | |
|--|----------------|----------|-----------------|
| <i>Outcome:</i> | | | |
| | Sim | Não | Não sei |
| 1. Ausência de alocação oculta | | | |
| 2. Ausência de ocultação | | | |
| 3. Inclusão incompleta de doentes | | | |
| 4. Reporte seletivo de <i>outcomes</i> | | | |
| 5. Estudo interrompido por benefício | | | |
| 6. Uso de sub-rogados não validados | | | |
| | | | |
| 7. Risco de viés | | | |
| | Elevado | Moderado | Baixo |
| 8. No caso de responder sim ao ponto 7, como classifica o risco de viés | | | |
| | | | |
| | Sim | Não | |
| 9. Baixar a classificação da qualidade da evidência? | | | |
| | | | |
| | Baixar 1 nível | | Baixar 2 níveis |
| 10. No caso de responder sim ao ponto 9, escolha uma das seguintes opções: | | | |

Nome do Avaliador: _____

Assinatura do Avaliador: _____

Data ___/___/___

FORMULÁRIO DE AVALIAÇÃO GLOBAL (GRUPO DE AVALIAÇÃO) DA QUALIDADE DA EVIDÊNCIA SUBMETIDA (FORM 006-QE-RV)

Avaliação de benefício adicional da intervenção:

| Risco de Viés | | | |
|--|----------------|----------|-----------------|
| <i>Outcome:</i> | | | |
| | Sim | Não | Não sei |
| 1. Ausência de alocação oculta | | | |
| 2. Ausência de ocultação | | | |
| 3. Inclusão incompleta de doentes | | | |
| 4. Reporte seletivo de <i>outcomes</i> | | | |
| 5. Estudo interrompido por benefício | | | |
| 6. Uso de sub-rogados não validados | | | |
| | | | |
| 7. Risco de viés | | | |
| | Elevado | Moderado | Baixo |
| 8. No caso de responder sim ao ponto 7, como classifica o risco de viés | | | |
| | | | |
| | Sim | Não | |
| 9. Baixar a classificação da qualidade da evidência? | | | |
| | | | |
| | Baixar 1 nível | | Baixar 2 níveis |
| 10. No caso de responder sim ao ponto 9, escolha uma das seguintes opções: | | | |

Assinatura do coordenador do Grupo de Avaliação:

Data __/__/__

FORMULÁRIO DE AVALIAÇÃO INDIVIDUAL (POR ELEMENTO DO GRUPO DE AVALIAÇÃO) DA QUALIDADE DA EVIDÊNCIA SUBMETIDA (FORM 007-QE-IMP)

Avaliação de benefício adicional da intervenção:

| Imprecisão | | | |
|--|----------------|-----------------|-----|
| <i>Outcome:</i> | | | |
| | Sim | Não | ND* |
| 1. O intervalo de confiança a 95% é suficientemente estreito e exclui o efeito nulo | | | |
| 2. O número de participantes incluídos nos estudos em análise é igual ou superior ao “tamanho ótimo de informação” | | | |
| | | | |
| 3. Imprecisão (os resultados são imprecisos?) | | | |
| | | | |
| | Sim | Não | |
| 4. Baixar a classificação da qualidade da evidência? | | | |
| | | | |
| | Baixar 1 nível | Baixar 2 níveis | |
| 5. No caso de responder sim ao ponto 4, escolha uma das seguintes opções: | | | |

ND – Não determinado

Nome do Avaliador: _____

Assinatura do Avaliador: _____

Data ___/___/___

FORMULÁRIO DE AVALIAÇÃO GLOBAL (GRUPO DE AVALIAÇÃO) DA QUALIDADE DA EVIDÊNCIA SUBMETIDA (FORM 008-QE-IMP)

Avaliação de benefício adicional da intervenção:

| Imprecisão | | | |
|--|----------------|-----------------|-----|
| <i>Outcome:</i> | | | |
| | Sim | Não | ND* |
| 1. O intervalo de confiança a 95% é suficientemente estreito e exclui o efeito nulo | | | |
| 2. O número de participantes incluídos nos estudos em análise é igual ou superior ao “tamanho ótimo de informação” | | | |
| | | | |
| 3. Imprecisão (os resultados são imprecisos?) | | | |
| | | | |
| | Sim | Não | |
| 4. Baixar a classificação da qualidade da evidência? | | | |
| | | | |
| | Baixar 1 nível | Baixar 2 níveis | |
| 5. No caso de responder sim ao ponto 4, escolha uma das seguintes opções: | | | |

ND – Não determinado

Assinatura do Coordenador do Grupo de Avaliação:

Data ___/___/___

FORMULÁRIO DE AVALIAÇÃO INDIVIDUAL (POR ELEMENTO DO GRUPO DE AVALIAÇÃO) DA QUALIDADE DA EVIDÊNCIA SUBMETIDA (FORM 009-QE-INC)

Avaliação de benefício adicional da intervenção:

| Inconsistência | | | |
|---|----------------|-----|-----------------|
| <i>Outcome:</i> | | | |
| | Sim | Não | ND |
| 1. As estimativas de efeito variam substancialmente entre estudos | | | |
| 2. Os intervalos de confiança das estimativas de efeito não mostram qualquer sobreposição ou mostram apenas uma sobreposição mínima | | | |
| 3. Os testes estatísticos de heterogeneidade (teste Q) mostram um valor de P significativo | | | |
| 4. O I ² apresenta um valor elevado | | | |
| | | | |
| 5. Inconsistência (os resultados são inconsistentes?) | | | |
| | | | |
| | Sim | Não | |
| 6. Baixar a classificação da qualidade da evidência? | | | |
| | | | |
| | Baixar 1 nível | | Baixar 2 níveis |
| 7. No caso de responder sim ao ponto 6, escolha uma das seguintes opções: | | | |

ND – Não determinado

Nome do Avaliador: _____

Assinatura do Avaliador: _____

Data ___/___/___

FORMULÁRIO DE AVALIAÇÃO GLOBAL (GRUPO DE AVALIAÇÃO) DA QUALIDADE DA EVIDÊNCIA SUBMETIDA (FORM 010-QE-IMP)

Avaliação de benefício adicional da intervenção:

| Inconsistência | | | |
|---|----------------|-----|-----------------|
| <i>Outcome:</i> | | | |
| | Sim | Não | ND |
| 1. As estimativas de efeito variam substancialmente entre estudos | | | |
| 2. Os intervalos de confiança das estimativas de efeito não mostram qualquer sobreposição ou mostram apenas uma sobreposição mínima | | | |
| 3. Os testes estatísticos de heterogeneidade (teste Q) mostram um valor de P significativo | | | |
| 4. O I ² apresenta um valor elevado | | | |
| | | | |
| 5. Inconsistência (os resultados são inconsistentes?) | | | |
| | | | |
| | Sim | Não | |
| 6. Baixar a classificação da qualidade da evidência? | | | |
| | | | |
| | Baixar 1 nível | | Baixar 2 níveis |
| 7. No caso de responder sim ao ponto 6, escolha uma das seguintes opções: | | | |

ND – Não determinado

Assinatura do Coordenador do Grupo de Avaliação:

Data __/__/__

FORMULÁRIO DE AVALIAÇÃO INDIVIDUAL (POR ELEMENTO DO GRUPO DE AVALIAÇÃO) DA QUALIDADE DA EVIDÊNCIA SUBMETIDA (FORM 011-QE-IND)

Avaliação de benefício adicional da intervenção:

| Evidência indireta | | |
|--|----------------|-----------------|
| <i>Outcome:</i> | | |
| | Sim | Não |
| 1. A população de estudo é diferente da população de interesse? | | |
| 2. A intervenção testada é diferente da intervenção de interesse? | | |
| 3. O estudo utiliza <i>outcomes</i> sub-rogados? | | |
| 4. A comparação do efeito da intervenção de interesse e do comparador não é feita diretamente? | | |
| | | |
| 5. Evidência indireta (existe evidência indireta?) | | |
| | | |
| | Sim | Não |
| 6. Baixar a classificação da qualidade da evidência? | | |
| | | |
| | Baixar 1 nível | Baixar 2 níveis |
| 7. No caso de responder sim ao ponto 6, escolha uma das seguintes opções: | | |

Nome do Avaliador: _____

Assinatura do Avaliador: _____

Data ___/___/___

FORMULÁRIO DE AVALIAÇÃO GLOBAL (GRUPO DE AVALIAÇÃO) DA QUALIDADE DA EVIDÊNCIA SUBMETIDA (FORM 012-QE-IND)

Avaliação de benefício adicional da intervenção:

| Evidência indireta | | |
|--|----------------|-----------------|
| <i>Outcome:</i> | | |
| | Sim | Não |
| 1. A população de estudo é diferente da população de interesse? | | |
| 2. A intervenção testada é diferente da intervenção de interesse? | | |
| 3. O estudo utiliza <i>outcomes</i> sub-rogados? | | |
| 4. A comparação do efeito da intervenção de interesse e do comparador não é feita diretamente? | | |
| | | |
| 5. Evidência indireta (existe evidência indireta?) | | |
| | | |
| | Sim | Não |
| 6. Baixar a classificação da qualidade da evidência? | | |
| | | |
| | Baixar 1 nível | Baixar 2 níveis |
| 7. No caso de responder sim ao ponto 6, escolha uma das seguintes opções: | | |

Assinatura do Coordenador do Grupo de Avaliação:

Data __/__/__