Summary HTA

german agency for HTA of
German Institute of Medical
Documentation and Information

HTA-Report | Summary

Indirect comparisons of therapeutic interventions

Schöttker B, Lühmann D, Boulkhemair D, Raspe H

German Agency for HTA
at DIMDI (DAHTA)
Waisenhausgasse 36-38a
50676 Köln
Germany

Tel.: +49 221 4724-525
Fax: +49 221 4724-444
dahta@dimdi.de
www.dimdi.de

All HTA reports are available for free of charge as full texts in the DAHTA database (only in German) and at German Medical Science (GMS).

Health political background

In the system of statutory health insurance coverage decisions are increasingly based on the results of effectiveness or cost-effectiveness analyses conducted in the context of Health Technology Assessments (HTA).
Randomised controlled head-to-head trials which directly compare the effects of different therapies are considered the gold standard methodological approach for the comparison of the efficacy of medical interventions. While research progresses, more and more treatment options are being developed for certain indications. As concerns pharmacological interventions, proven positive effects compared to placebo may be sufficient to attain market approval. Therefore manufacturers rarely see the need to test the effects of new interventions against the effects of interventions that are in the market already. Given multiple therapeutic options for an indication, there will hardly be a head-to-head trial testing all options in parallel. Statements on comparative efficacy have to rely on indirect comparisons.

Scientific background

Comparisons are defined as indirect if the effects of interventions are compared to each other by their performance against a common comparator. This may be an active intervention (usually standard care) or placebo. Up to date many questions concerning the validity of indirect comparisons remain unanswered. In 2005 a British HTA-report was published, containing a comprehensive systematic overview of available methods for indirect comparisons and their validity. The report, which refers to publications up to 1999, introduces three methodological approaches for indirect comparisons: unadjusted and adjusted indirect comparisons, and metaregression-analyses. The authors conclude that discrepancies between the results of direct and indirect comparisons are considerable but their direction cannot be foreseen. It is pointed out that unadjusted indirect comparisons are highly prone to bias. Contrasting, adjusted indirect comparisons and metaregression-analyses provide a higher degree of validity.
On the basis of these results the current report gives an updated review of indirect comparisons by means of five research questions. It focuses on the comparative efficacy of medical interventions on the basis of high-quality randomised controlled trials (RCT).

Within the scope of the

Bundesministerium
für Gesundheit

Research questions

1. What methodological approaches for indirect comparisons of therapeutic interventions are available today (March 2008) and under what circumstances may they be applied?
2. What methodological approaches for indirect comparisons have been applied in systematic reviews and how often?
3. What is the validity of results from indirect comparisons compared to the results of direct comparisons and do both arrive at the same conclusions?
4. What is the validity of results from indirect comparisons compared to the results of direct comparisons if results from head-to-head trials are included in the indirect comparison?
5. Is it possible to identify a “gold standard methodology“ for indirect comparisons of competing interventions?

Methods

Systematic literature searches are conducted with two purposes:

1. Identification of papers describing methodological approaches for indirect comparisons.
2. Identification of systematic reviews which apply indirect comparisons (exclusively, or in addition to information from direct comparisons.

The basis of relevant references is extracted from the systematic review of Glenny et al. which covers the relevant literature up to 1999. To identify papers published after 1999 all medical databases of the German Institute of Medical Documentation and Information (DIMDI) and the ISI Web of Knowledge® are searched using the search strategy of Glenny et al. with minor modifications.
In addition, reference lists of the main methodological papers and systematic reviews as well as the homepages of the member institutions of the International network of agencies for Health Technology Assessment (INAHTA) are screened for relevant papers.
The description of the different methodological approaches for indirect comparisons is based as far as possible on information from methodological papers and completed by information from methods chapters of published applications. Their application frequency is calculated by counting the number of applications in all systematic reviews with indirect comparisons published 1999 to 2008.
Indirect comparisons which use metaanalysis techniques are validated empirically on the basis of systematic reviews that report results of direct as well as indirect comparisons. For every methodological approach the following hypothesis is tested: the results of the indirect comparison do not differ significantly from the results of the direct comparison. In order to test this hypothesis the difference in the results of a direct and an indirect comparison for the same intervention is calculated. This difference is named discrepancy. In order to make discrepancies from different reviews comparable, they are transformed into z-scores. The final validity check for the different methodological approaches for indirect comparisons was performed in four steps.

1. Test for systematic over- or underestimation: Are the z-scores nor-

mally distributed with an average value of z = 0 (Kolmogorov-Smirnov-Test, $p \leq 0.05$)?

2. Quantification of the amount of discrepancy: Calculation of the mean absolute value of z ($|\overline{z}|$).
3. Determination of the share of statistically discrepant z-scores ($|\overline{z}| \geq 1.96$) among all z-scores.
4. For data sets with statistically significant discrepant z-scores: Homogeneity testing of the underlying study pool for the direct and indirect comparisons.

Finally it is reported in how many cases the direct and indirect comparisons arrive at the same conclusions.
While it is assumed that inclusion of head-to-head trials into indirect comparisons may level out discrepancies between direct and indirect comparison, the validity check (main analysis) is repeated in a subgroup of data sets (subgroup analysis), which do not include results from head-to-head trials into indirect comparisons.

Results

Method descriptions

Literature reveals that all methodological approaches for indirect comparisons are based on the same assumption: The observed variability among the results of studies that are going to be included into an indirect comparison is solely due to random error or - in other words - no meaningful between-study heterogeneity is present.
Four frequently applied methodological approaches for indirect comparisons, which use metaanalytical methods, are identified:

1. In an unadjusted indirect comparison the comparison of an intervention A with an intervention B is prepared by metaanalytically pooling the results of all study arms treated with A to get a summary estimate θ_A and by doing the same in a second metaanalysis with all study arms treated with B to get θ_B. This procedure is called “unadjusted indirect comparison” because the indirect comparison is not adjusting for events in the control group. There are four ways of comparing the summary effect estimates θ_A and θ_B: calculation of a summary effect estimate $\theta_{A \text{ versus } B}$; testing the difference between θ_A and θ_B for statistical significance; check the confidence intervals around θ_A and θ_B for overlap or a narrative comparison of the efficacy of A and B.
2. To perform an adjusted (for events in the comparator arms) indirect comparison the summary effect estimates $\theta_{A \text{ versus comparator}}$ and $\theta_{B \text{ versus comparator}}$ are calculated by conventional metaanalytic methods. For the comparison of the two summary effect estimates the same four methods as introduced in point 1 are applicable.
3. In metaregression-analyses the summary effect estimates $\theta_{A \text{ versus comparator}}$ and $\theta_{B \text{ versus comparator}}$ are estimated separately in two regression equations. In addition to adjusting for effects in the comparator arms the regression models can adjust for the effects of further covariates (which are regarded as the origin of heterogeneity – like i. e. age of study population or severity of illnesses). Again, the comparison of $\theta_{A \text{ versus comparator}}$ and $\theta_{B \text{ versus comparator}}$ is performed by the above mentioned four methods (see 1.).
4. Mixed treatment comparison (MTC) is a collective term for methodological approaches for indirect comparisons comparing more than

two interventions simultaneously and possibly including head-to-head studies. MTC are able to rank an unlimited number of therapeutic options according to their efficacy. For that purpose Bayesian statistics are applied to successively pool all available evidence from RCT in order to gain summary effect estimates for all possible comparisons of the interventions of interest.

Indirect comparisons without metaanalysis are performed if there is only one trial available for the options of interest or if available studies are highly heterogeneous. Indirect comparisons without metaanalysis also follow the principles of adjusted or unadjusted comparisons and may be performed by the four methods introduced in 1.

Application frequency of different methodological approaches for indirect comparisons

In 106 systematic reviews published between January 1999 and February 2008 found by the literature searches, one metaanalytic method of an indirect comparison is applied (exception: Vandermeer et al. 2007 applied three different methods). The considerably most frequently applied method is the adjusted indirect comparison (60 times), followed by metaregression-analyses (17 times), unadjusted indirect comparisons (14 times), MTC (twelve times) and other approaches which cannot be allocated to the four main methodological groups (five times). In 2006 a steep rise in the utilisation of MTC is observed (ten examples from 2006 until 2007).

Validity check

For the validity check of the indirect approaches a total of 248 paired results from direct and indirect comparisons (z-scores) are available from 57 systematic reviews.

The test for systematic over- or underestimation reveals that none of the approaches for indirect comparisons systematically over- or underestimates the results of a corresponding direct comparison. Nevertheless, differences in the mean absolute z-scores are observed among the indirect methods: The largest are found with the unadjusted indirect comparisons ($|\overline{z}|$ = 1.63 [95 %-CI: 1.20; 2.07]) while adjusted indirect comparisons ($|\overline{z}|$ = 0.95 [95 %-CI: 0.80; 1.09]), metaregression-analyses ($|\overline{z}|$ = 0.99 [95 %-CI: 0.20; 1.79]) and MTC (n = 57; $|\overline{z}|$ = 0.59 [95 %-CI: 0.45; 0.73]) provide lower values. For the MTC a higher average z-score is observed in the subgroup analysis without inclusion of head-to-head trials (n = 12; $|\overline{z}|$ = 0.83 [95 %-CI: 0.40; 1.26]) while the results of the main and subgroup analyses are concordant for the other methods. It is to be noted though that the variance of the mean absolute z-scores differs considerably across the methods. The number of outstandingly high z-scores ($|\overline{z}| > 1.96$) varies among the indirect methodological approaches: the unadjusted indirect comparison provides a share of 25.5 % (n = 47; 95 %-CI: 13.1 %; 38.0 %) of statistically significant discrepant z-scores, the adjusted indirect comparison of 12.1 % (n = 116; 95 %-CI: 6.1 %; 18.0 %), the metaregression-analysis of 16.7 % (n = 6; 95 %-CI: -13.2 %; 46.5 %) and the MTC of 1.8 % (n = 57; 95 %-CI: 2.1 %; 34.3 %). The results from the main and subgroup analysis are concordant. Summarising all indirect methods, 32 of 248 comparisons provide statistically significant discrepancies (12.9 % [95 %-CI: 8.7%; 17.1%]).

For 15 of the 32 statistically significant discrepancies (z-scores) no information concerning heterogeneity of the pooled studies is given by the original review authors. Proof of significant heterogeneity is found by the original review authors in eleven of the statistically significant discrepant comparisons but not in the remaining six.

Congruence of conclusions

In about half of the 248 comparisons of interventions no statistically significant difference in therapeutic efficacy is found - neither by direct nor by indirect comparison (49.2 %; 95 %-CI: 43.0 %; 55.4 %). In 21.8 % (95 %-CI: 16.6 %; 26.9 %) of cases one intervention is found to perform significantly better than the other by both the direct and the indirect comparison. In another 29 % (95 %-CI: 23.4 %; 34.7 %) of the analysed comparisons the conclusions of the direct and indirect comparison are not concordant. However the feared case that the direct comparison prefers the one and the indirect comparison the other intervention with statistical significance is observed rarely (five cases; corresponding to a share of 2 % (95 %-CI: 0.3 %; 3.8 %) among all cases.

Precision of indirect comparisons

In the analysed sample (n = 248) the confidence intervals around the effect estimates of the indirect comparisons are found to be slightly smaller than those around the direct estimates (median difference: 9 % (25th percentile: -34 %; 75th percentile: 30 %) while the indirect comparisons include six times more studies than the direct comparisons (median: 6 (25th percentile: 4; 75th percentile: 13)). It may therefore be stated that for the analysed sample a six to one ratio of included studies (with an approximately equal number of participants) for the indirect and direct comparison yields almost comparable precision of effect estimates. This supports the claim of Glenny et al. that an indirect comparison must include four times as many studies (of equal size) as a direct comparison to yield the same precision.

Discussion

In decision making whether, and if so, which approach of indirect comparisons should be applied, four criteria should be taken into consideration:

1. Validity of the methodological approach

Compared to the results of head-to-head trials unadjusted indirect comparison provide the lowest validity. Some authors blame the method for breaking the randomisation of the included RCT because effects are not adjusted for events in the control groups. Therefore results are easily distorted by all types of biases that are normally typical for observational studies (i. e. selection bias and confounding).

In contrast the adjusted indirect comparison, the metaregression and the MTC adjust for events in the control groups and hereby preserve the randomisation of the included RCT. However, a selection bias on the meta-level may still appear if the included studies for one intervention use different inclusion criteria than the studies for the other intervention. The resulting unevenly distributed patient characteristics may, if they are associated with the outcome, act as confounders. Therefore the introduced methods for indirect comparisons should be applied only if the results that are going to be pooled are extracted from homogeneous studies. This prerequisite holds not only for the methodological approaches to indirect comparisons but for conventional metaanalyses as well.

These theoretical aspects are supported by the results of the empirical validity check. Adequate numbers of data were available to support the hypothesis that – provided a homogeneous pool of studies – adjusted indirect comparisons may arrive at the same results as direct comparisons.

Likewise a high validity can be ascribed to MTC, if they include head-to-head studies with the interventions of interest. The validity of metaregression-analyses, MTC without included head-to-head trials and the rarely used other methods cannot be appraised yet due to a limited number of available applications.

2. Number of therapies to compare

If only two interventions are to be compared indirectly the adjusted indirect comparison seems to be the most appropriate methodological approach considering the validity data and the limited methodological effort. If more than two interventions are to be compared, only a MTC is applicable to rank them in order of their efficacy.

3. If results from head-to-head trials are to be included

Beside MTC the three other methods for indirect comparisons also provide methodological extensions for the inclusion of head-to-head trials into an indirect comparison. However there haven't been sufficient data for a check of their validity. It can only be stated yet that MTC which include head-to-head trials yield similar results as the head-to-head trial alone. Their additional advantage is the possible increase in precision of the effect estimate by combining the results of direct and indirect comparisons.

4. Heterogeneous trials

The indirect comparison by metaregression-analysis cannot yet be regarded a sufficiently validated method that trustworthily adjusts for factors that cause heterogeneity. Likewise adjusting for covariates in MTC by introduction of inconsistency factors has not been validated due to the limited number of applications. In conclusion: If considerable heterogeneity is present among the trials, the risk of bias in indirect comparisons is high – regardless of what methodological approach is used. In cases of low heterogeneity a conservative estimate may be calculated by a random effects model. Fixed effects models should only be applied in homogenous pools of studies. Both models are applicable in all methodological approaches for indirect comparisons described.

Conclusions

There are a number of methodological approaches available for indirect comparisons which differ in their ability to summarize the evidence from different pools of studies.

The empirical investigation reveals that mainly the results of unadjusted indirect comparisons differ from the results of direct comparisons. The other indirect methods may provide concordant results with direct comparisons, especially if the summarized studies are characterized by low heterogeneity. For that reason adjusted indirect comparisons, metaregression-analyses and MTC should only be used when study results are homogeneous. In the context of HTA and the development of clinical guidelines they are valuable tools, if direct evidence for a comparison of efficacy of interventions is not available.

Before indirect comparisons can be applied more broadly, it remains to be defined at which amount of heterogeneity (and inconsistency) they provide effect estimates of acceptable validity - because a perfectly homogeneous pool of studies is rarely found in real life.