

## 4.0. Veranschaulichung der theoretischen Vorgehensweise:

Im folgenden soll das Prinzip der „Evidence-based medicine“ erläutert werden und exemplarisch aufzeigen, wie diese theoretischen Ansätze in einem chirurgischen Fach in die Praxis umgesetzt werden können. Die praktische Umsetzung im jeweils nachfolgenden Beispiel wird sich auf Therapiemöglichkeiten in der spinalen Chirurgie mittels operativer Verfahren beschränken. Es wird dabei ein Vergleich der zur Zeit angewandten Operationen angestrebt, damit abgewogen werden kann, bei welcher Indikationsstellung eine bestimmte Therapie als nützlich erscheint.

### 4.1. Formulierung der Ausgangsfrage

In der praktischen Anwendung sollte zunächst festgestellt werden, welche Thematik in der Behandlung „evidence-based“ analysiert werden sollte. Dies soll veranschaulichen, welcher Abschnitt der Behandlung Fragen aufwirft. Strittige Punkte können darauf hinweisen, daß eine große Behandlungsvarianz besteht, daß Uneinigkeit über die Behandlungsmethoden vorliegen oder daß eine Unsicherheit über Nutzen und Wirkung der verfügbaren Verfahren gegeben ist.

Allgemein kann sich die Ausgangsfrage auf folgende Bereiche ausdehnen<sup>55</sup>:

- **Befund:** Wie ist die Anamnese und die klinische Untersuchung zu interpretieren?
- **Ätiologie:** Gründe für die Entstehung der Krankheit.
- **Differentialdiagnose:** Einschätzung der Wahrscheinlichkeit, Zuverlässigkeit und Behandlungsmöglichkeit der potentiellen Erkrankungen.
- **Diagnostischer Test:** Auswahl und Interpretation eines Tests zum Aus-/Ein- Schluß einer Diagnose basierend auf Präzision, Akzeptanz des Verfahrens, Kosten und Sicherheit.
- **Prognose:** Abschätzung des klinischen Verlaufs, Voraussage der möglichen Komplikationen.
- **Therapie:** Auswahl der optimalen Behandlung unter Berücksichtigung des Kosten/Nutzen- Effekts.

- **Prävention:** Senkung der Erkrankungswahrscheinlichkeit durch Identifikation der Risikofaktoren und der möglichen Screeningtests.
- **Selbstverbesserung:** Wie können klinische Fähigkeiten verbessert werden und eine höhere Effektivität in der Behandlung erzielt werden?

Die Formulierung der Ausgangsfrage sollte das Problem des individuellen Patienten möglichst präzise und direkt ausdrücken, damit klare Antworten gefunden werden können.

Der Patient oder das Problem wird so charakterisiert, daß eine Zuteilung zu einer bestimmten Behandlungsgruppe möglich ist. Damit wird die Thematik präzisiert und eingekreist. Danach erfolgt die Festlegung für die vorgesehene Intervention und eine Festlegung der wichtigen Alternativmethoden. Hierdurch wird die Art der Therapie, die diagnostische Methode oder ein prognostischer Faktor klar definiert und gleichzeitig mit einer anderen Methode verglichen. Letztendlich sollten auch die zu erwartenden Ziele mit den jeweiligen Vor- und Nachteilen festgehalten werden <sup>55</sup>.

Unter Berücksichtigung der Vorgeschichte des Patienten, der Begleiterkrankungen, des sozialen Umfeldes und der schon erfolgten Therapien und inzwischen auch Nutzen und Kosten (Belastung der Solidargemeinschaft), kann nun die praktische Umsetzung erfolgen.

Da sich in der Behandlung eines Patienten häufig mehrere Fragen über die Art und Weise der Behandlung ergeben, soll an dieser Stelle darauf hingewiesen werden, daß aufgrund Zeitmangels Fragen ausgerichtet werden können und sich dann lediglich auf eine Thematik beschränken, z.B. auf die wichtigste Frage für das Wohlergehen eines Patienten, oder welche Fragen sind überhaupt in der zur Verfügung stehenden Zeit zu bearbeiten?

**Praktisches Beispiel:****Anamnese und Klinik:**

Ein 45-jähriger Patient mit leichtgradiger Fuß- und Großzehheberparese sowie einer Hypästhesie im Dermatome L5 rechts, hat starke Schmerzen bei Belastung, zusätzlich Dysästhesien (seit 3 Wochen bestehend). Nervendehnungsschmerz (Laseque'sches Zeichen) rechts bei 35° auslösbar, an der gesamten Lendenwirbelsäule paravertebrale Myogelosen mit schmerzbedingter Bewegungseinschränkung in allen Ebenen. Der Patient hat keine weiteren Begleiterkrankungen.

**Diagnose:**

Nervenzwurzelkompression, von L5 rechts bei Bandscheibenvorfall L4/5 rechts.

**Therapie:**

Nach 6-wöchiger konservativer Therapie ergab sich keine Besserung.

Aufgrund der erfolglosen konservativen Therapie besteht die Indikation zu einem operativen Vorgehen.

Wahlweise stehen drei Operationsverfahren zur Verfügung:

1. Intradiskale Operationsverfahren
2. Mikrochirurgisches Vorgehen
3. Transforaminale endoskopische Sequesterentfernung

**Therapieziel:**

Langfristige Beschwerdefreiheit des Patienten.

**Fragestellung:**

Welches der möglichen Operationsverfahren ist für meinen Patienten am besten geeignet?

## 4.2. Auswahl thematisch relevanter Studien, Literatur und Veröffentlichungen.

Der erste Schritt zur Literatursuche sollte in der schriftlichen Festlegung der Ein- und Ausschlusskriterien für Studien bestehen. Damit wird das Prozedere reproduzierbar und stellt außerdem eine Entscheidungshilfe für den Bearbeitenden dar, der sich bei der Recherche an einer klaren Definition orientieren kann.

Die sich nun anschließende Suche nach dem entsprechenden „external evidence“\* beschränkt sich im allgemeinen auf wenige Medien. Im einfachsten Fall stammen die Informationen von einem Fachkollegen, sie können Lehrbüchern und Zeitschriften entnommen werden oder aus Fortbildungsveranstaltungen resultieren. Das modernste Medium ist zur Zeit das Internet mit einem einladend großen Angebot. Dem kritischen Leser aber zeigen sich jedoch gelegentlich Mängel in den verfügbaren Informationen.

Die gewonnenen Informationen sind teilweise veraltet (insbesondere Lehrbücher), das Material wurde selektiv ausgewählt entsprechend der publizierten Thematik, Studien sind teilweise nicht transparent und damit nicht nachvollziehbar. Einige Aussagen sind nicht umfassend genug und unterliegen nicht vollständig zu vermeidenden Bias. Nach dem Selektionsprinzip der „Evidence-based medicine [...] entfällt die übergroße Mehrheit aller Publikationen als irreführend oder bestenfalls wertlos aus.“<sup>13</sup>

Leider kann eine Verzerrung des Gesamtbildes auch durch den Bearbeitenden entstehen, insbesondere einem Experten auf einem Spezialgebiet, da dieser mit einer vorgefaßten Meinung die Thematik bearbeitet und so die Informationssuche einseitig gestaltet. Bei wichtigen schon lokalisierten Studien sollte nicht versäumt werden, die angegebenen Referenzen nach weiteren interessanten Informationen zu durchsuchen.

Vorteilhaft hat sich die Verwendung von randomisierten kontrollierten Studien erwiesen, die bei einer ausreichenden Anzahl von Teilnehmern und guten Ein- und Ausschlusskriterien verlässliche Daten produzieren. Ersichtlich wird dies in der Tabelle 1 (nachfolgende Seite), in der die Studientypen in Hinblick auf die Beweiskraft hierarchisch geordnet dargestellt sind.

\*„external evidence“: klinisch anwendbare wissenschaftliche Ergebnisse in Bezug auf die Behandlung.

Eine Anleitung zur systematischen, manuellen Suche nach Artikeln über kontrollierte klinische Prüfungen wird vom Deutschen Cochrane Zentrum angeboten unter:

<http://cochrane.de/deutsch/cclitfr1.htm><sup>77</sup>.

In der nachfolgenden Tabelle 1 werden die Studien hinsichtlich ihrer Qualität und Verwendbarkeit für die „Evidence-based medicine“ hierarchisch geordnet.

„Type and Strength of Evidence“<sup>75</sup>, Hierarchie der wissenschaftlichen Evidenzen.

**Tabelle 1**

|     |  |
|-----|--|
| I   | Mindestens ein systematischer Review auf der Basis methodisch hochwertiger randomisierter kontrollierter Studien.          |
| II  | Mindestens eine ausreichend große, methodisch hochwertige randomisierte kontrollierte Studie.                              |
| III | Methodisch hochwertige Studien ohne Randomisierung bzw. nicht prospektiv. (Kohorten-, Fall-Kontroll-Studien)               |
| IV  | Mehr als eine methodisch hochwertige nicht experimentelle Studie.  |
| V   | Meinungen von Autoritäten, die auf klinischen Beweisen basieren, Berichte von Expertenkommissionen, beschreibende Studien. |
| VI  | „Someone once told me“. ...habe ich mal gehört.  |

In der Tabelle ist die Stärke der Evidenz von I-VI abfallend geordnet. Die Stärke des Beweises hängt von dem Grad ab, mit dem Bias oder „confounding“-Faktoren (siehe Glossar) kontrolliert werden können. Definitionsgemäß bedeutet das, daß großangelegte Studien den stärksten Beweis erbringen, weil sie Bias kontrollieren, vor allem, wenn diese eine ausreichend große Teilnehmerzahl aufweisen.

Wenn keine Übersichtsarbeiten (Reviews) vorhanden sind (Evidenzgrad I), ist zu beachten, daß elementare Hauptstudien berücksichtigt werden, da diese häufig die einzige Quelle von wissenschaftlichen Beweisen darstellen können.

### **Praktisches Beispiel:**

#### **Einschluß- und Ausschlußkriterien für Studien und Literatur bei degenerativer lumbaler Nervenwurzelkompression.**

- Veröffentlichungszeitraum sollte die letzten 5 Jahre umfassen, bei ergebnisloser Suche kann ein Zeitraum von 10 Jahren angenommen werden.
- Die Thematik der Studien sollte sich auf Operationsverfahren der degenerativen lumbalen Nervenwurzelkompressionen beschränken und möglichst die Effektivität der möglichen Verfahren vergleichend darstellen.
- Evidenzgrade I-III werden eingeschlossen.

#### **4.2.1. Hinweise zu Informations- und Bezugsquellen der wissenschaftlichen Evidenz:**

Zu beachten ist bei der Suche im Internet nach dem „external evidence“, daß in Anbetracht der weitläufigen Zugänglichkeit, Studien, Artikel oder aber auch Reviews mit minderer Qualität vorkommen<sup>27</sup>. Einer Qualitätssicherung in diesem Sinne unterliegen Organisationen wie z.B. das Cochrane Collaboratorium. Nachfolgend werden einige Bezugsquellen aufgeführt.

- Cochrane Collaboration: Informationen über das Cochrane Collaboratorium und „Evidence-based medicine“ im Internet: <http://www.imbi.uni-freiburg.de/cochrane/> Das Ziel der Cochrane Collaboration ist es, systematische Übersichtsarbeiten zu erstellen, die insbesondere randomisierte kontrollierte Studien umfassen, welche sich fast ausschließlich auf das Gesundheitswesen beschränken<sup>44</sup>.
- Bezug der Cochrane Library: Update Software, Summertown Pavilion, Middle Way, Summertown, Oxford OX2 7LG, UK.
- University of York, NHS Centre for Reviews and Dissemination (CRD), im Internet: <http://www.york.ac.uk/inst/crd/>
- Deutsches Institut für Medizinische Dokumentation und Information im Internet (insbesondere für Medikamente): <http://www.dimdi.de>
- National Library of Medicine im Internet: <http://www4.ncbi.nlm.nih.gov/PubMed/clinical.html>
- Current Controlled Trials: [www.controlled-trials.com](http://www.controlled-trials.com)

- Bezug der Zeitschrift „Evidence-based medicine“; Englisch: BMJ Publishing Group, BMA House, Tavistock Square, London WC1H 9JR, UK. Deutsch: W. Zuckschwerdt Verlag, Industriestraße 17, 82110 München.
- ACP Journal Club im Internet: <http://www.acponline.org/journals/acpj/jcmenu.htm>
- Weitere Hinweise über Verknüpfungen mit „Evidence-based medicine“ sind im Internet verfügbar unter „Netting the Internet.“ A SchARR Inroduction to Evidence Based Practice on the Internet (University of Sheffield). <http://www.shef.ac.uk/~scharr/ir/netting.html>.

### **Zur Suchstrategie:**

Bei der Anwendung von „Suchmaschinen“ sollten die Vorgehensweisen im Umgang von Verknüpfungswörtern und deren Terminologie bekannt sein, um möglichst effektiv zu suchen. Bei Unkenntnis kann dies aber schnell erlernt werden.

Beispielsweise:

Verknüpfung zweier Wörter mit „AND“ verkleinern das Suchfeld, mit „OR“ hingegen wird es vergrößert. Die Terminologie zur Suche von Studien über Behandlungen kann die Begriffe „clinical trial; randomized controlled trial; therapeutic use,“ etc. umfassen. Weitere nützliche Hinweise können bei der Einweisung zur Verwendung der Internetdatenbanken erfragt werden.

### **Praktisches Beispiel:**

Suche im Internet: „Thesaurus“ medizinischer Index gibt Aufschlüsselung über Synonyma, die in „Grateful Med.“ unter „Medical Subject Headings“ enthalten sind.

### **Studientyp (Publication Types):**

*randomized controlled trial*

### **Suchbegriffe (Query Terms):**

*percutaneous discectomy or chemonucleolysis or microsurgery and therapy*

**Languages:** *English*

**Study Groups:** *Human*

**Age Groups:** *All Adult (19+years)*

**Year range, begin year:** *1990 through*

**Gender:** *all*

**Journals:** *all*

**End year:** *1999*

### 4.3. Kritische Beurteilung der Studien

Zu falschen Schlußfolgerungen führen unweigerlich Verzerrungen des Behandlungseffekts (Bias), daher sollte jede einzelne Studie auf diesen Aspekt hin überprüft werden. Sobald eine systematische Verzerrung ausgeschlossen werden kann, sollte der Behandlungseffekt nach Größe und Präzision beurteilt werden. Zu beachten ist auch die Anwendbarkeit auf den individuellen Patienten, unter Berücksichtigung des Risikos/Nutzen, der Konsequenz der Nichtbehandlung und inwiefern sich die Ergebnisse der gefundenen Studien mit den Problemen des Patienten überschneiden<sup>4</sup>.

Die Validität einer Studie wird durch Evaluation des Studientyps und die Überprüfung der Daten auf Vollständigkeit festgestellt.

Schließlich kann die Effektgröße durch die RRR (relative Risikoreduktion), der ARR (absolute Risikoreduktion), der Odds-ratio (Chancenverhältnis) oder der NNT (number needed to treat) bestimmt werden<sup>4</sup>.

Fragen, die zur Beurteilung der Studien herangezogen werden sollten, sind nachfolgend in Primär- und Sekundärfragen gegliedert. Können die Primärfragen positiv beantwortet werden, handelt es sich um eine valide Studie. Die Sekundärfragen schlüsseln die Qualität der Studie nochmals genauer auf, d.h. je nach Zeitinvestition kann eine erweiterte oder verkürzte Bewertung erfolgen.

#### **Primärfragen<sup>32</sup>:**

- **War die Zuordnung der Patienten zu den Behandlungsgruppen zufällig (randomisiert) und verdeckt?**
- **Wurden die Daten aller Patienten in den Schlußfolgerungen der Studie berücksichtigt?**
- **War das Follow-up komplett? (Wurden Angaben über fehlende Daten gemacht?)**
- **Wurden die Patienten in ihrer, durch Randomisierung festgelegten, Gruppe analysiert (Intention-to-treat Prinzip)?**



Die Überprüfung der Validität einer Studie erfolgt in erster Linie durch Bewertung der angewandten Methodik. Randomisierte Studien haben den Vorteil, daß diese weniger willkürlichen Verzerrungen unterliegen, da die Studienteilnehmer zufällig ihrer Behandlungsgruppe zugeteilt werden. Studien, welche keinem randomisierten Verfahren unterliegen, haben häufig zu fatalen Trugschlüssen geführt. Das bekannteste Beispiel ist in der Neurochirurgie die intrakranielle-extrakranielle Bypass Operation. Das Ergebnis der Studie wird in dem Kapitel „Potentieller Nutzen der „Evidence-based medicine“ in der Neurochirurgie“ ausführlich dargestellt.

Mit der Vollständigkeit der Daten ist gemeint, daß eine schlüssige Begründung für fehlende Daten und ausgeschlossene Patienten angegeben werden sollte. Patienten, die aus einer Studie ausscheiden, können dafür verschiedene Gründe haben.

Im negativsten Fall könnte es das Versterben derselben sein oder im positiven Fall die völlige Genesung. Um diese Begebenheiten beurteilen zu können, ist ein komplettes Follow-up und eine gute Dokumentation dringend notwendig.

Die Studienteilnehmer sollten in den zugeteilten Gruppen analysiert werden, in die sie anfänglich randomisiert wurden, auch bei nicht Compliance oder nicht erfolgter Teilnahme am eigentlichen Experiment (an einer Operation). Dies wird als „Intention- to- treat“ bezeichnet. Beispielsweise könnte ein Patient aus einer chirurgischen Studie ausgeschlossen werden, wenn aufgrund seiner Verfassung eine Teilnahme an einer Operation nicht zumutbar ist. Würde dieser dann der Kontrollgruppe zugeordnet, um ihn nicht komplett auszuschließen, erscheint die Operation insgesamt als effektiv. In diesem Fall würden nur „kerngesunde“ zu der Experimentalgruppe (Operationsgruppe) hinzugezogen werden und der Kontrollgruppe auch die Patienten, welche an einer unabhängigen Begleiterkrankung leiden. Wenn jede randomisierte Person aber in ihrer Gruppe verbleibt, sind bekannte und unbekannte prognostische Faktoren ausbalanciert.

### **Sekundärfragen<sup>32</sup>:**

- **Handelt es sich um eine Doppelblindstudie?**
- **Waren die Behandlungsgruppen in der Studie vergleichbar?**
- **Wurden die Gruppen abgesehen vom experimentellen Teil gleich behandelt?**

Eine weitere Beurteilung der Validität erfolgt durch genaueres Betrachten der Vorgehensweisen in der Studie. Eine Doppelblindstudie ermöglicht es vorgefaßte Meinungen von Probanden, Klinikern und dem Personal nicht in die Studie mit einfließen zu lassen. Studien, die nicht blind durchgeführt werden konnten, sollten daraufhin untersucht werden, ob ein unabhängiger ‚verblindeter‘ Beobachter die Therapieergebnisse ausgewertet hat.

Die Vergleichbarkeit der Behandlungsgruppen für alle Faktoren, die das Therapieergebnis beeinflussen könnten, sichert, daß beim Start der Studie jeder potentiell die gleiche Therapie erhalten könnte oder aber zumindest gleiche Chancen auf Besserung hat (z.B.: Wurden Patienten voroperiert, am gleichen Defekt?). Allerdings gibt es statistische Methoden, Gruppen anzugleichen, wenn diese von Anfang an nicht gleichzusetzen waren. Diese Methode ist nur unter Vorbehalt anzuwenden und bedarf einer Beratung durch einen Biostatistiker, daher wird an dieser Stelle verzichtet, näher darauf einzugehen.

Die Behandlungsgleichheit, abgesehen vom experimentellen Teil, ist insofern wichtig, daß Verzerrungen der Studie vermieden werden können, wenn darauf geachtet wird, daß z.B. bei der Behandlungsgruppe kein genaueres Follow-up als in der Kontrollgruppe erfolgt (z.B. gleiche Krankheitspflege durch das Personal und gleiche therapiebegleitende Behandlung).

### **Praktisches Beispiel:**

#### **Kritische Beurteilung einer Studie aus „Spine“ Vol.20, 1995, Nummer 6, Seiten 734-738.**

20

Autoren: Sandip Chatterjee, FRCS, Patrick M. Foy, FRCS, and Gordon F. Findlay, FRCS.

„Report of a Controlled Clinical Trial Comparing Automated Percutaneous Lumbar Discectomy and Microdiscectomy in the Treatment of Contained Lumbar Disc Herniation“

#### **Zusammenfassung der wichtigsten Daten:**

|                       |   |
|-----------------------|---|
| <b>Studientyp:</b>    | randomisierte kontrollierte Studie.   |
| <b>Studienumfang:</b> | ursprünglich 160 Personen. Abbruch der Studie aus ethischen Gründen bei 71 untersuchten und behandelten Personen, weil die Mikrodiskektomie der APLD weitgehend überlegen erschien. |
| <b>Studiendauer:</b>  | 3 Jahre.  |
| <b>Vorbehandlung:</b> | Minimum 6 Wochen konservative Therapie.   |

|                               |  |
|-------------------------------|--|
| <b>Hauptsymptom:</b>          | Wurzelkompressionsschmerz.   |
| <b>Diagnostik:</b>            | MRI, T1 axial und T2 sagittal.   |
| <b>Einschlußkriterien:</b>    | Bildgebende Verfahren, die eine Bandscheibenhernie in einer Höhe zeigen und weniger als 30% der Höhe des sagittal- kranialen Umfanges aufweisen.   |
| <b>Ausschlußkriterien:</b>    | Patienten, die als hauptsächliches Symptom Rückenschmerzen der Lendenwirbelsäule angaben. Außerdem wurden Patienten mit Bandscheibenextrusionen, Sequestern, subarticular oder foraminal Stenosen, sowie Bandscheibenhernien auf mehreren Ebenen ausgeschlossen. Falls nach dem bildgebenden Verfahren Zweifel über die Beschaffenheit des Prolaps bestanden, wurden diese auch ausgeschlossen.  |
| <b>Operationsverfahren:</b>   | Mikrodiskektomie und /oder perkutane Diskektomie unter Aspiration (APLD= automated percutaneous lumbar discectomy).  |
| <b>Follow-up:</b>             | Nach 3 Wochen, nach 2 Monaten und nach 6 Monaten wurden Nachuntersuchungen durchgeführt. Die Untersuchung erfolgte jeweils von einem unabhängigen, verblindeten „observer“ und dem Operateur.  |
| <b>Statistische Methoden:</b> | Anwendung des $\chi^2$ -Test und der Yate´s correlation (siehe Glossar).   |
| <b>Interventionen:</b>        | Zu dem Zeitpunkt, an dem der „observer“ einen Mißerfolg feststellte (1.Follow-up nach 3 Wochen), wurde zum Ausschluß morphologischer Veränderungen nochmals ein MRI durchgeführt. Bei einer mißlungenen APLD wurde dem Patienten eine Mikrodiskektomie angeboten, der er sich, nach Einwilligung, 6 Wochen (Spannweite: 4-7 Wochen) später unterziehen konnte. 20 von 22 Patienten willigten ein.  |
| <b>Ergebnisevaluation:</b>    | Der „observer“ wurde verblindet, indem durch Bedeckung der Lendenwirbelsäule des Patienten die Operationsartefakte nicht sichtbar waren. Eine Zweitbeurteilung erfolgte durch den Operateur. Die Beurteilungskriterien wurden nach MacNab <sup>43</sup> (siehe Glossar) festgelegt. Die endgültige Ergebnisanalyse erfolgte nach einem Minimum von 6 Monaten (entsprechend dem Follow-up). Aufteilung der Ergebnisgruppen in 1) excellent and good und 2) fair and poor. |

### Zusätzliche Daten aus der Studie:

Nachfolgende Daten wurden der Originalstudie entnommen und wegen der Übersichtlichkeit in Form einer **Tabelle 2** dargestellt.

**Tabelle 2**

|  | APLD                    | Mikrodiskektomie        |
|--|-------------------------|-------------------------|
| Ebene L4-L5                              | 12 Personen             | 17 Personen             |
| Ebene L5-S1                              | 19 Personen             | 23 Personen             |
| Beschwerdedauer vor OP                   | Durchschnitt 18 Monate  | Durchschnitt 33 Monate  |
| Alter                                    | Durchschnitt 38,9 Jahre | Durchschnitt 41,3 Jahre |
| Radikuläre Schmerzen<br>(radicular pain) | Durchschnitt 13 Wochen  | Durchschnitt 20 Wochen  |
| Geschlecht                               | 49% männlich            | 60% männlich            |
| Teilnehmer                               | 31 Personen             | 40 Personen             |
| Erfolgreiches Ergebnis                   | 9 Personen (29%)        | 32 Personen (80%)       |
| Krankenhausaufenthalt                    | Ambulant                | 3,5-5,3 Tage            |

Verwendete Abkürzungen:

APLD = automated percutaneous lumbar discectomy

OP = Operation

APLD versus Mikrodiskektomie: Signifikanz  $\chi^2=16,56$   $p<0,001$ , d.h. es besteht ein bedeutender Behandlungsunterschied zwischen diesen zwei Methoden.

APLD versus Mikrodiskektomie mit Aufteilung in den Ebenen L4-L5 und L5-S1:

Signifikanz für L4-L5:  $\chi^2= 5,249$   $p< 0,05$

Signifikanz für L5-S1:  $\chi^2= 11,45$   $p< 0,001$

**Tabelle 3**

| Procedure                   | Excellent/Good | Fair/Bad    |
|-----------------------------|----------------|-------------|
| Micro.                      | 32/40 (80%)    | 8/40 (20%)  |
| APLD alone                  | 9/31 (29%)     | 22/31 (71%) |
| APLD⇒Micro.                 | 13/20 (65%)*   | 7/20 (35%)  |
| APLD alone +<br>APLD⇒Micro. | 22/31 (71%)    | 9/31 (29%)  |

Micro. = Microdiscectomy

APLD= automated percutaneous lumbar discectomy.

\*Zwanzig von zweiundzwanzig (20/22) Patienten mit Mißerfolg der APLD wurden auch der Mikrodiskektomie unterzogen.

Tabelle, ("Comparison of outcome") wurde aus der Originalstudie entnommen<sup>20</sup>.

Die eingangs erwähnten Primär- und Sekundärfragen werden nachfolgend auf die Studie angewendet und anhand der gegebenen Informationen beantwortet.

### **Praktisches Beispiel:**

#### **Beantwortung der Primärfragen:**

#### **War die Zuordnung der Patienten zufällig und gab es eine verdeckte Randomisierung?**

Dieser Frage kann zugestimmt werden, da es sich um eine randomisierte Studie handelt, bei der ein unabhängiger „observer“ die Personen in die Studie aufnahm, ohne zu wissen, welche der beiden Therapien sie erhalten würden.

#### **Wurden die Daten aller Patienten, die in die Studie aufgenommen wurden, in den Schlußfolgerungen berücksichtigt?**

Da diese Studie abgebrochen wurde, wurden die Daten der ursprünglich 160 Patienten nicht aufgeführt. Die Daten der verbliebenen und untersuchten 71 Personen werden aber vollständig in das Ergebnis einbezogen.

### **War das Follow-up komplett?**

Es sind keine Angaben über nicht erfolgte Nachuntersuchungen gemacht worden oder über Patienten, die keine Folgeuntersuchung erhalten konnten. In der Ergebnisanalyse wurden die Zwischenergebnisse von 3 Wochen und 2 Monate nach der Therapie leider nicht dargestellt. Diese Frage, soweit beurteilbar, kann nicht nachgeprüft werden.

### **Wurden die Patienten in ihrer durch Randomisierung festgelegten Gruppe analysiert?**

Patienten, die durch die Randomisierung der Mikrodiskektomie zugeteilt wurden, sind bis zum Ende der Studie in ihrer Gruppe beobachtet worden und dementsprechend beurteilbar. Patienten, die der APLD zugeteilt wurden, erhielten jedoch, je nach Ergebnis der Operation, eine weitere Behandlung. Bei negativem Ergebnis wurde dem Patienten die Mikrodiskektomie angeboten, welche von 20 (von insgesamt 22) Patienten angenommen wurde. Demnach ist das Verfahren der APLD mit der Mikrodiskektomie nur bis zu dem Zeitpunkt vergleichbar, wo die erste Nachuntersuchung erfolgte. Die erste Nachuntersuchung erfolgte 3 Wochen nach Operation, fehlgeschlagene APLDs wurden nach Einwilligung 4-7 Wochen später der Mikrodiskektomie unterzogen.

Da sich 90% (20 von 22) der Patienten mit mißlungener APLD einer weiteren Operation unterzogen, sind diese nicht mehr der Gruppe zuzuordnen, zu der sie ursprünglich randomisiert wurden, sondern einer Subgruppe (APLD + Micro.). Demnach besteht kein „Intention-to-treat“ (siehe Glossar).

Bei positiven Ergebnissen der APLD und der Mikrodiskektomie wurden die Patienten bis zum Ende der Studie in ihren Gruppen analysiert.

Die eingangs formulierte Frage sollte dementsprechend verneint werden, es wurden nicht alle Patienten in ihrer ursprünglich festgelegten Gruppe analysiert. Wenn der komplette Zeitraum im Follow-up betrachtet werden soll, ist ein Wechsel der Patienten zwischen den Gruppen bei Mißerfolg ungünstig, da kein Vergleich zwischen den Gruppen vollzogen werden kann.

Aus ethischen Gründen ist diese Handlung nachzuvollziehen, jedoch aus statistischer Sichtweise ist die Verkürzung der Studie nachteilig.

Da die letzte Frage verneint werden mußte (Beurteilung der Kontinuität in dem Randomisierungsvorgang), kann diese Studie nur als mäßig valide angesehen werden. Wünschenswert wäre auch bei Operationsmißerfolg eine Langzeitevaluation. Fraglich ist, ob dies ethisch vertretbar wäre.

## **Beantwortung der Sekundärfragen:**

### **Handelt es sich um eine Doppelblindstudie?**

Es handelt sich nicht um eine Doppelblindstudie, aber in Anbetracht der Operationsverfahren ist dies aus der Sicht des Patienten auf jeden Fall zu rechtfertigen. Es erfolgte jedoch bei allen Patienten eine Beurteilung des Operationserfolges durch einen unabhängigen „observer“, der die Patienten, zusätzlich „verblindet“ (durch Bedecken der Operationsartefakte), untersuchte.

### **Waren die Behandlungsgruppen in der Studie vergleichbar?**

In Anbetracht der vorhandenen Informationen sind die Gruppen vergleichbar, abgesehen von den Patienten, die sich beiden zu vergleichenden Operationen unterzogen. Es stehen aber keine Angaben über möglicherweise vorhergehende Operationen einzelner Patienten oder aber Begleiterkrankungen, die das Ergebnis beeinflussen könnten, zur Verfügung.

### **Wurden die Gruppen abgesehen vom experimentellen Teil gleich behandelt?**

Dies ist aus der Originalstudie nicht zu entnehmen, dementsprechend auch nicht zu beantworten.

Die vorliegende Studie gibt leider keinen Aufschluß über Interventionen oder Faktoren, die das Ergebnis zusätzlich beeinflussen können. Positiv zu bewerten ist die Beurteilung der Operationsergebnisse durch eine unabhängige, verblindete Person.

### **Zusammenfassende Beurteilung der Studie:**

Es handelt sich um eine mäßig valide Studie, die bis zu einem Zeitraum von 4-7 Wochen nach der Operation valide Informationen liefert, danach aber durch Aufhebung des Randomisierungsprozesses nicht weiter beurteilbar ist. Positiv ist das Randomisierungsverfahren an sich zu bewerten und die Verwendung von standardisierten Meßgrößen.

Des Weiteren fehlen nützliche Zusatzinformationen über die Ausgangslage des Patienten (klinische Symptome, neurologische Defizite), Behandlungsgleichheit und Begleiterkrankungen, sowie soziodemographische Daten.

#### **4.4. Dokumentation der recherchierten Ergebnisse.**

Die Studien, welche anhand der Ein- und Ausschlußkriterien in die engere Wahl gekommen sind, sollten mit Titel, Autor und der Quellenangabe aufgeführt werden.

Jede Studie sollte zusätzlich eine kritische Beurteilung erhalten, die mittels Primär- und Sekundärfragen herausgefunden wurde. Das gesichtete Material bedarf aufgrund der zu erwartende Menge einer logischen und strukturierten Verwaltung, so daß jeder Zeit ein einfacher Zugriff möglich ist, und Daten nicht abhanden kommen.

An dieser Stelle ist zu überlegen, ob eine elektronische Datenverwaltung eingesetzt werden soll, damit eine bequemere und übersichtlichere Handhabung der gewonnenen Informationen möglich ist. Außerdem können nützliche Artikel direkt aus dem Internet kopiert und leicht in die neue Datenbank integriert werden.

Studien, die in die engere Wahl gekommen sind, werden nachfolgend auf ihre klinische Relevanz hin überprüft.

#### **4.5. Interpretation der Ergebnisse.**

Im folgenden wird eine Studie auf ihre Relevanz und ihre Präzision untersucht und damit das Ergebnis auf Nützlichkeit interpretiert. Dazu werden zunächst die verwendeten statistischen Hilfsmittel erklärt und nachfolgend an einem Beispiel praktisch angewendet (Studie aus Kapitel 1.4.3.).

Klinische Studien beschränken sich häufig nur auf die Angabe der relativen Risikoreduktion und der statistischen Signifikanz für die Unterschiede zwischen zwei Gruppen. Eine höhere Aussagekraft hat aber die absolute Risikoreduktion, weil in ihr das Ausgangsrisiko für das Ereignis in der Studienpopulation mit einfließt.

Des weiteren ist der Kehrwert der absoluten Risikoreduktion als Zahl zu verwenden, die beschreibt, wie viele Patienten behandelt werden müssen, um einen positiven Effekt zu erreichen (number to be treated, NNT). Dieser Wert kann wiederum mit Werten aus anderen medizinischen Verfahren verglichen werden.

Ebenso kann das Ausmaß möglicher Nebenwirkungen oder Begleiterscheinungen berechnet werden (number needed to treat to produce one episode of harm, NNH). Dazu wird eine prozentuale Angabe über eine Begleiterscheinung benötigt, über deren Kehrwert dann die NNH



errechnet wird. Durch die Ermittlung der NNT und der NNH kann eine Abwägung erfolgen, mit der festgestellt werden kann, daß für jeden Therapieerfolg eine bestimmte Anzahl von Begleiterscheinungen auftreten. Im folgenden wird genauer auf die statistischen Möglichkeiten eingegangen.

#### **4.5.1. Wie groß war der Behandlungseffekt?**

Die Abschätzung des Interventionseffekts (Behandlungseffekts) wird offensichtlich durch die Darstellung der relativen Risikoreduktion, der Odds- Ratio, der „number needed to be treated“ und der absoluten Risikoreduktion<sup>34</sup>.

Die relative Risikoreduktion ist ein Maß, welches die Senkung der Rate an negativen Ereignissen in der experimentellen Gruppe mit dem der Kontrollgruppe vergleicht. Im folgenden abgekürzt mit RRR. Mit anderen Worten, ein Faktor, der die Mißerfolgswahrscheinlichkeit einer Methode 1 gegenüber einer Methode 2 beschreibt.

Die Odds- Ratio (auch Chancenverhältnis) beschreibt das Verhältnis zwischen einer Chance eines Patienten aus der Experimentalgruppe und einer Chance eines Patienten aus der Kontrollgruppe, einem Ereignis zu erleiden. Die Odds- Ratio wird im folgenden mit OR abgekürzt.

Die „number to be treated“ gibt die Zahl der Patienten an, die behandelt werden müßte, um ein negatives Ereignis (z.B. mißlungene Operation) zu verhindern. Im folgenden auch abgekürzt mit NNT.

Die absolute Risikoreduktion beschreibt die Risikorate unter der Standardmethode abzüglich der Risikorate unter der experimentellen Methode. Nachfolgend abgekürzt mit ARR.

Die RRR, ARR sowie die OR und die NNT müssen aus den vorhandenen Daten errechnet werden, wenn sie nicht angegeben sind. Die Herangehensweise zur Berechnung dieser Parameter wird nachfolgend geschildert.

Eine Zusammenfassung über die statistischen Termini, die hier verwendet werden, ist auch im Internet abrufbar unter: <http://www.ihs.ox.ac.uk/casp/statistic.html> (von CASP, siehe Glossar)<sup>15</sup>.

Übersichtshalber sollten die Daten einer vergleichenden Studie in eine Vierfeldertafel eingetragen werden, sobald dies möglich ist. Die Vierfeldertafel ist besonders leicht anwendbar, wenn es sich um die Betrachtung zweier zu vergleichender Behandlungsmethoden handelt.

**Tabelle 4:**

|               |   |   |          |
|---------------|---|---|----------|
| Methode       | 1 | 2 | $\Sigma$ |
| Erfolg (+)    | A | B |          |
| Mißerfolg (-) | C | D |          |
| $\Sigma$      |   |   | N=       |

Unter der entsprechenden Methode wird jeweils die Anzahl der Personen mit einem Merkmal eingetragen. Außerdem wird in jeweils eine Spalte die Summe der teilgenommenen Personen für Methode 1 und 2, sowie die Gesamtsumme aller teilnehmenden Personen eingetragen.

Mit einem signifikanten  $\chi^2$ -Wert macht es Sinn, die Unterschiede zwischen den Behandlungsmethoden statistisch detaillierter zu untersuchen. Mit der OR wird mittels eines Zahlenwertes die Größe des Unterschiedes verdeutlicht.

#### **Berechnung der relativen Risikoreduktion (RRR):**

Die Anzahl der negativen Ergebnisse in der Kontrollgruppe wird zunächst dividiert durch die gesamte Anzahl der Patienten in dieser Gruppe, wodurch sich die sogenannte „control event rate“(CER) ergibt.

Anzahl der neg. Ereignisse/Anzahl der Patienten = CER.

Anschließend wird die Anzahl der negativen Ergebnisse in der Experimentalgruppe durch die Anzahl der Patienten in dieser Gruppe dividiert. Damit ergibt sich die „experimental event rate“ (EER).

Anzahl der neg. Ereignisse/Anzahl der Patienten = EER

Abschließend wird die RRR durch nachfolgende Formel berechnet:

$$RRR = \frac{(CER - EER)}{CER}$$

**Praktisches Beispiel:**

Umsetzung der statistischen Methode anhand Studie von Kapitel 1.4.3.  
(aus „Spine“ Vol.20, Nummer 6; Seiten 734-738)<sup>20</sup>.

Eintrag der Daten in eine Vierfeldertafel:

**Tabelle 5:**

| Methode       | APLD     | Mikro.   | $\Sigma$ |
|---------------|----------|----------|----------|
| Erfolg (+)    | 9 ( a )  | 32 ( b ) | 41       |
| Mierfolg (-) | 22 ( c ) | 8 ( d )  | 30       |
| $\Sigma$      | 31       | 40       | N=71     |

APLD= automated percutaneous lumbar discectomy; Mikro. = Mikrodiskektomie

Ein signifikanter  $\chi^2$ -Test von 16,56 (nach Yates) wurde in der Studie angegeben, so da zwi-  
schen den Methoden ein statistisch bedeutsamer Unterschied festzustellen ist.

**Berechnung der RRR:**

Standardformel: 
$$RRR = \frac{(CER - EER)}{CER}$$

CER= control event rate

(Anzahl der Personen fr ein negatives Ereignis in der Kontrollgruppe)

EER= experimental event rate

(Anzahl der Personen fr ein negatives Ereignis in der Experimentalgruppe)

Da in dieser Studie dargelegt werden soll, welchen Stellenwert die APLD gegenber der Mikrodiskektomie aufweist, entspricht die APLD der Experimentalgruppe und die Mikrodiskektomie der Kontrollgruppe.

Folglich entspricht die CER: ( d ) = 8 Patienten oder 20% ( 8 von 40)  
 die EER: ( c ) = 22 Patienten oder 71% ( 22 von 31)

$$RRR = \frac{(CER - EER)}{CER} = \frac{(20\% - 71\%)}{20\%} = -2,55$$

Anhand des Beispiels ist ersichtlich, daß die APLD gegenüber der Mikrodiskektomie eher einen Schaden anrichtet (negative RRR). Wenn die RRR = 0 entsprechen würde, bedeutet dies, daß keinerlei Nutzen von diesem Verfahren zu erwarten ist.

### **Berechnung der Odds Ratio (OR)<sup>14</sup>:**

$$\frac{a * d}{b * c} = OR$$

Interpretativ entspricht die Odds Ratio einem Faktor, um den die Wirksamkeit einer Methode 1 der Wirksamkeit einer Methode 2 überlegen ist.

OR = 1 bedeutet, die Behandlung hat keinen Effekt.

OR > 1 bedeutet die Behandlung ist effektiver als diejenige in der Kontrollgruppe.

OR < 1 bedeutet die Behandlung hat einen negativen Effekt im Vergleich zur Kontrollgruppe.

### **Berechnung der Odds Ratio am praktischen Beispiel:**

Durch die vorherige Ermittlung der RRR ist zu erkennen, daß die APLD der Mikrodiskektomie unterlegen ist. Einen weiteren Hinweis gibt die angegebene Signifikanz mit  $\chi^2=16,56$ , folglich weist die Methode Mikrodiskektomie eine höhere Erfolgsrate auf.

Für die Berechnung der Odds Ratio werden Erfolgs- und Mißerfolgsquoten zweier Behandlungsmethoden ins Verhältnis gesetzt. Es besteht die Möglichkeit diese über die Erfolgswahrscheinlichkeiten oder eine Formel direkt zu errechnen.

Vierfeldertafel für eine vergleichende Evaluation zweier Behandlungsmethoden:

**Tabelle 6:**

| Methoden      | 1 (Behandlungsgruppe) | 2 (Kontrollgruppe) | $\Sigma$ |
|---------------|-----------------------|--------------------|----------|
| Erfolg (+)    | 9 ( a )               | 32 ( b )           | 41       |
| Mißerfolg (-) | 22 ( c )              | 8 ( d )            | 30       |
| $\Sigma$      | 31                    | 40                 | N= 71    |

Für die OR ergibt sich unter Verwendung der Werte der Vierfeldertafel:

$$OR = \frac{a * d}{b * c} = \frac{9 * 8}{32 * 22} = 0,102$$

Da die  $OR < 1$  ergibt sich keinerlei Vorteil für die neue Methode.

### **Berechnung der „number needed to be treated“ (NNT) und der absoluten Risikoreduktion (ARR):**

Nach dem Errechnen der CER (control event rate) und der EER (experimental event rate) muß zunächst die absolute Risikoreduktion (ARR) errechnet werden.

Die ARR berechnet sich durch folgende Formel:

$$(CER - EER) = ARR$$

Der reziproke Wert aus der ARR ergibt die NNT.

$$\frac{1}{ARR} = NNT$$

Oder wenn es sich um prozentuale Angaben handelt:

$$\frac{100}{ARR} = NNT$$

Ein großer Effektunterschied zwischen beiden Behandlungsmethoden führt demzufolge zu einer kleinen NNT und ein kleiner Unterschied zu einer großen NNT. Gibt es keinen Behandlungseffekt, ist die  $ARR = 0$  und die  $NNT = \infty$  (siehe <sup>2</sup>).

Die NNT ist insbesondere bei randomisierten kontrollierten Studien eine geeignete Zahl, das Ergebnis darzustellen. Sie kann immer dann berechnet werden, wenn es sich um den Vergleich zweier Studien mit binären Ergebnissen handelt<sup>2</sup>. Da es sich auch hierbei um eine Einschätzung handelt, sollte sie auf jeden Fall mit einem Konfidenzintervall untermauert werden. Dadurch wird die Spannweite dargestellt, um den ein Wert schwankt.

### **Praktisches Beispiel:**

Zunächst werden die Angaben der Studie wieder in eine Vierfeldertafel eingetragen.

**Tabelle 7:**

| Methode       | APLD      | Mikro.    |
|---------------|-----------|-----------|
| Erfolg (+)    | 29% ( a ) | 80% ( b ) |
| Mierfolg (-) | 71% ( c ) | 20% ( d ) |
| $\Sigma$      | 100%      | 100%      |

APLD= automated percutaneous lumbar discectomy; Mikro. = Mikrodiskektomie.

Es folgt die Berechnung der absoluten Risikoreduktion (ARR):

Standardformel:  $ARR = (CER - EER)$

CER = ( d ) entsprechen 20% oder 0,20

EER = ( c ) entsprechen 71% oder 0,71

Folglich ist die

$$ARR = (0,20 - 0,71) = -0,51$$

Reziproker Wert der ARR entspricht der NNT:

$$1 / ARR = NNT = 1 / -0,51 \approx -2$$

Das negative Vorzeichen bedeutet, da mit der Behandlung von zwei Patienten mit der Mikrodiskektomie ein Mierfolg unter der APLD verhindert werden kann.

### Das Konfidenzintervall der ARR und der NNT:

Nachfolgend wird die Bestimmung des jeweiligen Konfidenzintervalles erklärt. Das 95% ige Konfidenzintervall der ARR kann über den Standardfehler berechnet werden. Dies wird ausführlich im nachfolgenden Absatz dargestellt. In diesem Beispiel reicht das Konfidenzintervall der ARR von  $-30,79\%$  bis  $-71,21\%$ , demnach kann das 95%ige Konfidenzintervall für die NNT folgendermaßen berechnet werden:

$$\frac{100}{-30,79} \approx -3 \quad \text{bis} \quad \frac{100}{-71,21} \approx -1,4$$

Da die NNT in diesem Fall ein negativer Bereich ist, kann sie auch als NNH (number needed to harm) bezeichnet werden<sup>55</sup>.

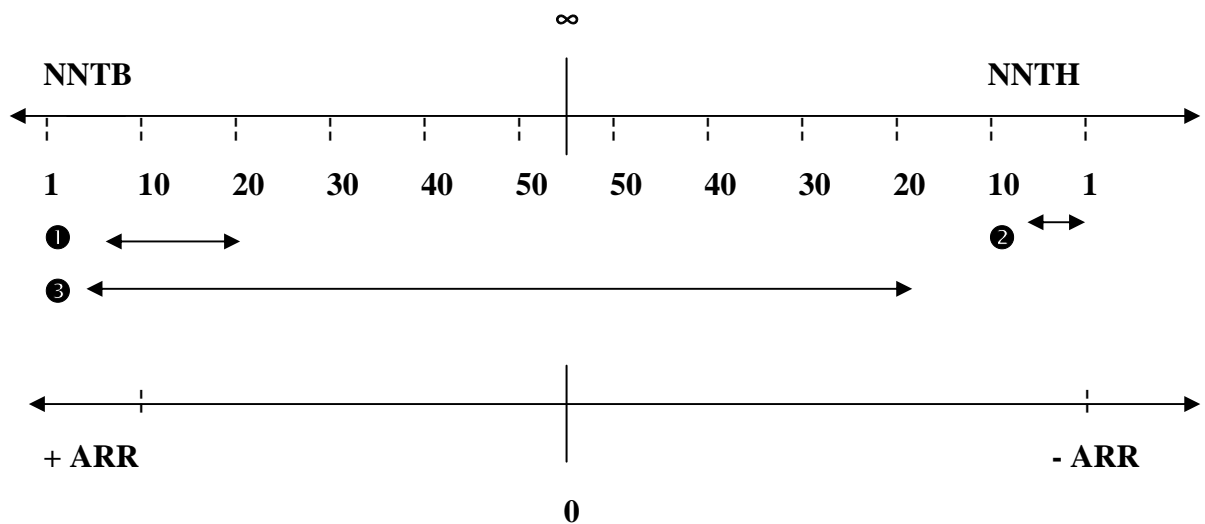
In der Anwendung wird also unterschieden zwischen **NNTB** (engl. : benefit, Nutzen) und **NNTH** (engl.: harm, Schaden), damit können Zahlen ohne negativen Vorzeichen verwendet werden. Dies macht insofern Sinn, da es praktisch unmöglich ist, eine negative Anzahl von Patienten zu behandeln. Dementsprechend bedeutet das für das zuvor genannte Beispiel, daß bei der Behandlung von 1,3 bis 4 Patienten, entsprechend dem Konfidenzintervall, mit dem neuen Verfahren ein positiver Effekt bei durchschnittlich einem Patient weniger festgestellt wird als bei dem alten Verfahren. Aus dieser kleinen NNT- Zahl ist folglich ein großer Behandlungseffekt für das alte Verfahren abzuleiten.

Problematisch wird die Verwendung des Konfidenzintervalls für die NNT bei nicht-signifikanten Studien, bei denen sich das Intervall von einem negativen bis zu einem positiven Bereich spannt<sup>2</sup>.

### Praktisches Beispiel:

Graphische Darstellung mit drei Beispielen für verschiedene Möglichkeiten:

1. Positiver Bereich des 95% Konfidenzintervalls: ARR=10%; CI 95% = 5% bis 15% und NNT=10; CI 95%= 6,7 bis 20 (siehe <sup>2</sup>).
2. Negativer Bereich des 95% Konfidenzintervalls: ARR= -0,51%; CI 95% = -30,79% bis -71,21% und NNT= -2; CI 95% = -3 bis -1,4 .
3. Positiver bis negativer Bereich des 95% Konfidenzintervalls: ARR=10%; CI 95% = -5% bis 25% und NNT= 10; CI 95% = -20 bis 4 (siehe <sup>2</sup>).



Entsprechend der Definition, daß sich bei ARR=0, die NNT gegen unendlich bewegt, gibt es nachweislich keinen Behandlungseffekt in der Mitte der Skala, da  $ARR = 1/NNT$ .



#### 4.5.2. Wie präzise war die Einschätzung des Behandlungseffekts?

Die Präzision des Behandlungseffekts kann zum einen über den Vertrauensbereich (Konfidenzintervall) oder zum anderen mit der Sicherheitsschwelle (Signifikanzniveau) eingeschätzt werden. Das Konfidenzintervall beantwortet die Frage, wie genau die Näherung an den wahren Wert ist, das Signifikanzniveau ist Ausdruck der zulässigen Überschreitungswahrscheinlichkeit, zur Testung der Nullhypothese<sup>34</sup>.

Die wahre Risikoreduktion kann nicht berechnet werden, daher ist nur eine Annäherung an den wahren Wert möglich. Beachtet werden sollte aber, daß ein statistisch signifikanter Wert klinisch völlig irrelevant sein kann. Ebenso werden Studien mit einem P-Wert von  $>0,05$  als statistisch nicht signifikant bezeichnet und damit fälschlicher Weise impliziert, daß im Vergleich zweier Behandlungsmethoden kein Unterschied festzustellen ist<sup>3</sup>. Statistisch wird aber lediglich gezeigt, daß es keine Beweise für einen existierenden Unterschied gibt. Da in der Wissenschaft der Grundsatz gilt, daß jede Annahme zu beweisen ist, darf bei nicht signifikanten Studien nicht davon ausgegangen werden, daß kein Zusammenhang zwischen den Merkmalen besteht<sup>3</sup>.

Dementsprechend sollte auf die Verwendung von Konfidenzintervallen und einem angemessenen großen Studienumfang nicht verzichtet werden.

Da der Einfluß von zufälligen Fehlern um so größer ist, je kleiner die Stichprobe, die postulierten Unterschiede zwischen den Gruppen und je größer die Varianz der Einzelmessung ist<sup>4</sup>.

Nachfolgend soll die Signifikanz und das Konfidenzintervall in ihren Eigenschaften definiert werden.

##### **Definition der Signifikanz ( $\alpha$ -Fehler)<sup>14</sup>:**

In der Erkennung des  $\alpha$ -Fehlers geht es darum, welches Risiko angenommen werden kann, wenn die Nullhypothese ( $H_0$ ) verworfen wird und die Alternativhypothese ( $H_1$ ) angenommen wird, obwohl  $H_0$  in Wirklichkeit zutrifft. Das bedeutet für den Kliniker, das ein diagnostischer Test positiv anzeigt, obwohl reell keine Krankheit vorliegt und somit eine fehlerhafte Aussage beinhaltet.

**Tabelle 8:**

|                                       | Krankheit vorhanden (+)           | Krankheit nicht vorhanden (-)      |
|---------------------------------------|-----------------------------------|------------------------------------|
| Diagnostischer<br>Test <b>positiv</b> |                                   | <b><math>\alpha</math>- Fehler</b> |
| Diagnostischer<br>Test <b>negativ</b> | <b><math>\beta</math>- Fehler</b> |                                    |

In der Statistik haben sich zwei Werte für das maximal tolerierbare  $\alpha$ -Fehlerrisiko eingebürgert, dies entspricht dem Signifikanzniveau oder auch dem P-Wert.

$\alpha = 0,05$  und  $\alpha = 0,01$

Die Wahl des Signifikanzniveaus (auch Sicherheitsschwelle genannt) muß sich am Forschungsgegenstand orientieren. Das Signifikanzniveau sollte um so niedriger angesetzt werden, je gravierender die Konsequenzen einer Fehlentscheidung sind.

Ein P-Wert von  $<0,05$  bedeutet, daß die Wahrscheinlichkeit, über fälschlicherweise sich unterscheidende Ergebnisse zu berichten, geringer als 5% ist.

An dieser Stelle sei angemerkt, daß ein  $\beta$ -Fehler vorhanden ist, wenn klinisch eine Krankheit vorliegt und der diagnostische Test negativ ausfällt.  $1-\beta$  wird die Power eines Testes genannt. Die Power beschreibt die Fähigkeit des Testes einen Unterschied zu entdecken, wenn dieser wirklich vorliegt. Somit ist in der Analogie zum diagnostischen Test  $1-\beta$  analog zur Sensitivität und  $1-\alpha$  analog zur Spezifität zu setzen.

#### **Definition des Konfidenzintervalls (CI)<sup>4</sup>:**

Gewöhnlich wird ein Konfidenzintervall von 95% benutzt.

Das Konfidenzintervall beschreibt den Bereich, in dem mit 95% iger Sicherheit der wahre Wert der Effektgröße für die gesamte Population erwartet werden kann. Das Konfidenzintervall wird mit zunehmender Patientenzahl enger, d.h. die Effektgröße kann präziser eingegrenzt werden. Ist das Konfidenzintervall (CI) nicht bekannt, kann nach mehreren Möglichkeiten vorgegangen werden.

- Durch Angabe des Standardfehlers (SE) (siehe Glossar) von der relativen Risikoreduktion errechnet sich das CI, weil sich die Obergrenze und die Untergrenze von 95% CI mittels der Formel  $\pm 2*SE$  ergibt.
- Wenn die RRR bekannt ist, kann das CI auch darüber errechnet werden.

### **Praktisches Beispiel:**

Da das Signifikanzniveau eine Zahl ist, die vor jeder Studie festgelegt wird, wird im folgenden nicht näher darauf eingegangen.

Das Konfidenzintervall kann berechnet werden, indem zuvor der Standardfehler durch folgende Formel ermittelt wird:

$$SE = \sqrt{\frac{CER(1 - CER)}{N1} + \frac{EER(1 - EER)}{N2}}$$

|     |   |               |
|-----|---|---------------|
| N1  | = Anzahl der Patienten in der Kontrollgruppe            | = 40          |
| N2  | = Anzahl der Patienten in der Experimentalgruppe        | = 31          |
| CER | = Anzahl der Mißerfolge in der Kontrollgruppe ( d )     | = 8 oder 20%  |
| EER | = Anzahl der Mißerfolge in der Experimentalgruppe ( c ) | = 22 oder 71% |

Daraus folgt:

$$SE = \sqrt{\frac{20\%(100\% - 20\%)}{40} + \frac{71\%(100\% - 71\%)}{31}} = 10,31\%$$

$$ARR = ERR - CER = -51\%$$

Das 95%ige Konfidenzintervall (CI) wird berechnet:

$$ARR \pm 1,96 * SE$$

Wobei die Zahl 1,96 angenommen wird bei unendlich vielen Freiheitsgraden bei einem zweiseitigen Test (aus einer t-Wert Tabelle) und  $\alpha = 0,05$ .

Daraus folgt:

$$95 \% CI = -51 \% + 1,96 * 10,31 \% = -30,79 \%$$

und

$$95 \% CI = -51 \% - 1,96 * 10,31 \% = -71,21 \%$$

also von -30,79 % bis -71,21 %.

Die Effektgröße des wahren Wertes liegt demnach mit 95% iger Sicherheit bei -30,79% bis -71,21%, und bedeutet aufgrund der negativen Vorzeichen, daß die Behandlung keinen Effekt hat.

## 4.6. Analyse und Repräsentation der Ergebnisse.

Übersichtshalber sollte eine inhaltliche Zusammenfassung aller überarbeiteter Studien angefertigt werden. Die Ergebnisse der einzelnen Studien werden analysiert über die Heterogenität in den Evidenzen, die Sensitivität, Publikationsbias und Aussagen, die über Subgruppenanalysen getroffen wurden.

Unter Heterogenität in den Evidenzen ist zu verstehen, daß in der Literatur durchaus widersprüchliche Aussagen auftauchen können, welches sich im Vergleich der Studienergebnisse widerspiegelt.

Es ist davon auszugehen, daß insbesondere negative Studien, also Studien über Mißerfolge selten publiziert werden. Diese Publikationsbias führen zu einer verzerrten Darstellung des Sachverhalts und sollten daher unbedingt herausgefunden werden.

Die Ergebnisrepräsentation, zeigt demzufolge die bestehende Evidenzlage, sowie die Qualität der vorhandenen Studien im Hinblick auf Verzerrung und Anwendbarkeit. Vorteilhaft wäre natürlich eine Verbreitung der gewonnenen Erkenntnisse, durch Vervielfältigung.

## 4.7. Anwendung auf den individuellen Patienten.

Im folgenden soll analysiert werden, ob sich die Ergebnisse aus einer oder mehreren Studien auf den individuellen Patienten anwenden lassen. Anhand der nachfolgenden Kriterien kann dabei wie folgt vorgegangen werden.

### **Einschlußkriterien:**

Eine grobe Einordnung ergibt sich mit der Frage: Kann der Patient die vorgegebenen Einschlußkriterien dieser Studie erfüllen?<sup>34</sup> Wenn das nicht der Fall ist, sollte entschieden werden, ob es zwingende Gründe gibt, warum die Studie nicht bei diesem Patienten angewendet werden sollte.

### **Anwendung der NNT:**

Entspricht der Patient den Einschlußkriterien, kann die NNT ohne Veränderung übernommen werden. Liegt eine Studie vor, deren Therapie unabhängig vom Ausgangsrisiko immer zu einem Erfolg führt, kann die NNT für einen individuellen Patienten errechnet werden. Angenommen wird ein Schätzwert, der das Ausgangsrisiko im Verhältnis zu den Probanden in der Studie beschreibt (z.B.: doppeltes oder dreifaches Risiko aufgrund einer Herzinsuffizienz). Diesen Wert bestimmt der behandelnde Arzt durch seine Erfahrung und Einschätzung der Situation. Um die individuelle Zahl der NNT zu erhalten, wird der Schätzwert (Faktor F) durch die NNT der Studie dividiert<sup>4</sup>.

### **Subgruppenanalysen:**

Schließlich ist es auch möglich, daß der Patient in eine in der Studie enthaltene Subgruppenanalyse fällt. Da Subgruppenanalysen häufig erst im Laufe einer Studie entstehen, sind auch hier Kriterien angebracht, die deren Validität beurteilen.

Es kann von einer glaubhaften Subgruppenanalyse ausgegangen werden, wenn folgende Punkte erfüllt sind:

- Wenn der Behandlungseffekt zwischen der Subgruppen und der Gesamtgruppe groß genug ist.
- Wenn das Auftreten des untersuchten Effekts nicht zufällig ist.
- Wenn die Effekte in der Subgruppe vor Studienbeginn als Hypothese formuliert wurden.
- Wenn der Effekt schon in anderen Studien bestätigt werden konnte.

**Studienergebnisse:**

Eine weitere Frage, die gestellt werden sollte, lautet: Wurden alle klinisch relevanten Ergebnisse berücksichtigt?<sup>34</sup>

Klinische Studien messen häufig ihren Therapieerfolg an physiologischen Parametern, wie z.B. Veränderungen in bildgebenden Verfahren (z.B. MRI, Röntgen), weil diese standardisiert, einfach und objektiv zu messen sind. Dies erlaubt aber nur einen indirekten Rückschluß auf den klinischen Verlauf und erlaubt keine Aussage über die Lebensqualität des Menschen. Daher sollte darauf geachtet werden, daß die Studienergebnisse dem Patienten direkt nützen, da die Feststellung eines Sachverhaltes oder eine allgemeine Feststellung als Studienergebnis in diesem Fall unbrauchbar ist. Beispielsweise dokumentieren chirurgische Studien eine verlängerte Lebenszeit, andererseits haben Patienten aber auch ein erhöhtes Risiko direkt nach der Operation zu sterben.

**Nutzen-Risikoeinschätzung:**

Ist der mögliche Therapienutzen groß genug, so daß die mit der Therapie verbundenen Risiken und Kosten in Kauf genommen werden können?<sup>34</sup>

Es sollte abgewogen werden, ob der Nutzen für den Patienten groß genug ist, um die Risiken in Kauf zu nehmen. Hilfreich ist auch hier die Bestimmung der „number to be treated“. Bevor sich endgültig zu einer Therapie entschieden wird, sollte auch das Risiko berechnet werden, welches für sein individuelles Leben entsteht, wenn der Patient unbehandelt bleibt.

**Wunsch des Patienten:**

Sind die Präferenzen und die persönliche Werteinschätzung bei der Wahl der Therapie berücksichtigt worden?<sup>34</sup>

Die evaluierten Ergebnisse und die Einschätzung des behandelnden Arztes sollen an dieser Stelle helfen, zusammen mit dem Patienten eine akzeptable Lösung zu finden. Entscheidend für die therapeutische Vorgehensweise ist, daß die Lebensqualität des Patienten immer im Vordergrund steht. Abgesehen von den genannten Aspekten, muß die Situation des individuellen Patienten durch den behandelnden Arzt beurteilt werden, so daß seine Erfahrung die maßgebliche Entscheidung bestimmt.

Des weiteren muß bei der Anwendung auf den individuellen Patienten beachtet werden, daß er pathophysiologische Unterschiede zu anderen Patienten aufweisen kann, eine Compliance nicht immer möglich ist oder aber Begleiterkrankungen bestehen.<sup>23</sup>