

Indirekte Vergleiche von Therapieverfahren

Indirect comparisons of therapeutic interventions

Abstract

Health political background

The comparison of the effectiveness of health technologies is not only laid down in German law (Social Code Book V, § 139 and § 35b) but also constitutes a central element of clinical guidelines and decision making in health care. Tools supporting decision making (e. g. Health Technology Assessments (HTA)) are therefore in need of a valid methodological repertoire for these comparisons.

Scientific background

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 interventions. Because this type of trial is rarely found, comparisons of efficacy often need to rely on indirect comparisons whose validity is being controversially debated.

Research questions

Research questions for the current assessment are: Which (statistical) methods for indirect comparisons of therapeutic interventions do exist, how often are they applied and how valid are their results in comparison to the results of head-to-head trials?

Methods

In a systematic literature research all medical databases of the German Institute of Medical Documentation and Information (DIMDI) are searched for methodological papers as well as applications of indirect comparisons in systematic reviews. Results of the literature analysis are summarized qualitatively for the characterisation of methods and quantitatively for the frequency of their application.

The validity of the results from indirect comparisons is checked by comparing them to the results from the gold standard – a direct comparison. Data sets from systematic reviews which use both direct and indirect comparisons are tested for consistency by of the z-statistic.

Results

29 methodological papers and 106 applications of indirect methods in systematic reviews are being analysed. Four methods for indirect comparisons can be identified:

1. Unadjusted indirect comparisons include, independent of any comparator, all randomised controlled trials (RCT) that provide a study arm with the intervention of interest.
2. Adjusted indirect comparisons and
3. metaregression analyses include only those studies that provide one study arm with the intervention of interest and another study

Ben Schöttker¹

Dagmar Lühmann¹

Dalila Boulkhemair¹

Heiner Raspe¹

¹ Institut für Sozialmedizin,
Universität zu Lübeck,
Lübeck, Deutschland

arm with a common comparator. While the aforementioned methods use conventional metaanalytical techniques,

4. Mixed treatment comparisons (MTC) use Bayesian statistics. They are able to analyse a complex network of RCT with multiple comparators simultaneously.

During the period from 1999 to 2008 adjusted indirect comparisons are the most commonly used method for indirect comparisons. Since 2006 an increase in the application of the more methodologically challenging MTC is being observed.

For the validity check 248 data sets, which include results of a direct and an indirect comparison, are available. The share of statistically significant discrepant results is greatest in the unadjusted indirect comparisons (25,5% [95% CI: 13,1%; 38%]), followed by metaregression analyses (16,7% [95% CI: -13,2%; 46,5%]), adjusted indirect comparisons (12,1% [95% CI: 6,1%; 18%]) and MTC (1,8% [95% CI: -1,7%; 5,2%]). Discrepant results are mainly detected if the basic assumption for an indirect comparison – between-study homogeneity – does not hold. However a systematic over- or underestimation of the results of direct comparisons by any of the indirectly comparing methods was not observed in this sample.

Discussion

The selection of an appropriate method for an indirect comparison has to account for its validity, the number of interventions to be compared and the quality as well as the quantity of available studies. Unadjusted indirect comparisons provide, contrasted with the results of direct comparisons, a low validity. Adjusted indirect comparisons and MTC may, under certain circumstances, give results which are consistent with the results of direct comparisons. The limited number of available reviews utilizing metaregression analyses for indirect comparisons currently prohibits empirical evaluation of this methodology.

Conclusions/Recommendations

Given the main prerequisite – a pool of homogenous and high-quality RCT – the results of head-to-head trials may be pre-estimated by an adjusted indirect comparison or a MTC. In the context of HTA and guideline development they are valuable tools if there is a lack of a direct comparison of the interventions of interest.

Zusammenfassung

Gesundheitspolitischer Hintergrund

Die vergleichende Nutzenbewertung von gesundheitsrelevanten Technologien ist nicht nur gesetzlich festgeschrieben (§ 139 und § 35b, SGB V; SGB = Sozialgesetzbuch), sondern ist auch ein zentrales Element von klinischen Leitlinien bzw. Entscheidungssituationen. Entscheidungsunterstützungsinstrumente wie Health Technology Assessments (HTA) sollten daher über ein valides methodisches Instrumentarium verfügen.

Wissenschaftlicher Hintergrund

Randomisierte kontrollierte Head-to-head-Studien, die Therapien direkt miteinander vergleichen, gelten als Goldstandard für den Wirksamkeitsvergleich. Da Studien dieses Typs nur begrenzt zur Verfügung stehen, sind Wirksamkeitsvergleiche auf indirekt vergleichende Methoden angewiesen, deren Validität allerdings noch kontrovers diskutiert wird.

Fragestellung

Fragestellungen für das vorliegende Assessment sind: Welche (statistischen) Methoden zur Durchführung indirekter Vergleiche therapeutischer Interventionen gibt es, wie häufig werden sie eingesetzt und wie ist ihre Validität im Vergleich zu den Ergebnissen direkter Vergleiche zu beurteilen?

Methodik

In einer systematischen Literaturrecherche werden die medizinnahen Datenbanken des Deutschen Instituts für medizinische Dokumentation und Information (DIMDI) nach Methodenpapieren und Anwendungen von indirekten Vergleichen in systematischen Reviews durchsucht. Die Literaturlauswertung erfolgt qualitativ beschreibend (Methoden) und quantitativ für die Häufigkeit ihres Einsatzes.

Eine Validitätsprüfung der Methoden für indirekte Therapievergleiche ist über den Vergleich ihrer Ergebnisse mit dem Goldstandard - den Ergebnissen von Head-to-head-Studien - möglich. In systematischen Reviews, in denen Therapieverfahren sowohl direkt als auch indirekt verglichen werden, werden diese Ergebnisse mithilfe der z-Statistik auf Übereinstimmung geprüft.

Ergebnisse

29 Methodenpapiere und 106 Methodenanwendungen werden ausgewertet. Aus diesen lassen sich vier Methoden für indirekte Vergleiche identifizieren:

1. Nicht-adjustierte indirekte Vergleiche schließen, unabhängig vom Komparator, alle randomisierten kontrollierten Studien (RCT), die einen Studienarm mit einer der interessierenden Therapieoptionen enthalten, ein.
2. Adjustierte indirekte Vergleiche und
3. Metaregressionen greifen nur auf Studien zurück, die einen Arm mit einer Therapieoption von Interesse und einen Arm mit einem gemeinsamen Komparator aufweisen.
4. Während die genannten Verfahren konventionelle Metaanalysetechniken einsetzen, können die mit Bayes'schen Methoden arbeitenden Mixed treatment comparisons (MTC), ein komplexes Netzwerk aus RCT mit multiplen Komparatoren simultan analysieren.

Im Zeitraum von 1999 bis 2008 werden die adjustierten indirekten Vergleiche am häufigsten angewendet. Seit 2006 ist auch ein deutlicher Anstieg der Verwendung des methodisch anspruchsvolleren MTC zu verzeichnen.

Für die Validitätsprüfung stehen 248 Datensätze mit Ergebnisgegenüberstellungen aus direktem und indirektem Vergleich zur Verfügung. Der Anteil diskrepanter Ergebnisse mit statistischer Signifikanz war am größten bei den nicht-adjustierten indirekten Vergleichen (25,5% [95%-KI: 13,1; 38]), gefolgt von dem der Metaregressionen (16,7% [95%-KI: -13,2; 46,5]), der adjustierten indirekten Vergleiche (12,1% [95%-KI: 6,1; 18]) und des MTC (1,8% [95%-KI: -1,7; 5,2]). Diskrepante Ergebnisse werden vor allem dann beobachtet, wenn die Voraussetzung für die Durchführung eines indirekten Vergleichs – eine homogene Studienlage – nicht gegeben ist. Eine systematische Über- oder Unterschätzung der Ergebnisse direkter Vergleiche durch den indirekten Vergleich zeigt sich in dieser Stichprobe bei keinem der genannten Verfahren.

Diskussion

Die Auswahl einer geeigneten Methode für einen indirekten Vergleich hat sich an deren Validität, der Anzahl der zu vergleichenden Therapieoptionen sowie an der Qualität und Quantität der verfügbaren Studien zu orientieren. Nicht-adjustierte indirekte Vergleiche weisen in Gegenüberstellung zu direkten Vergleichen eine geringe Validität auf. Adjustierte indirekte Vergleiche und MTC können dagegen unter bestimmten Voraussetzungen Ergebnisse liefern, die in den meisten Fällen denen direkter Vergleiche entsprechen. Die Validität von indirekten Vergleichen mittels Metaregression kann auf Basis der wenigen, bisher verfügbaren Anwendungsbeispiele noch nicht beurteilt werden.

Schlussfolgerung/Empfehlung

Bei Einhaltung der zentralen Voraussetzung – Anwendung an einem Pool homogener, qualitativ hochwertiger RCT – lassen sich die Ergebnisse von qualitativ hochwertigen Head-to-head-Studien durch den Einsatz adjustierter indirekter Vergleiche und MTC abschätzen. Im Kontext von HTA und Leitlinienerstellung stellen sie somit wertvolle Hilfsinstrumente dar, wenn direkte Evidenz für einen Wirksamkeitsvergleich von Therapieverfahren nicht zur Verfügung steht.

Schlüsselwörter: randomisierte kontrollierte Studien, Metaanalyse, Literaturübersicht

Executive Summary

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).

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 normally 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 (absolute value of the mean of z-scores).
3. Determination of the share of statistically discrepant z-scores (absolute value of the mean of z-scores ≥ 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 re-

gression 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 is 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, a 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 (absolute value of the mean of z-scores = 1.63 [95%-CI: 1.20; 2.07]) while adjusted indirect comparisons (absolute value of the mean of z-scores = 0.95 [95 %-CI: 0.80; 1.09]), metaregression-analyses (absolute value of the mean of z-scores = 0.99 [95%-CI: 0.20; 1.79]) and MTC (n=57; absolute value of the mean of z-scores = 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 ($n=12$; absolute value of the mean of z-scores = 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 (absolute value of the mean of z-scores > 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 meta-regression-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 meta-regression 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 meta-regression-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.

Kurzfassung

Gesundheitspolitischer Hintergrund

Entscheidungen über die Kostenübernahme für medizinische Maßnahmen durch die Solidargemeinschaft werden verstärkt auf Basis von Nutzen- oder Kosten-Nutzen-Bewertungen im Rahmen von Health Technology Assessments (HTA) gestützt.

Randomisierte kontrollierte Head-to-head-Studien, die Therapien direkt miteinander vergleichen, gelten als Goldstandard für den Wirksamkeitsvergleich. Mit dem medizinisch-wissenschaftlichen Fortschritt steigt jedoch die Zahl der Therapieoptionen für eine bestimmte Indikation, ohne dass die regulativen Vorgaben eine sorgfältige Prüfung ihrer Wirksamkeit gegen bereits etablierte Therapieformen vorsehen. In der Pharmakotherapie reicht häufig ein Vergleich der Wirksamkeit neuer Arzneistoffe gegen Placebo für eine Zulassung aus. Daher sehen die Arzneimittelhersteller bislang nur selten Veranlassung, ihre neuen Arzneistoffe gegen bereits im Markt befindliche Arzneistoffe zu testen. Bei multiplen Behandlungsmöglichkeiten für eine Indikation wird es deshalb in der Regel keine randomisierten kontrollierten Studien (RCT) geben, die sämtliche Therapiemöglichkeiten parallel, d. h. direkt miteinander vergleichen. Soll dennoch eine Aussage zur vergleichenden Wirksamkeit gemacht werden, muss auf indirekt vergleichende Verfahren zurückgegriffen werden.

Wissenschaftlicher Hintergrund

Unter indirekten Vergleichen versteht man die Gegenüberstellung der Wirksamkeit verschiedener Interventionen mithilfe eines gemeinsamen Komparators. Hierbei kann es sich um einen aktiven Komparator (meistens Standardtherapie) oder um Placebo handeln.

Viele Fragen zur Validität indirekter Vergleiche sind aber bisher noch ungeklärt. Eine umfassende systematische Übersicht über verfügbare Methoden für indirekte Vergleiche und ihre Validität wird in dem 2005 publizierten britischen HTA-Bericht von Glenny et al. gegeben. Dieser Bericht, der drei Methoden für indirekte Vergleiche (nicht-adjustierte und adjustierte indirekte Vergleiche sowie Metaregressionen) vorstellt, schließt Publikationen bis 1999 ein. Die Autoren kommen aufgrund des damaligen Kenntnis- und Erfahrungsstandes zu dem Schluss, dass zwischen den Ergebnissen direkter und indirekter Vergleiche deutliche Diskrepanzen sichtbar werden, deren Richtung aber nicht voraussehbar ist. Sie weisen auf eine hohe Anfälligkeit für systematisch verzerrte Ergebnisse in nicht-adjustierten indirekten Vergleichen hin. Adjustierte indirekte Vergleiche und Metaregressionen zeigen dagegen eine deutlich höhere Validität.

Auf der Grundlage dieser Vorarbeiten gibt der vorliegende Bericht anhand von fünf Forschungsfragen einen Überblick über den aktuellen Kenntnisstand. Dabei wird unter anderem aus pragmatischen Gründen auf indirekte Ver-

gleiche fokussiert, die Aussagen zur Wirksamkeit therapeutischer Verfahren machen und hierzu die Ergebnisse von RCT heranziehen.

Forschungsfragen

1. Welche Methoden zur Durchführung indirekter Vergleiche therapeutischer Interventionen existieren bisher (März 2008) und unter welchen Voraussetzungen können sie eingesetzt werden?
2. Welche Methoden des indirekten Wirksamkeitsvergleichs therapeutischer Interventionen wurden wie häufig in bisher publizierten systematischen Reviews eingesetzt?
3. Wie ist die Validität indirekter Vergleiche gegenüber den Ergebnissen von direkten Vergleichen zu beurteilen und führen sie zu den gleichen Schlussfolgerungen über Therapieeffektunterschiede?
4. Wie ist die Validität indirekter Vergleiche gegenüber den Ergebnissen von direkten Vergleichen zu beurteilen, wenn in den indirekten Vergleichen zusätzlich Ergebnisse von Head-to-head-Studien eingebracht werden?
5. Lässt sich ein Goldstandard identifizieren, nachdem indirekte Vergleiche der Wirksamkeit von therapeutischen Interventionen vorgenommen werden sollten?

Methodik

Es werden systematische Literaturrecherchen mit zwei (bzw. drei) Zielen durchgeführt:

1. Auffinden von Methodenpapieren zu indirekten Vergleichen.
2. Auffinden von systematischen Literaturübersichten, in denen indirekte Vergleiche eingesetzt werden (ausschließlich, oder in Ergänzung zu den Informationen aus direkten Vergleichen).

Der Grundstock an relevanten Referenzen wird dem systematischen Review von Glenny et al. entnommen, der die Literatur bis 1999 abdeckt. Für die nach diesem Datum erschienene Literatur werden die medizinischen Datenbanken des Deutschen Instituts für Medizinische Dokumentation und Information (DIMDI) und des ISI Web of Knowledge® durchsucht.

Zur Ergänzung werden Handsuchen in den Referenzlisten wichtiger Methodenpapiere, sowie eine Recherche auf den Internetseiten der Mitgliedsorganisationen des International network of agencies for Health Technology Assessment (INAHTA), durchgeführt.

Die Beschreibung der einzelnen indirekt vergleichenden Methoden erfolgt, soweit vorhanden, auf Basis der Methodenpapiere und wird ergänzt um Angaben aus den Methodenbeschreibungen von Anwendungsbeispielen. Die Bestimmung der Häufigkeit des Einsatzes der einzelnen indirekt vergleichenden Methoden erfolgt anhand der gefundenen systematischen Reviews mit indirekten Vergleichen.

Indirekt vergleichende Methoden, die mit metaanalytischen Techniken arbeiten, werden einer empirischen Validitätsprüfung unterzogen. Hierzu werden systematische Literaturübersichten analysiert, in denen Therapieoptionen sowohl anhand der Ergebnisse von Head-to-head-Studien als auch indirekt, unter Hinzuziehung eines gemeinsamen Komparators, gegenübergestellt werden. Für jede indirekt vergleichende Methode wird die Hypothese geprüft, ob sich die Ergebnisse des indirekten von denen des direkten Vergleichs signifikant unterscheiden. Um diese Hypothese zu testen, wird zunächst für jedes Ergebnis eines direkten und indirekten Vergleichs zu dem gleichen Therapievergleich (Datensatz) die Differenz der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich gebildet; in der Folge Diskrepanz genannt. Anschließend werden, um die Diskrepanzen miteinander vergleichbar zu machen, z-Werte berechnet und in vier Schritten die Validität der Ergebnisse der verschiedenen indirekt vergleichenden Methoden denen der direkten Vergleiche gegenübergestellt.

1. Test auf systematische Über- oder Unterschätzung: Sind die z-Werte (Diskrepanzen) normalverteilt um einen Mittelwert $z = 0$ (Kolmogorov-Smirnov-Test, $p \leq 0,05$)?
2. Quantifizierung des Ausmaßes der Diskrepanz: Berechnung der mittleren Beträge von z (Betrag des Mittelwerts der z-Werte).
3. Bestimmung des Anteils der statistisch signifikant diskrepanten Datensätze (Betrag des Mittelwerts der z-Werte $\geq 1,96$) an allen für die Validitätsprüfung einer Methode herangezogenen Datensätzen.
4. In statistisch signifikant diskrepanten Datensätzen: Überprüfung der Homogenität des Studienpools für sowohl den direkten als auch für den indirekten Vergleich.

Abschließend wird berichtet, in wie vielen Fällen der direkte und der indirekte Vergleich zu den gleichen Schlussfolgerungen kommen.

Da vermutet wird, dass der Einschluss von Head-to-head-Studien in indirekte Vergleiche deren Ergebnisse nivelliert (d. h. an die Ergebnisse der Head-to-head-Studien annähert), wird die Validitätsprüfung (Hauptanalyse) an einer Subgruppe von Datensätzen (Subgruppenanalyse) wiederholt, in denen keine Head-to-head-Studien in den indirekten Vergleich eingeschlossen wurden.

Ergebnisse

Methodenbeschreibungen

Die Durchsicht der Literatur ergibt, dass allen Methoden für indirekte Vergleiche die gleiche Annahme zugrunde liegt: Die Variabilität zwischen den Ergebnissen der einzuschließenden Studien ist ausschließlich zufällig bedingt, oder – in anderen Worten – es besteht keine bedeutsame Heterogenität zwischen den Studien.

Es werden vier häufig eingesetzte Methoden für die Durchführung indirekter Vergleiche identifiziert, die alle mit metaanalytischen Ansätzen arbeiten:

1. Nicht-adjustierter indirekter Vergleich: Für den Vergleich einer Therapie A mit einer Therapie B werden die Ergebnisse aller mit A behandelten Studienarme in einer Metaanalyse zu einem Gesamteffektschätzer θ_A zusammengefasst. Das gleiche wird in einer zweiten Metaanalyse für alle mit B behandelten Studienarme durchgeführt um θ_B zu erhalten. Dieses Vorgehen wird als „Nicht-adjustierter indirekter Vergleich“ bezeichnet, was eine Abkürzung der Bezeichnung „Nicht bezüglich der Ergebnisse aus den Kontrollgruppen adjustierter indirekter Vergleich“ darstellt. Für den Vergleich der zusammengefassten Gesamteffektschätzer θ_A und θ_B gibt es im Anschluss vier Möglichkeiten: Die Bildung eines Gesamteffektschätzers $\theta_{A \text{ versus } \theta_B}$; die Prüfung der Differenz zwischen θ_A und θ_B auf statistische Signifikanz mithilfe eines statistischen Tests; die Prüfung der Konfidenzintervalle um θ_A und θ_B auf Überlappung oder ein narratives Gegenüberstellen der Therapieeffekte von A und B.
2. Für den (hinsichtlich der Ereignisse in den Komparatorarmen) adjustierten indirekten Vergleich werden zunächst mithilfe konventioneller Metaanalysemethoden die Gesamteffektschätzer $\theta_{A \text{ versus Komparator}}$ und $\theta_{B \text{ versus Komparator}}$ gebildet. Zum Vergleich der beiden Gesamteffektschätzer sind die unter 1. skizzierten vier Methoden einsetzbar.
3. Bei der Metaregression werden zur Bildung der Gesamteffektschätzer θ_A und θ_B die Einzelstudienresultate "A versus Komparator" und "B versus Komparator" in zwei Regressionsgleichungen eingesetzt. Die Regressionsmodelle können darüber hinaus die Effekte von Kovariaten (die als Auslöser von Heterogenität gelten – wie z. B. Alter der Studienpopulationen oder Krankheitsschweregrade) bei der Bildung der Gesamteffektschätzer berücksichtigen. Der Vergleich von θ_A und θ_B erfolgt wieder mithilfe der vier genannten Methoden.
4. Mixed treatment comparison (MTC) steht als Sammelbegriff für indirekt vergleichende Methoden, die mehr als zwei Therapieoptionen simultan miteinander vergleichen und eventuell vorhandene Head-to-head-Studien mit aufnehmen können. Sie kann unbegrenzt viele Therapieoptionen in eine Wirksamkeitsrangfolge bringen. Dabei folgt sie mithilfe der Bayes'schen Statistik einem iterativem Konzept, das sukzessive die gesamte verfügbare RCT-Evidenz zu dem/n Therapievergleich/en zusammenfasst.

Indirekte Vergleiche ohne Metaanalyse werden durchgeführt, wenn nur je eine Studie mit Informationen zu Therapie A und eine zu Therapie B zur Verfügung steht bzw. wenn die vorhandenen Studien hochgradig heterogen sind. Auch die dann durchgeführten deskriptiven Vergleiche verfolgen entweder das Prinzip des nicht-adjustierten oder des adjustierten Vergleichs.

Relative Häufigkeit des Einsatzes der verschiedenen Methoden für indirekte Vergleiche

In dem ungefähr neunjährigen Zeitraum von Januar 1999 bis Februar 2008 werden in den gefundenen systematischen Übersichtsarbeiten 108-mal metaanalytische Methoden für indirekte Vergleiche angewendet. Ein zeitlicher Trend mit einer kontinuierlich ansteigenden Zahl der Publikationen mit indirekten Vergleichen ist erkennbar.

Die mit Abstand am häufigsten verwendete Methode ist der adjustierte indirekte Vergleich mit 60 Einsätzen. Es folgen die Metaregressionen (17 Einsätze), die nicht-adjustierten indirekten Vergleiche (14 Einsätze), der MTC (zwölf Einsätze) und die sonstigen Methoden, die sich nicht in die vier wichtigsten Methodengruppen einordnen lassen (fünf Einsätze). 2006 ist ein sprunghafter Anstieg bei der Häufigkeit der durchgeführten MTC zu verzeichnen (zehn Beispiele von 2006 bis 2007).

Validitätsprüfung

Insgesamt werden 248 Gegenüberstellungen von Metaanalysen mit direktem und indirektem Vergleich (Datensätze) aus 57 systematischen Reviews für die Validitätsprüfungen herangezogen.

Bei der Testung auf systematische Über- oder Unterschätzung stellt sich heraus, dass keine der untersuchten indirekt vergleichenden Methoden systematisch die Ergebnisse des direkten Vergleichs über- oder unterschätzt.

Dennoch sind Diskrepanzen zwischen den Ergebnissen direkter und indirekter Vergleiche nachweisbar, die mit Abstand am stärksten bei den nicht-adjustierten indirekten Vergleichen ausgeprägt sind (Betrag des Mittelwerts der z-Werte = 1,63 [95%-KI: 1,20; 2,07]). Die durchschnittlichen Diskrepanzen der adjustierten indirekten Vergleiche (Betrag des Mittelwerts der z-Werte = 0,95 [95%-KI: 0,80; 1,09]), der Metaregressionen (Betrag des Mittelwerts der z-Werte = 0,99 [95%-KI: 0,20; 1,79]) und der MTC (n=57; Betrag des Mittelwerts der z-Werte = 0,59 [95%-KI: 0,45; 0,73]) liegen auf einem vergleichbaren Niveau. Der niedrige mittlere Betrag der MTC nivelliert sich in der Subgruppenanalyse, in der keine Head-to-head-Studien mit in die MTC eingeschlossen sein durften (n = 12; Betrag des Mittelwerts der z-Werte = 0,83 [95%-KI: 0,40; 1,26]). Bei den anderen Methoden sind die Ergebnisse aus Haupt- und Subgruppenanalyse konkordant.

Es muss allerdings berücksichtigt werden, dass die Streuung um die Mittelwerte der durchschnittlichen Diskrepanz bei den verschiedenen Methoden stark variiert. Es werden unterschiedlich viele extreme Werte mit statistisch signifikanter Diskrepanz (Betrag des Mittelwerts der z-Werte > 1,96; entspricht p=0,05) gefunden: Beim nicht-adjustierten indirekten Vergleich beträgt der Anteil statistisch signifikant diskrepanter Datensätze an allen ausgewerteten Datensätzen 25,5% (n=47; 95%-KI: 13,1%; 38,0%), beim adjustierten indirekten Vergleich 12,1% (n=116; 95%-KI: 6,1%; 18,0%), bei der Metaregression

16,7% (n=6; 95%-KI: -13,2%; 46,5%) und beim MTC 1,8% (n=57; 95%-KI: 2,1%; 34,3%). Die Ergebnisse der Hauptstudien sind mit denen der Subgruppenanalyse konkordant.

Zusammengefasst weisen 32 der 248 Gegenüberstellungen von Ergebnissen aus direktem und indirektem Vergleich statistisch signifikante Diskrepanz auf (12,9% [95%-KI: 8,7%; 17,1%]).

Für 15 der 32 diskrepanten Datensätze lagen keine Angaben zur Heterogenität des zugrunde liegenden Studienpools vor. Bei elf der diskrepanten Datensätze war unter den eingeschlossenen Studien signifikante Heterogenität nachweisbar und bei sechs der diskrepanten Datensätze nicht.

Kongruenz in den Schlussfolgerungen

In knapp der Hälfte der 248 Gegenüberstellungen liegt sowohl im direkten als auch im indirekten Vergleich kein signifikanter Therapieeffektunterschied vor (49,2%; 95%-KI: 43,0%; 55,4%). Am zweithäufigsten wird sowohl vom indirekten als auch vom direkten Vergleich die gleiche Therapieoption als signifikant überlegen erkannt (21,8%; 95%-KI: 16,6%; 26,9%). In den restlichen 29% (95%-KI: 23,4%; 34,7%) der Fälle erzielen der direkte und der indirekte Vergleich keine Übereinstimmung in ihren Schlussfolgerungen. Der gefürchtete Fall, dass direkter und indirekter Vergleich jeweils das andere Therapieverfahren als signifikant überlegen erachten, kann allerdings nur in fünf Fällen (Anteil 2%; 95%-KI: 0,3%; 3,8%) beobachtet werden.

Präzision indirekt vergleichender Methoden

In der untersuchten Stichprobe (n=248) besitzt der Effektschätzer des indirekten Vergleichs eine im Median um 9% geringere Konfidenzintervallweite als der direkte Vergleich (25. Perzentil: -34%; 50. Perzentil (Median): -9%; 75. Perzentil: 30%). Gleichzeitig beinhalten diese indirekten Vergleiche im Median exakt sechsmal so viele Studien wie die direkten Vergleiche (25. Perzentil: 4; 75. Perzentil: 13). In der untersuchten Stichprobe ergibt sich somit, bei einem Verhältnis an eingeschlossenen Studien von sechs zu eins zwischen indirektem Vergleich und direktem Vergleich, ein geringfügig engeres Konfidenzintervall beim indirekten Vergleich. Dies unterstützt die Aussage von Glenny et al., dass ein indirekter Vergleich viermal mehr ungefähr gleich große Studien einschließen müsse als ein direkter, um die gleiche Präzision zu erhalten.

Diskussion

Bei der Entscheidung ob, und wenn ja welche, Methode des indirekten Vergleichs eingesetzt werden kann, sollten vier Kriterien berücksichtigt werden:

1. Validität der Methode

Die geringste Validität im Vergleich zu direkt vergleichenden Methoden weist der nicht-adjustierte indirekte Vergleich auf. Da die Ergebnisse der Kontrollgruppen nicht mit in der Analyse berücksichtigt werden, wird von einem Bruch der Randomisation gesprochen. In der Analyse besteht ein großes Risiko, dass Biasformen, die für unkontrollierte Studien typisch sind (u. a. Selektionsbias und Confounding), die Resultate verzerren.

Der adjustierte indirekte Vergleich, die Metaregression und der MTC berücksichtigen dagegen die Ergebnisse in den Kontrollgruppen der Einzelstudien und bewahren somit die erzielte Strukturgleichheit in den Einzelstudien. Ein Selektionsbias auf der Metaebene kann jedoch auch bei diesen Methoden auftreten, wenn die Studien unterschiedliche Einschlusskriterien aufweisen. Die resultierenden ungleich verteilten Patientencharakteristika können, wenn sie mit dem Therapieeffekt assoziiert sind, zu Confoundern werden. Confounder täuschen einen Therapieeffekt vor, der nicht ursächlich durch die Therapie, sondern durch den Confounder verursacht wird.

Aufgrund dieser Verzerrungsrisiken sollten die genannten indirekt vergleichenden Methoden nicht bei heterogener Studienlage eingesetzt werden. Diese, alle indirekt vergleichenden Methoden betreffende Voraussetzung, gilt auch für konventionelle Metaanalysen und stellt im Prinzip keine Besonderheit der indirekten Vergleiche dar. Die theoretischen Betrachtungen werden durch die empirische Validitätsprüfung der indirekt vergleichenden Methoden gestützt. Ausreichend Datensätze zur Untermauerung der Hypothese, dass – eine homogene Studienlage vorausgesetzt – adjustierte indirekte Vergleiche die gleiche Ergebnisvalidität wie direkte Vergleiche aufweisen können, liegen vor. Ebenso können MTC, wenn diese Head-to-head-Studien mit den Therapieoptionen von Interesse einschließen, eine hohe Ergebnisvalidität bescheinigen werden.

Die Validität von Metaregressionen, MTC ohne Einschluss von Head-to-head-Studien und der unter dem Punkt „Sonstige Methoden“ gelisteten Techniken, kann aufgrund mangelnder Daten nicht bewertet werden.

2. Anzahl der zu vergleichenden Therapien

Sollen nur zwei Therapieverfahren indirekt miteinander verglichen werden, bietet sich unter Berücksichtigung der Ergebnisvalidität und des methodischen Aufwandes vor allem der Einsatz eines adjustierten indirekten Vergleichs an. Sollen jedoch mehr als zwei Therapien einander gegenüber gestellt werden, ist nur der MTC einsetzbar, um diese in eine Rangfolge hinsichtlich ihrer Wirksamkeit zu bringen.

3. Einschluss von Head-to-head-Studien

Neben den MTC liefern auch die anderen drei indirekt vergleichenden Methoden Ansätze für den Einschluss von Head-to-head-Studien in den indirekten Vergleich. Es

liegen allerdings keine ausreichenden Daten für eine Validitätsprüfung vor. Für die bisher publizierten Beispiele gilt, dass der MTC, der Head-to-head-Studien mit einschließt, zu den gleichen Ergebnissen kommen kann, wie die Head-to-head-Studien allein. Sein zusätzlicher Vorteil liegt darin, dass durch Hinzunahme des indirekten Vergleichs die Präzision des Gesamteffektschätzers erhöht werden kann.

4. Heterogene Studienlage

Der indirekte Vergleich mittels Metaregression stellt noch keine ausreichend validierte Methode dar, um mit ihr verlässlich für Heterogenitätsfaktoren adjustieren zu können. Auch die Adjustierung für Kovariaten beziehungsweise die Berücksichtigung von Heterogenität mit einem Inkonsistenzfaktor in MTC wurden noch nicht ausreichend validiert.

Liegt klinisch bedeutsame Heterogenität vor, kann derzeit kein indirekt vergleichendes Verfahren empfohlen werden, da das Risiko für systematische Fehler hoch ist. Bei geringer Heterogenität, kann über ein Random effects-Modell eine konservative Schätzung vorgenommen werden. Fixed effects-Modelle sollten nur bei weitgehend homogenen Studiensätzen eingesetzt werden. Für alle in diesem HTA beschriebenen Methoden für indirekte Vergleiche stehen sowohl Fixed als auch Random effects-Modelle zur Verfügung.

Schlussfolgerungen

Es bleibt festzustellen, dass für indirekte Vergleiche eine Reihe von Methoden zur Verfügung stehen, die sich hinsichtlich ihrer Möglichkeiten, die Evidenz verschiedener Studienlagen zusammenzufassen, unterscheiden.

In der empirischen Überprüfung zeigt sich, dass vor allem die Ergebnisse nicht-adjustierter indirekter Vergleiche häufig von denen direkter Therapievergleiche abweichen. Die übrigen indirekt vergleichenden Methoden können insbesondere dann konkordante Ergebnisse liefern, wenn die zusammengefassten Studien wenig Heterogenität aufweisen. Aus diesem Grund sollten adjustierte indirekte Vergleiche, Metaregressionen und MTC nur bei einer homogenen Studienlage eingesetzt werden. Zudem muss ein indirekter Vergleich, damit er einen ähnlich präzisen Effektschätzer wie ein direkter Vergleich liefern kann, ca. sechsmal so viele ungefähr gleich große Studien wie der direkte Vergleich enthalten.

Für den Kontext von HTA und Leitlinienerstellung stellen indirekte Vergleiche dann ein wertvolles Hilfsinstrument dar, wenn direkte Evidenz für einen Wirksamkeitsvergleich von Therapieverfahren nicht zur Verfügung steht.

Da eine homogene Studienlage in der Praxis allerdings nur selten vorliegt, bleibt vor einer breiteren Verwendung von indirekten Therapievergleichen zu definieren, bei welchem Ausmaß an Heterogenität (und Inkonsistenz) es aus Validitätsgründen noch vertretbar ist, sie zur Beantwortung von Forschungsfragen heranzuziehen.

Korrespondenzadresse:

Dr. med. Dagmar Lühmann
Institut für Sozialmedizin, Universität zu Lübeck,
Beckergrube 43-47, 23552 Lübeck, Tel.: 0451/7992538
Dagmar.Luehmann@uk-sh.de

Bitte zitieren als

Schöttker B, Lühmann D, Boulkhemair D, Raspe H. Indirekte Vergleiche von Therapieverfahren. *GMS Health Technol Assess.* 2009;5:Doc71.

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/hta/2009-5/hta000071.shtml>

Der vollständige HTA-Bericht steht zum kostenlosen Download zur Verfügung unter:

http://portal.dimdi.de/de/hta/hta_berichte/hta243_bericht_de.pdf

Copyright

©2009 Schöttker et al. Dieser Artikel ist ein Open Access-Artikel und steht unter den Creative Commons Lizenzbedingungen (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.de>). Er darf vervielfältigt, verbreitet und öffentlich zugänglich gemacht werden, vorausgesetzt dass Autor und Quelle genannt werden.