

Departement Innere Medizin

Journal Club DIM

Fortbildung zu Evidence Based Medicine in der Inneren Medizin

R. L. Galeazzi

(modifiziert P. Vernazza)

1. Einführung
2. Artikel-Auswahl
3. Die einzelnen Studientypen
4. Meta-Analysen
5. Glossar

1. Einführung

Ziel

Der Journal Club gibt den Assistenten Gelegenheit zu lernen, Originalartikel aus guten Zeitschriften selber kritisch zu lesen und zu beurteilen.

Organisation

Der Journal-Club findet einmal wöchentlich (Dienstag, 8:20-8:50) für alle Rotationsassistenten des DIM statt. Das Programm wird von PD Dr. P. Vernazza, FB-Leiter Infektiologie, zusammengestellt und ist auf dem Internet unter www.infekt.ch/agenda.htm für alle ersichtlich.

Die präsentierten Artikel sind ebenfalls über das Internet für alle Teilnehmer des Journal Clubs als pdf-file verfügbar. Eine Powerpoint-Vorlage (Download durch Klick auf JC-DIM-Symbol) erleichtert die Vorbereitung für die Vorstellung des Artikels. Die Präsentation soll **max. 15 Minuten** dauern.

Jede(r) Assistent(in) kann selber einen Artikel auslesen und meldet diesen – wenn immer möglich mit Weitergabe des PDF-Files an den Programmleiter. Andernfalls (d.h. wenn bis 6 Wochen vor dem Termin kein selbstgewählter Artikel genannt wurde) wählt der Programmleiter einen Artikel aus.

Der Chefarzt AIM und die Oberärzte AIM und Infektiologie unterstützen den Programmleiter bei der Auswahl von geeigneten Artikeln.

Auswahl der Artikel

Es sollen Artikel aus dem gesamten Gebiet der Inneren Medizin gewählt werden:

A. Klinisch-therapeutische Studien

(Antwort auf die Frage: Nützt eine perorale Antikoagulation bei Vorhofflimmern, welches nicht durch einen rheumatischen Herzfehler verursacht wird?)

B. Epidemiologische Studien über Häufigkeit und Ursachen von internistischen Krankheiten

(Frage: Führt das Rauchen wenig nikotinhaltiger Zigaretten gleich häufig zu Lungenkrebs wie das Rauchen normaler Zigaretten?)

C. Studien über die Wertigkeit neuer oder alter diagnostischer Verfahren und Tests

(Frage: Wie nützlich ist das Lungenzintigramm in der Diagnose von Lungenembolien?)

D. Studien zur Pathophysiologie und Pathogenese internistischer Krankheiten

(Frage: Wie führt eine Lungenembolie zu atemabhängigen Schmerzen?)

E. Meta-Analysen

(Frage: Wie sieht die Evidenz aus nach Durchsicht aller gemachter Studien?)

Präsentation der Studie

Ein Assistent / in bereitet den Artikel vor. Die Präsentation sollte 15-20 Minuten dauern. Dabei geht es nicht in erster Linie darum, eine perfekte Präsentation zu bieten sondern den Artikel kritisch zu beleuchten. Ziel ist es, dass der Wert der Arbeit anschliessend von allen Teilnehmern diskutiert werden kann. Wenn immer möglich, wird ein erfahrenes Kadermitglied einen besonderen Aspekt der Arbeit etwas theoretisch ausführen.

Die Vorstellung der Studie soll in etwa nach folgendem Schema geschehen:

- **Vorstellung** des *Titels*, der *Autoren* und der *Institution*, an welcher die Studie durchgeführt wurde. Handelt es sich um eine *Multizenterstudie*? Sind *Firmen* beteiligt, oder haben die Studie finanziert?
- Vorstellen des *klinischen und wissenschaftlichen Kontexts* der Studie. Kann eine klare *Hypothese* genannt werden? Handelt es sich um Fragestellungen, wie sie auch in unserer Klinik vorkommen? Ist die Studie überhaupt im grossen und ganzen *sinnvoll*?
- Vorstellen der *Methoden, Art der Studie* (randomisierte, Fallkontrollstudie etc.). Ist die *Auswahl der Patienten* sinnvoll und klar definiert? Werden *Ein- und Ausschlusskriterien* genannt? Sind die *Kontrollen* vernünftig gewählt? Ist die Zahl der Patienten und Kontrollen genügend etc.?
- Vorstellen der *wichtigsten Resultate*, eventuell mit eigenen Tabellen oder mit Fotokopien aus dem Artikel.
- **Diskussion** wie sie im Artikel erwähnt ist, anschliessend
- **Kritik der Studie**: Vorstellen der *wichtigsten Resultate*, eventuell mit eigenen Tabellen oder mit Fotokopien aus

Im Prinzip geht es bei jeder Studie um die Beantwortung von 4 Fragen:

1. Welches **Ziel** hat die Studie? d.h. was wird gemessen?
2. Worauf beruht das **Studienkonzept**? d.h. bei wem und wie wird gemessen?
3. Wie lauten die **Resultate**? d.h. wie wurden die Messungen analysiert? beurteilt?
4. Wie **interpretiert** man die Resultate? (Allgemeingültigkeit? Uebereinstimmung mit Biologie? etc.)

Einige Hilfen zur Vorbereitung eines Journal-Clubs

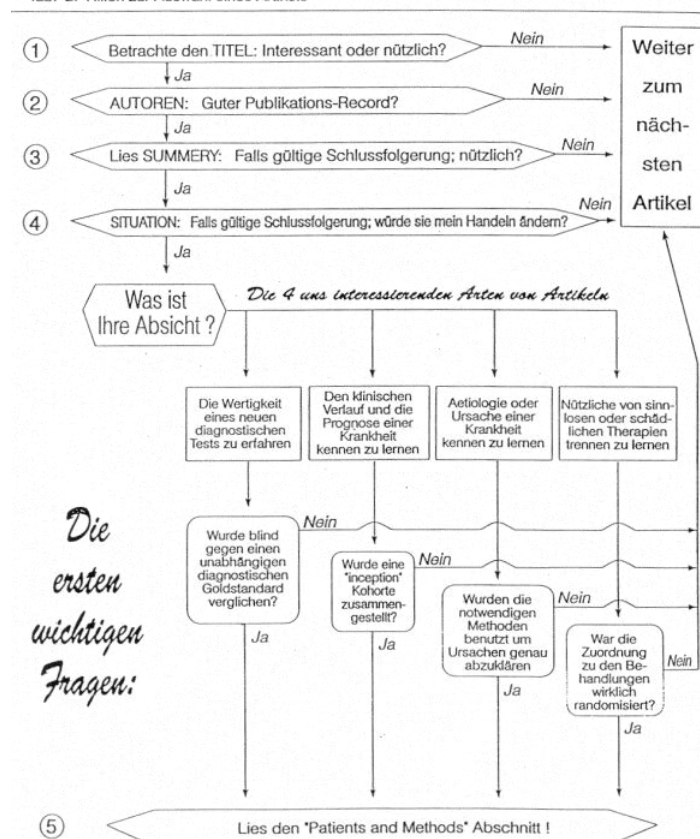
Warum müssen wir eigentlich immer medizinische Literatur lesen?

Tab: 1 10 Gründe um klinische Zeitschriften zu lesen

1. Um andere zu beeindrucken
2. Um beruflich auf dem neuesten Stand des Wissens zu sein
3. Um die Pathobiologie zu verstehen
4. Um herauszufinden, wie ein erfahrener Kliniker ein bestimmtes Problem löst
5. Um zu entscheiden ob ein bekannter oder ein neuer diagnostischer Test eingesetzt werden soll
6. Um die klinische Präsentation und den Verlauf einer Krankheit zu erlernen
7. Um die Ursache einer Krankheit oder den Zusammenhang mit andern Faktoren zu erfahren
8. Um nützliche von unnützen „therapeutischen“ Massnahmen unterscheiden zu können
9. Um den Anspruch auf den Einsatz, Qualität und Kosten-Effektivität von medizinischen und anderen Massnahmen des Gesundheitswesens beurteilen zu können
10. Um sich an den „Letters to the editor“ zu erfreuen

2. Überlegungen bei der Auswahl des Artikels

Tab. 2: Hilfen zur Auswahl eines Artikels



3. Studientypen

(A) Klinisch therapeutische Studien

Die wichtigsten Fragen sind in der folgenden Tabelle zusammengestellt.

Tab: 3 Die wichtigsten Fragen
<p>Design:</p> <ol style="list-style-type: none"> 1. Ist das Ziel der Studie genau beschrieben? 2. Sind die Einschlusskriterien genau beschrieben? 3. Aus welchem Pool wurden die Probanden ausgewählt? 4. Wurden zeitgleiche (oder historische) Kontrollen ausgewählt? 5. Wurde die Intervention / Therapie genau (Dosierung, Ablauf etc.) beschrieben? 6. Wurde die Intervention randomisiert zugeteilt? 7. Auf welche Art wurde randomisiert? (Kriterien?) 8. Nach welchen Kriterien wurde stratifiziert? 9. Ist die Art der Intervention dem Patienten und/oder Therapeuten bekannt, d.h. [doppel]blind? 10. Sind die Kriterien zur Beurteilung der Intervention bekannt? 11. Sind die Messmethoden geeignet, die Studienziele zu beantworten? 12. Wurde die Anzahl (Probanden/Interventionen) aufgrund einer „power-calculation“ festgelegt? 13. Wurde die Dauer der Nachkontrolle („follow-up“) genügend lang gewählt?
<p>Messverfahren:</p> <ol style="list-style-type: none"> 1. Sind Behandlungs- & Kontrollgruppe vergleichbar bezüglich relevanter Kriterien? (Implementation) 2. Ist die Rate der aus der Studie gefallenen Personen („drop outs“) vertretbar klein? 3. Sind die ausgeschiedenen Personen genau beschrieben? 4. Verteilen sich die „drop-outs“ auf allen Studienarme gleichmässig? 5. Wie wurden Nebenwirkungen erfasst und dokumentiert?
<p>Analyse:</p> <ol style="list-style-type: none"> 1. Wurden alle klinisch relevanten Interventions-Resultate beschrieben? 2. Wurden die statistischen Techniken genau beschrieben und/oder zumindestens entsprechende Literaturreferenzen angegeben? 3. Sind die statistischen Berechnungs Techniken vernünftig gewählt? 4. Wurden prognostische Faktoren adäquat berücksichtigt? 5. Wurden für die wichtigsten Resultate Vertrauensintervalle angegeben?
<p>Interpretation:</p> <ol style="list-style-type: none"> 1. Sind die Schlussfolgerungen aus der Studie gerechtfertigt aus den gemachten Beobachtungen? 2. In welchem Ausmass können die Befunde dieser Studie veralgemeinert werden? 3. Lassen sich neue Hypothesen ableiten aus den Hauptbefunden oder aus Untergruppen-Analysen?

Kommentar

Eine wichtige Frage am Schluss eines solchen Artikels ist die Frage nach der klinischen Relevanz bei gegebener statistischer Signifikanz. So führt z.B. eine Senkung des Cholesterinspiegels durch medikamentöse Massnahmen zu einer signifikanten Reduktion der nicht tödlichen Myokardinfarkte. Die Gesamtmortalität ist aber in beiden Gruppen gleich. Hat die medikamentöse Senkung des Cholesterinspiegels einen Sinn? Oder: Die Therapie einer milden Hypertonie (diastolisch 95-100) führt zu einer signifikanten Reduktion cerebraler Infarkte. Es müssen aber 800 Patienten behandelt werden, um einen einzigen Hirnschlag zu verhindern. Sinnvoll?

RRR = relative risk reduction

$$= \frac{\text{Rate of event without intervention} - \text{Rate of event with intervention}}{\text{Rate of event without intervention}}$$

ARR = absolute risk reduction

$$= \text{Rate of event without intervention} - \text{Rate of event with intervention}$$

NNT = number to give intervention to avoid event in 1 patient

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

Besonders bei negativem Unterschied (nicht signifikante Unterschiede) muss man sich fragen, ob die Grösse der Studienzahl adäquat gewählt worden ist. Bei einem Signifikanztest wird immer der Unterschied in einem Mittelwert in Beziehung gebracht zu einem Mass der wahrscheinlichen Variabilität in den beiden Gruppen. Je sicherer nun die Variabilität der beiden Gruppen von einander unterschieden werden kann, desto wahrscheinlicher wird auch ein kleiner Unterschied statistisch signifikant. Wenn daher zu kleine Gruppen gewählt werden, kann die Variabilität in den einzelnen Gruppen genau geschätzt und daher zu wenig genau zwischen den Gruppen unterschieden werden, und so fällt auch eventuell ein signifikanter Unterschied als insignifikant aus. Diesen Fehler nennt man **Fehler zweiter Ordnung II** oder ein **Typ β -error** (= falsch negativ).

Tab: 4: Richtige Entscheidung und Fehler beim Testen der „Null-Hypothese“

Schlussfolgerung am Experiment (Studie)	Realität (Gesamtpopulation)	
	<i>Realer</i> Unterschied zwischen den 2 Gruppen (oder >)	<i>Kein</i> Unterschied zwischen den 2 Gruppen (oder >)
Unterschied <i>festgestellt</i> zwischen den 2 Gruppen (oder >)	<i>wahr positiv</i> korrekte Schlussfolgerung 1-b (= power)	<i>falsch positiv</i> Fehler I. Ordnung a (= p-Wert)
<i>Kein</i> Unterschied festgestellt zwischen 2 Gruppen (oder >)	<i>falsch negativ</i> Fehler II. Ordnung b	<i>wahr negativ</i> korrekte Schlussfolgerung 1-a

(B) Studien über diagnostische Tests

Bei der Beurteilung eines Artikels über einen diagnostischen Test soll man sich folgende Fragen stellen:

Tab: 5 Elemente einer sorgfältigen Evaluation eines diagnostischen Tests

1. Wurde der zu prüfende Test „blind“ verglichen gegen einen „Gold-Standard“ der entsprechenden Diagnostik?
2. Enthielten die Patienten Proben ein vernünftiges Spektrum an milder und schwerer, behandelter und unbehandelter Krankheit?
3. Wurden auch Proben von Patienten mit ähnlichen, häufig mit der gesuchten verwechselte Krankheit?
4. Wurde die Studienanlage wie auch die Einschlusskriterien der Probanden genau beschrieben?
5. Wurde die Reproduzierbarkeit der Testresultate (Präzision) und deren Interpretation (Beurteiler Variation) bestimmt?
6. Wurde der Begriff „normal“ fein genug festgelegt?
7. Falls der Test als Teil einer Abklärungssequenz oder -Gruppe angesehen wird; wurde sein Beitrag zur Wertigkeit der ganzen Abklärungssequenz untersucht und beschrieben?
8. Wurden die technische Durchführung des Tests genau genug beschrieben, so dass eine Reproduktion möglich wäre?
9. Wurde die „Nützlichkeit“ des Tests bestimmt?

Beurteilung und Aussagekraft eines diagnostischen Tests

Häufig angegeben und für die „technische“ Qualität eines Test sicher entscheidend sind Sensitivität und Spezifität eines Testes

Tab: 6 4-Felder Tabelle: „blind“ versus „Gold-Standard“

		Gold Standard		
		Patient hat die Krankheit	Patient hat die Krankheit nicht	
Test Resultat Schlussfolgerung gezogen aus dem Test-Resultat	Positiv Patient scheint die Krankheit zu haben	Richtig positiv a	Falsch positiv b	a + b
	Negativ Patient scheint die Krankheit nicht zu haben	Falsch negativ c	Richtig negativ d	c + d
		a + c	b + d	a+b+c+d

Stabile Verhältnisse:

$$a/(a+c) = \text{Sensitivität}$$

$$d/(b+d) = \text{Spezifität}$$

Häufigkeits-abhängige Verhältnisse:

$$a/(a+b) = \text{positive predictive value}^*$$

$$d/(c+d) = \text{negative predictive value}$$

$$(a+d)/(a+b+c+d) = \text{accuracy}$$

$$(a+c)/(a+b+c+d) = \text{Prävalenz}$$

*der positive predictive value kann auch anders berechnet werden, z.B. nach dem Bayes' Theorem:

$$\frac{\text{Prävalenz} \times \text{Sensitivität}}{(\text{Prävalenz} \times \text{Sensitivität}) + (1 - \text{Prävalenz}) \times (1 - \text{Spezifität})}$$

Sensitivity: Gibt an wieviele der Kranken einen positiven Test haben.

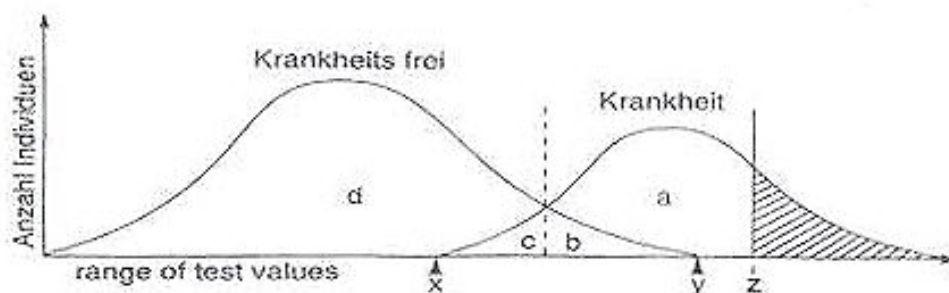
Specificity: Gibt an wieviele derjenigen, die die Krankheit nicht haben, einen negativen Test haben.

Normalerweise gilt, je sensitiver ein Test ist, desto weniger spezifisch ist er, d.h. wenn wir mit einem Test alle Kranken erfassen können, dann werden wir auch viele erfassen, bei denen der Test auch positiv ist, die die Krankheit aber nicht haben. Je spezifischer der Test ist, d.h. je sicherer wir sind, dass ein Patient mit positivem Test die Krankheit auch tatsächlich hat (wenig falsch positive), desto schwieriger ist es auch, mit diesem Test alle Patienten mit der Krankheit zu erfassen. Für alle diagnostischen Tests ist daher die Sensitivity und Specificity wichtig. Auch der sogenannte "positive predictive value" und "negative predictive value" sind zu kennen.

Positive predictive value (PPV): Gibt an, wieviele der Testpositiven auch tatsächlich die Krankheit haben. Oder: Wahrscheinlichkeit, bei positivem Test die Krankheit zu haben, abhängig von Prävalenz der Pathologie (je kleiner Prävalenz, desto geringer PPV, je grösser, desto höher PPV).

Negative predictive value (NPV): Gibt an, wieviele der Negativen tatsächlich die Krankheit nicht haben. Oder: Wahrscheinlichkeit, bei negativem Test die Krankheit nicht zu haben.

Fig. 7:



Schematische Darstellung des Zusammenhangs zwischen Sensitivity und Specificity: Wird der diagnostische Test so gewählt, dass alle Werte rechts vom X als positiv und alle Werte links als negativ gelten, so ist der Test sehr sensibel (alle Kranken werden als positiv erfasst). Aber er ist unspezifisch, da auch ein grosser Teil der Nichtkranken als positiv gewertet wird. Wird ein Test als positiv gewählt bei einem Wert rechts von Y, dann ist der Test recht spezifisch (keine Nichtkranken werden erfasst), aber nicht sensibel, da recht viele Kranke nicht erfasst werden können. Der Kompromiss liegt dazwischen

Beurteilung der Reproduzierbarkeit eines Messwertes

Basis jeglicher Wertung eines diagnostischen Tests bleibt die Genauigkeit der den untersuchten Daten zugrunde liegenden Messwerte. Wenn *ein* Untersucher dieselbe Sache mehrfach misst und diese Messwerte miteinander vergleicht, so beurteilt er die **Intraobserver Variabilität**. Wenn 2 oder mehrere Untersucher dieselbe Sache untersuchen und messen, so ergibt ein Vergleich dieser Werte ein Mass für die **Interobserver Variabilität**. Ein häufiges Mass, die Verlässlichkeit von erhobenen Messwerten anzugeben ist der kappa-Wert (κ). Er berechnet sich folgendermassen:

		Untersucher 2		
		Abnormal	Normal	
Untersucher 1	Abnormal	20	15	35
	Normal	10	55	65
		30	70	

20 % (resp. 55%) Übereinstimmung \Rightarrow total 75 %

dabei werden allerdings rein zufällig übereinstimmende Messwerte vernachlässigt. Erwünscht wäre damit ein Mass, welches den möglichen Einfluss des Zufalls miteinschliesst. Der kappa-Wert ist nun definiert als:

$$K = \frac{\text{Übereinstimmung ausserhalb des Zufalls}}{\text{maximal mögliche Übereinstimmung ausserhalb des Zufalls}}$$

$$K = \frac{O - C}{1 - C} \quad \begin{array}{l} O = \text{beobachtete Übereinstimmung} \\ C = \text{zufällige Übereinstimmung} \end{array}$$

Aus dem Beispiel in obiger Tabelle berechnen sich:

O: $20 + 55 = \mathbf{75}$ [%]

C: - maximal mögliche *zufällige Übereinstimmung*

$$30 \times 35 / 100 = 10.5$$

- maximal mögliche *negative Übereinstimmung*

$$70 \times 65 / 100 = 45.5$$

$$\rightarrow C = 10.5 + 45.5 / 100 = 0.56$$

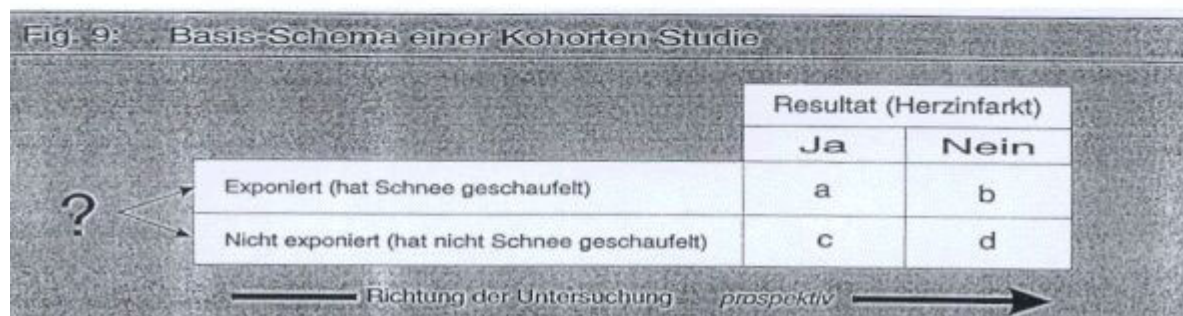
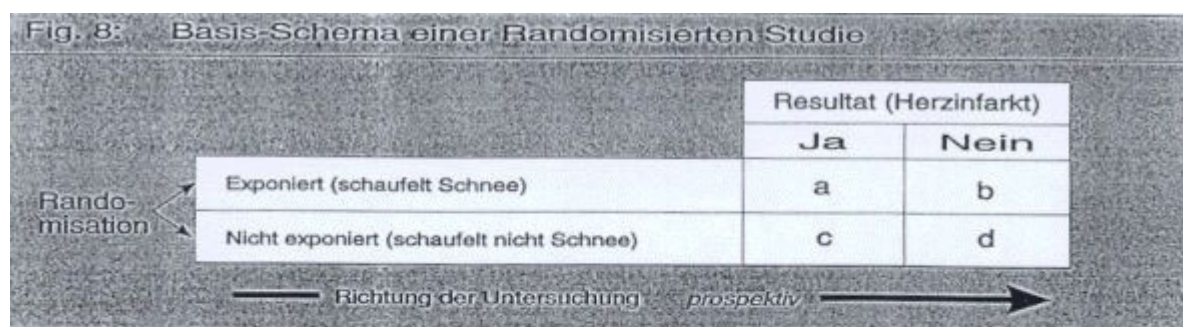
$$K = \frac{0.75 - 0.56}{1 - 0.56} = 0.43$$

(C) Epidemiologische Studien

Die erste Frage, die man sich zu stellen hat, ist nach der Art der Studie. Die Stärke der Argumentation für einen Zusammenhang eines bestimmten Faktors mit einer Krankheit hängt von der Art der Studie ab.

1. Stärkste Studienart: randomisierte, kontrollierte prospektive Studie

Eine randomisierte ausgewählte Gruppe wird einem bestimmten Faktor unterworfen, die andere ebenfalls randomisiert ausgewählte Gruppe wird dem Faktor nicht unterworfen.



Das relative Risiko berechnet sich aus = $\frac{a}{a+b} : \frac{c}{c+d}$

2. Zweitstärkste Studienart: Kohortenstudie

Dabei wird eine repräsentative Gruppe, welche sich dem Faktor unterwirft oder ihm unterworfen ist (z.B. alle Leute einer Stadt, *die Schnee schaufeln*), prospektiv daraufhin untersucht, ob sie eine bestimmte Krankheit über die Zeit erleiden (z.B. einen Myokardinfarkt machen). Sie werden verglichen mit einer andern repräsentativen Gruppe (z.B. alle Leute einer Stadt, welche *nicht* Schnee schaufeln). Dabei kann ebenfalls ein relatives Risiko berechnet werden (nämlich gleich wie beim randomisierten Versuch). Die Schwäche dieser Studie besteht darin, dass der zu untersuchende Faktor nicht randomisiert auf die beiden Gruppen verteilt werden kann. Es besteht daher die Gefahr sogenannter "*confounding variables*". Confounding variables sind

Faktoren, welche nicht kontrolliert werden können (z.B. die Möglichkeit, dass nur jene Leute Schnee schaufeln, die sowieso zu einem Myokardinfarkt neigen).

3. Schwächere Studienart: Fallkontroll-Studie (case-control Study):

Dabei wird bei einer Gruppe von Erkrankten (also bei einer Gruppe von Patienten mit Myokardinfarkt) gesucht, wieviele Patienten einem bestimmten Faktor unterworfen waren (also wie viele Schnee geschaufelt haben). Sie werden verglichen mit einer so ähnlich wie möglich ausgesuchten Kontroll-Gruppe, welche nicht an der Krankheit erkrankt sind, aber sonst so identisch wie möglich mit der erkrankten Gruppe sind; es wird nachträglich, retrospektiv eine Assoziation einer Krankheit mit einem bestimmten Faktor gesucht. Dabei können nicht mehr relative Risiken berechnet werden, da die Gesamtheit der Exponierten (also $a + b$) und die Gesamtheit der Nichtexponierten (also $c + d$) nicht bekannt sind. Es kann nur die sogenannte "odd's ratio" berechnet werden, also:

Fig: 10 Tabellen-Aufstellung für Relatives Risiko (RR) und odd`s ratio (OR)

	Krankheit	Keine Krankheit	
Risikof Faktor vorhanden	a	b	a+b
Risiko Faktor nicht vorhanden	c	d	c+d
	a+c	b+d	

$$\text{Odds`s ratio:} = \frac{a}{c} : \frac{b}{d} = \frac{a}{b} = \frac{c}{d} = \frac{a d}{c b}$$

Dieses Verhältnis gibt nicht an, wieviele Leute, welche einem bestimmten Faktor unterworfen sind, eine bestimmte Krankheit erleiden, sondern wie viele mehr Kranke (Fälle), als Nicht-Kranke (Kontrollen) einem bestimmten Faktor unterworfen waren. Diese Form der Argumentation, die retrospektive Form, ist weniger stark als die prospektive Form einer Kohortenstudie.

4. Schwächste Form: Fallserie (case series)

Sie beschreibt einen bestimmter Prozentsatz von Patienten, die vorher einem bestimmten Faktor unterworfen waren. Solche Studien sind nie beweisend, sie können aber Ausgangspunkt interessanter Hypothesen sein, welche anschliessend mit einer Fallkontroll-, einer Kohortenstudie oder einem randomisierten Versuch untersucht werden können.

Tab: 11 Vor- & Nachteil der **Case-Control Methode***Vorteile:*

1. Gut geeignet zur Untersuchung seltener Krankheiten oder solcher mit langer Latenz
2. Relativ rasch entworfene und durchgeführte Studienform
3. Relativ billig
4. Braucht vergleichsweise wenig Probanden
5. Bereits existierende Daten können gelegentlich verwendet werden
6. Kein zusätzliches Risiko für die Probanden (da „Intervention“ bereits erfolgt!)
7. Erlaubt Untersuchung mehrerer möglicher Ursachen einer Krankheit

Nachteile:

1. Basiert auf Erinnerung oder Daten über frühere Expositionen
2. Validierung der benutzten Daten ist schwierig oder manchmal unmöglich
3. Kontrolle zusätzlicher (wesensfremder) Variablen kann unvollständig sein
4. Selektion einer angemessenen Kontrollgruppe kann schwierig sein
5. Das Vorkommen einer Krankheit in exponierten und nicht exponierten Individuen kann nicht bestimmt werden
6. Methode relativ unvertraut unter Medizinern und schwierig zu erklären
7. Detaillierte Studie von Mechanismen kaum möglich

Tab: 12 Vor- & Nachteil der **Kohorten-Methode***Vorteile:*

1. Gewährt im Prinzip eine detaillierte Beschreibung des Verlaufs nach der Exposition, inklusive Progressionsrate, Stadium der Erkrankung und natürlichen Verlauf
2. Erlaubt Untersuchung multipler möglicher Effekte einer gegebenen Exposition, wodurch sowohl Information über möglichen Nutzen als auch Schaden anfällt
3. Erlaubt die Berechnung der Krankheitsrate unter exponierten als auch nicht exponierten Individuen
4. Erlaubt Flexibilität bezüglich Auswahl von zu registrierenden Variablen
5. Erlaubt gründliche Qualitäts Kontrolle im Messen von Studien Variablen

Nachteile:

1. Grosse Anzahl von Patienten notwendig (Problem bei seltenen Krankheiten!)
2. Möglicherweise lange Nachkontrollen notwendig
3. Aktuelle Handhabung, Gebrauch oder Exposition zu Studienfaktoren kann sich ändern, womit Befunde irrelevant werden
4. Relativ teure Studienform
5. Einhalten der Nachkontrolle kann schwierig sein
6. Kontrolle zusätzlicher Variablen kann unvollständig sein
7. Detaillierte Studie von Mechanismen ist kaum möglich

4. Meta-Analysen

Meta-Analysen klinischer Studien werden in den letzten Jahren immer häufiger. Dabei werden die Ergebnisse ähnlicher Studien zusammen gefasst, um statistisch aussagekräftigere und für eine grössere Patientengruppe gültige Angaben über die Wirksamkeit einer Therapie zu erhalten. Mit der Entwicklung geeigneter statistischer Methoden haben Meta-Analysen seit dem Anfang der achtziger Jahre zunehmend an Bedeutung gewonnen. Der meta-analytische Ansatz blieb jedoch von heftiger Kritik nicht verschont.

Gründe für Meta-Analysen

Viele Studien sind nicht gross genug, um kleine, aber klinisch bedeutsame Unterschiede zu erfassen. Eine Studie kann *falsch negativ* ausfallen, das heisst keinen Unterschied aufzeigen, wenn in Wirklichkeit ein solcher besteht. In diesem Fall spricht man von einem *Fehler II. Art*, dessen Wahrscheinlichkeit (β) berechnet werden kann. Eine Auswertung von 71 „negativen“ randomisierten Therapiestudien hat gezeigt, dass ein klinisch relevanter Unterschied zwischen Kontrolltherapie und experimenteller Therapie selten mit Sicherheit ausgeschlossen werden kann: β betrug in 62 der 71 Studien mehr als 20%. Besser bekannt sind *Fehler I. Art*: Die Resultate einer Studie können einen Unterschied zeigen, der ausschliesslich durch Zufall zustande gekommen ist. Diese Wahrscheinlichkeit (α) wird mit dem p-Wert ausgedrückt; liegt α unter 5% ($p < 0.05$), so spricht man in der Regel von einem signifikanten Resultat.

Während eine Einzelstudie nur Schlüsse auf eine verhältnismässig eng umschriebene Auswahl von Patienten zulässt, lassen sich die Ergebnisse von korrekt durchgeführten Meta-Analysen oft auf einen grösseren Personenkreis anwenden. Meta-Analysen ergeben zudem Hinweise auf den in Zukunft am meisten Erfolg versprechenden oder auf den dringendsten Forschungsansatz. Sie erlauben auch eine genauere Berechnung der dafür notwendigen Anzahl Patienten. Im Vergleich mit konventionellen „Übersichten“ zeichnen sich Meta-Analysen durch die Anwendung quantitativer Methoden und damit möglicherweise eine objektivere Beurteilung aus.

Prinzipien einer Meta-Analyse

Wie jede Studie muss auch der Meta-Analyse eine klare Definition der zu untersuchenden *Fragestellung* vorangehen. Besonders wichtig ist dabei die genaue *Definition der Einschlusskriterien* für die zu berücksichtigenden Studien. Die Suche nach relevanten Studien erfolgt heute meist mit Hilfe elektronischer Datenbanken (MEDLINE etc.) Computerdatenbanken allein identifizieren jedoch weniger als zwei Drittel aller relevanter Studien, weshalb zusätzliche Quellen („Current Contents“, Übersichtsartikel, Standardlehrbücher, Experten, Register unpublizierter Studien etc.) herangezogen werden müssen. Jede Studie muss dann in allen Einzelheiten analysiert werden; besonders wichtig sind Studienanlage, *Therapieschematas*, *Patientencharakteristik* und *Erfassung des Therapieerfolgs*. Weiter geht es darum, die Resultate von Studien, welche die Einschlusskriterien erfüllen, in eine standardisierte Form zu bringen, damit ein Vergleich zwischen den Studien überhaupt möglich wird. Zu diesem Zweck werden oft die „*Odds Ratio*“ berechnet.

Beispiel: Wenn die Chancen 4 zu 6 stehen, dass innerhalb einer bestimmten Zeit unter einem Antibiotikum Durchfall auftritt, so entspricht dies einem „Odds“-Wert von 4/6. Die „Odds-Ratio“ (für Durchfall) berechnet sich dann als Quotient aus den „Odds“ in der Antibiotika-Gruppe (4/6) und den „Odds“ in einer Vergleichsgruppe ohne das betreffende Antibiotikum (z.B. 1/9).

Die „Odds-Ratio“ beträgt damit 6: Die Chance, an Durchfall zu erkranken, ist in der Antibiotikagruppe 6-mal grösser als in der Vergleichsgruppe. Hier geht man also von einem bestimmten Ereignis aus.

Im Gegensatz dazu errechnet sich das *Risiko* eines bestimmten Effektes aus dem Verhältnis der Betroffenen zur Gesamtheit der Behandelten. In unserem Beispiel wäre daher das Risiko, unter dem Antibiotikum an Durchfall zu erkranken, 4/10 oder 40%. In der Kontrollgruppe ergäbe sich ein Risiko von 1/10 oder 10%. Das relative Risiko, unter dem Antibiotikum an Durchfall zu erkranken, ist genau 4 x grösser als ohne Antibiotikum. Hier geht man nicht vom Ereignis, sondern von der Population aus.

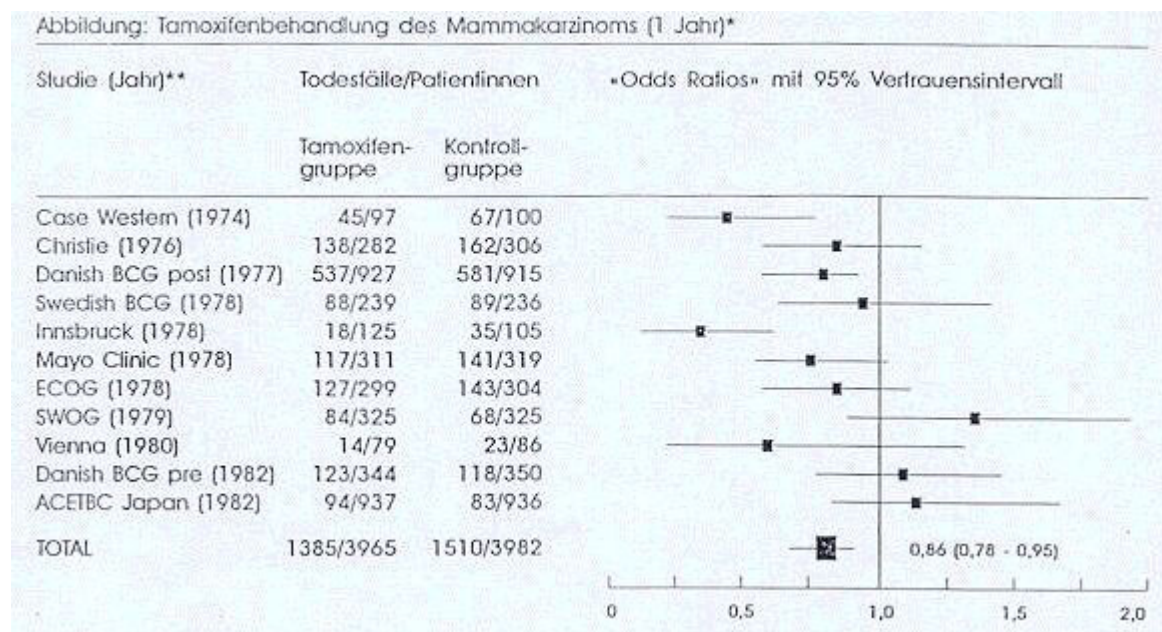
Besteht kein Unterschied zwischen 2 Vergleichsgruppen, so beträgt der Wert für die „Odds-Ratio“ und das relative Risiko genau 1. Solange das untersuchte Ereignis relativ selten auftritt, ergeben „Odds-Ratio“ und relatives Risiko ähnliche Werte. Tritt das Ereignis aber häufiger auf, so ergibt sich - wie das Beispiel zeigt - für die „Odds-Ratio“ einen deutlich grösseren Wert als für das relative Risiko. Im Vergleich mit dem relativen Risiko werden also vor- oder nachteilige Ereignisse mit der „Odds-Ratio“ überschätzt.

Der grosse Vorteil der „Odds-Ratio“ für die Meta-Analysen liegt aber darin, dass man damit mathematisch einfach einen Summationswert aller Studien berechnen und auf seine statistische Signifikanz hin testen kann. Die „Odds-Ratios“ werden zusammen mit dem zugehörigen Vertrauensintervall graphisch dargestellt. Die Ergebnisse von Studien werden bekanntlich durch das Spiel des Zufalls beeinflusst. Das Vertrauensintervall zeigt den Bereich, in dem der wahre Therapieeffekt liegen muss. Das gebräuchliche Vertrauensintervall von 95% gibt den Bereich an, in dem der wahre Therapieeffekt mit einer Wahrscheinlichkeit von 95% zu finden ist. Wenn das 95%-Vertrauensintervall den Wert 1 einschliesst, dann ist der Unterschied in der Wirksamkeit beider Therapie statistisch nicht signifikant, d.h. P ist grösser als 0.05.

Der letzte Schritt besteht in der Berechnung des Summationswertes mit Vertrauensintervall, in dem alle Studien vereint werden. Dafür stehen verschiedene statistische Methoden zur Verfügung, die jedoch meistens zu sehr ähnlichen Resultaten führen. Allgemein werden dabei Studien mit grösseren Patientenzahlen, die mehr Information beisteuern, stärker gewichtet als kleine Studien. Dieser Ansatz ist offensichtlich der Berechnung eines einfachen arithmetischen Mittelwertes überlegen.

Beispiel: Tamoxifen bei Mammacarcinom. 1992 wurden die Ergebnisse einer grossangelegten Meta-Analyse aus der Onkologie veröffentlicht. (Peto R. et al; Lancet 339; p 1-15) Es ging um die Bedeutung der *adjuvanten Therapie* nach der chirurgischen Entfernung des Primärtumors bei Mammacarcinom ohne Fernmetastasen. Insgesamt wurden 133 randomisierte Studien analysiert, in denen die Wirkung von verschiedenen systemischen Therapien (Hormon-, Chemo- oder

Immuntherapie) geprüft wurde. Das folgende Beispiel greift einen Teilaspekt dieser Meta-Analyse heraus. Es betrifft Studien, in denen die Wirkung einer einjährigen Therapie mit einem Antiöstrogen Tamoxifen untersucht worden war. In sieben Studien war die Kombination *Chemotherapie/Tamoxifen* mit *Chemotherapie* allein und in 4 Studien *Tamoxifen* mit *keiner* adjuvanten Therapie verglichen wurde.



* modifiziert nach Early Breast Cancer Trialist' Collaborative Group. Lancet 1992; 339: 1-15

Obenstehende Abbildung zeigt die Sterblichkeit aus verschiedenen Studien als „Odds ratio“ mit einem Vertrauensintervall von 95%. Acht Studien ergaben keinen statistischen Effekt (die Vertrauensintervalle - markiert durch die Linie - kreuzen den Punkt 1,0!). Nur 3 Studien („Case Western“, „Danish-BCG post“, „Innsbruck“) zeigen eine der Kontrolltherapie signifikant überlegene Wirkung von Tamoxifen („odds-ratio“ unter 1,0, die 95%-Vertrauens-intervalle kreuzen die 1,0-Linie *nicht*).

Man könnte nun leicht zum Schluss kommen, dass eine 1-jährige Tamoxifen Gabe keine Vorteile bietet, da die überwiegende Mehrheit der Studien keinen eindeutigen Nutzen ergaben. Eine Meta-Analyse aller Studien zeigt jedoch eine signifikante Wirkung mit einer „odds-ratio“ von 0.86. Die einjährige Tamoxifen Gabe kurz nach Diagnosestellung hat somit die „odds“, in den folgenden Jahren an einem metastasierenden Mamma-Ca zu sterben, um 14 % reduziert. Das entspricht einer Reduktion des Risikos um etwa 10%. Eine noch bessere Wirkung (Reduktion der Sterblichkeit um etwa 20%) wird durch eine längerdauernde Tamoxifen-Therapie erreicht.

Meta-Analysen von Tamoxifen-Studien haben schon vor Jahren zu einer Änderung der therapeutischen Richtlinien geführt, die ohne derartige quantitative Auswertungen möglichst vieler individueller Studienresultate kaum zustande gekommen wäre. (vgl. Consensus conference. JAMA 1985; 254:3461-3). Die Richtigkeit dieses Entscheides wird durch neuere Daten klar bestätigt.

Probleme von Meta-Analysen

Obwohl Meta-Analysen die Erfassung von „kleineren“ Wirkungen zu einem frühen Zeitpunkt ermöglichen können, darf diese Methode nicht unkritisch angewendet werden, d.h. nicht ohne Qualitätskontrollen. Die wichtigsten zu berücksichtigenden Fragen sind:

Ist die Qualität der berücksichtigten Studien genügend?

⇒ Als Mindestanforderung gilt, dass nur korrekt randomisierte Studien mit vollständigen Angaben über alle initial in die Studie aufgenommenen Personen meta-analysiert werden. Der Randomisierungsvorgang, der leider oft ungenügend beschrieben ist, verdient darum besondere Aufmerksamkeit. Die Intervention (z.B. Therapie) und die Beurteilungsverfahren müssen vergleichbar sein. Bei Medikamenten Studien muss z.B. gefordert werden, dass die Beurteilung der Therapieergebnisse blind erfolgt.

Wurden alle relevanten Studien berücksichtigt?

⇒ Da „positive“ Studien mit grösserer Wahrscheinlichkeit publiziert werden als Studien, die keinen oder einen „negativen“ Effekt zeigen, sollte versucht werden, auch unpublizierte Studien einzuschliessen oder zumindest bei der Diskussion anzugeben.

Haben die berücksichtigten Studien teilweise entgegengesetzte Resultate erbracht?

⇒ Der Variabilität der Ergebnisse wird oft nicht genügend Beachtung geschenkt. Das Problem der Heterogenität von Studienresultaten kann nicht durch die Anwendung statistischer Tests gelöst werden. Wenn stark voneinander abweichende Resultate vorliegen, müssen biologische und klinische Überlegungen in die Bewertung / Diskussion der Meta-Analyse mit einbezogen werden. Der Verdacht auf einen Selektions- und/oder Bewertungs-Bias drängt in derartigen Fällen auf.

Wie „robust“ sind die Ergebnisse der Meta-Analyse?

⇒ Bei jeder Meta-Analyse gibt es mehr oder weniger arbiträre Entscheide (Ein-/Ausschluss von Studien, Wahl der statistischen Methoden, Interpretation resp. Gewichtung bestimmter, insbesondere divergierender Resultate). Das Resultat einer Meta-Analyse sollte von solchen Entscheiden möglichst unabhängig sein.

5. Glossar

Actual analysis (=life table analysis / = Cutler-Ederer Methode)

Methode um Ueberlebenszeiten zu analysieren für Beobachtungen (censored/noncensored), deren Dauer zu Intervallen gruppiert wurden.

Bevölkerung (Population, Gesamtheit)

1. Alle Einwohner eines Landes oder einer Untereinheit
2. Bei Stichproben: Alle Einheiten, aus welchen eine Stichprobe gezogen werden kann; nicht nur Personen, die Einheiten können auch Krankengeschichten oder Ereignisse etc. sein.

Bias

(Ein Merkmal eines Verfahrens, welches zu einer systematischen Differenz zwischen der Wirklichkeit und der resultierenden Information führt.). Systematische, unbewusste **Verzerrung**. Das Wort Bias hat sich im Deutschen auch eingebürgert. Nach Ursprung der Verzerrung werden verschiedene Arten unterschieden, die wichtigsten sind:

1. Selektionsbias
2. Messverzerrung (z.B. recall-bias, observation-bias, interviewer bias)
3. Confounding

Censored observation

Beobachtung, deren Wert für die Auswertung nicht berücksichtigt werden kann, weil die Beobachtungszeit nicht lang genug (gemäß individuellen Studienbedingungen!), oder das geforderte Ereignis (z.B. Rezidiv oder Tod) noch nicht eingetreten ist.

Compliance

Mass für die Bereitschaft, eine medizinische Empfehlung oder Anordnung zu befolgen. Analog dazu wird bei Nichtbefolgen medizinischer Ratschläge und Therapien von Non-Compliance gesprochen.

Confounding

Systematische Verzerrung, hervorgerufen durch eine Vermischung von Effekten, welche zu einer falschen Berechnung des angestrebten Resultates führt. Der Fehler resultiert aus einem Zusammenhang zwischen einen oder mehreren nicht berücksichtigten Faktoren mit den Expositionsvariablen sowie dem "outcome" (Krankheit, Heilung etc.)

Design (Studienplan)

Anlage einer Studie, die der Fragestellung soweit als möglich gerecht werden sollte. Sollte vor Beginn jeder Studie schriftlich festgehalten werden.

Deskriptive Epidemiologie

Befasst sich mit der Beschreibung der **Häufigkeit** bestimmter Erkrankungen oder Gesundheitsstörungen und deren Verteilung in der Bevölkerung. (Oft Basis für den Ausgang einer Hypothese.)

Drop-out

Studienteilnehmer, die ausfallen oder ausscheiden, meist ein Problem bei **Follow-Up-Studien** und **randomisierten klinischen Studien**.

Experimentelle Epidemiologie

Teil der Epidemiologie, der sich mit der Testung eines experimentellen, kontrollierbaren Faktors (Intervention im Rahmen kontrollierter Studien, **randomisierte klinische Studien** oder bevölkerungsbezogene **Interventionsstudien**) befasst.

Fall

In der **Epidemiologie** eine Person in der **Bevölkerung** oder der Studiengruppe mit der definierten Krankheit oder Gesundheitsstörung, die untersucht wird.

Fall-Kontroll-Studie

Studie, welche identifizierte Fälle im Bezug auf wichtige Charakteristika mit möglichst identischen **Kontrollen** vergleicht (sie ist immer **retrospektiv**). Die Analyse ist meist retrospektiv, die Studie kann aber auch prospektiv angelegt werden.

Dieser Vergleich hilft Faktoren zu finden, welche unterschiedlich in den beiden Gruppen vorkommen und das Auftreten der Krankheit den Patienten erklären.

Fehler 1. Art (α -Fehler)

Ein statistisch signifikanter Unterschied wird gefunden, obwohl er in Realität in der **Bevölkerung** nicht vorhanden ist (Rückweisung der **Nullhypothese** zu Unrecht). Der p-Wert gibt die Wahrscheinlichkeit des Fehlens der 1. Art an.

Feldstudie

Begriff, der manchmal in den Studien gebraucht wird, die ausserhalb der Kliniken (in Arztpraxen oder in der **Bevölkerung**) durchgeführt werden.

Gauss'sche Verteilung

Siehe Normalverteilung.

Gipfelwert (Mode)

In einer Verteilung derjenige Wert, welcher am häufigsten vorkommt.

Inzidenz (Neuerkrankungsziffer)

Zahl neuauftretener Fälle in einer definierten Bevölkerung pro Zeiteinheit (meist pro Jahr), bezogen auf die gleiche Bevölkerung (meist pro 1'000 oder pro 100'000).

Irrtumswahrscheinlichkeit

Ausgedrückt als p (probability). Sie ist ein Mass für die Wahrscheinlichkeit, dass ein beobachteter Unterschied auf Zufall beruht, d.h. dass die **Nullhypothese** zutrifft.

p quantifiziert die Wahrscheinlichkeit, mit welcher ein Fehler 1. Art gemacht wird.

 κ (=Kappa-Wert)

Mass für die Reproduzierbarkeit eines Messwertes unter Berücksichtigung zufälliger/beobachteter Übereinstimmung.

$$\kappa = \frac{O - C}{1 - C}$$

O = beobachtete Übereinstimmung

C = zufällige Übereinstimmung

Kaplan-Meier Kurven

Methode um Ueberlebenszeiten zu analysieren und graphisch darzustellen für Beobachtungen (censored/noncensored), wobei deren exakte Dauer verwertet wird.

Kohorte

Eine Bevölkerungsgruppe, welche durch eine gemeinsame Erfahrung oder **Exposition** gekennzeichnet ist.

Kontrollen

Probanden oder Patienten, welche bei analytischen Studien in möglichst allen Kriterien gleich wie die Fälle oder bei experimentellen Studien, wie die Interventionsgruppe sind, bei welchen aber nichts unternommen wird.

Lead time

Zeitgewinn zwischen Diagnose einer Krankheit und dem untersuchten Endpunkt (z.B. Tod), der ausschliesslich durch vorzeitige Diagnose-Stellung zustande kommt und damit keinen echten Ueberlebensgewinn darstellt.

Letalität

Verhältnis der Todesfälle an einer Krankheit zur Zahl der Erkrankungsfälle an derselben Krankheit. Mass für die Gefährlichkeit einer Erkrankung.

Medianwert

Derjenige Wert einer Verteilung oberhalb und unterhalb dessen gleich viele Einzelwerte liegen (auch Fünfteiler **Perzentile**).

Mittelwert = x (Durchschnittswert oder arithmetischer Mittelwert)

Siehe auch **Normalverteilung**, errechnet aus der Summe aller Einzelwerte (x_i), dividiert durch die Anzahl Einzelwerte(n):

$$x = \frac{\sum x_i}{n}$$

Mode

Siehe Gipfelpunkt.

Morbidität

Mass für die **Häufigkeit** von Krankheiten in der **Bevölkerung** ohne Unterscheidung zwischen **Inzidenz** und **Prävalenz**.

Mortalität (Sterblichkeit)

In einer **Bevölkerung** gestorbene Personen, bezogen auf diese **Bevölkerung**. Ausgedrückt als Sterbeziffer

$$\text{(rohe) Mortalität} = \frac{\text{In einem Jahr gestorbene Einwohner eines Gebietes} \times 1'000}{\text{durchschnittliche Bevölkerung in demselben Gebiet und Jahr}}$$

(alters-)spezifische Mortalität = Mortalität einer bestimmten (Alters-)Gruppe

Normalverteilung (Gauss'sche Verteilung)

Kontinuierliche, symmetrische Verteilung, deren Enden beidseits ins Unendliche reichen, bestimmt durch zwei Parameter: Mittelwert und Standardabweichung. Mittelwert, Medianwert und Gipfelpunkt sind bei der Normalverteilung gleich.

Normalverteilung charakterisiert durch:

$$\text{Mittelwert: } \bar{x} = \frac{\sum X_i}{n} \quad \begin{array}{l} x_i = \text{Einzelwert} \\ n = \text{Zahl der Einzelwerte} \end{array}$$

und Standardabweichung (standarddeviation SD):

$$\text{SD} = \frac{\sqrt{\sum (x_i - \bar{x})^2}}{n-1}$$

Normwert

1. = Normalwert
2. = Das Uebliche (was man erwartet)
3. = Das Angestrebte, z.B. die Norm sollte sein, dass alle Autofahrer sich angurten.

Nullhypothese

Hypothese, deren Irrtumswahrscheinlichkeit mit statistischen Tests untersucht wird. Sie besagt, dass beobachtete Unterschiede in den Resultaten von Studien oder Untersuchungen nur durch Zufall zustande gekommen sind.

Odds

Verhältnis von Kranken zu Nicht-kranken in einer Population; auch:

$$\frac{\text{Prävalenz}}{1 - \text{Prävalenz}}$$

Odds-Ratio

Verhältnis von 2 Odds: Wird in Fall-Kontroll-Studien gebraucht:

$$\frac{\text{exponierte Fälle}}{\text{nicht-exponierte Fälle}} : \frac{\text{exponierte Kontrollen}}{\text{nicht-exponierte Kontrollen}}$$

Ergibt wieviel mal häufiger die Exposition unter den Fällen als unter den Kontrollen ist.

Perzentil

Einteilung von Messungen aufgrund ihrer prozentualen Verteilung. Häufig gebraucht bei Beurteilung von Verteilungen wie Gewicht oder Längenzuwachs bei Kindern (z.B. ein Gewicht unter der 5er **Perzentile** heisst, das Kind weist ein Gewicht auf, welches kleiner ist als dasjenige von 95% der Kinder seiner Grösse oder Altersgruppe).

Prävalenz

Bestand an Fällen (Vorkommen) gegenüber Auftreten (Inzidenz) einer bestimmten Krankheit zu einem bestimmten **Zeitpunkt**, bezogen auf die Einwohnerzahl (auch Punkt-

Prävalenz genannt). Wird die Zeiteinheit länger gewählt, so kann z.B. über eine Wochenprävalenz gesprochen werden (Periodenprävalenz).

Prospektive Studien

(besser **Langzeitstudien**, **longitudinale Studien** oder **Kohortenstudien**) Personen werden hinsichtlich des Auftretens von Krankheiten oder **Risikofaktoren** langfristig beobachtet.

Randomisierung

Zuordnung nach Zufallsprinzip zu Untersuchungs- oder Kontrollgruppe, in experimentellen oder kontrollierten Studien.

Randomisierte klinische Studie (RKS oder RCT = randomised controlled trials)

Ein sorgfältig geplantes, in Übereinstimmung mit ethischen Richtlinien durchgeführtes Experiment zur Überprüfung von Hypothesen mittels einer Zufallszuteilung der Studienteilnehmer in mindestens zwei Gruppen.

Relatives Risiko

Verhältnis der kumulativen **Inzidenz** exponierter Individuen gegenüber derjenigen nichtexponierter Individuen.

$$RR = \frac{I_E}{I_N}$$

IE = Inzidenz Exponierter

IN = Inzidenz Nicht-Exponierter

Sensitivität (Empfindlichkeit)

einer Untersuchung ist das Mass für das Vermögen der Untersuchung, wirklich Kranke korrekt zu identifizieren.

Signifikanz

Statistisches Mass für die Irrtumswahrscheinlichkeit, ausgedrückt als p. Im deutschen Sprachgebrauch wird signifikant manchmal für "bedeutend" verwendet - diese Ausdrucksweise sollte in wissenschaftlichen Publikationen vermieden werden. Signifikanzteste werden gebraucht um auszudrücken, wie wahrscheinlich die Nullhypothese (meist H_0 = kein Unterschied zwischen den verglichenen Gruppen) ist. Ist $p < 0.05$, d.h. die Wahrscheinlichkeit, dass die Nullhypothese zutrifft, ist kleiner als 5%, so wird sie verworfen und die alternative Hypothese angenommen (H_1 = es besteht ein Unterschied zwischen den verglichenen Gruppen).

Spezifität

einer Untersuchung ist das Mass für das Vermögen der Untersuchung, Gesunde korrekt zu erkennen.

Standardabweichung (SD)

Siehe Normalverteilung, errechnet sich als Quadratwurzel aus der Varianz.

Stichprobe

Auswahl von Probanden mittels eines (meist zufälligen) Verfahrens, das heisst jeder Proband hat eine bestimmte, vorher definierte Chance, in die Stichprobe aufgenommen zu werden. Ziel ist, ein repräsentatives Abbild der Bevölkerung zu erhalten.

Trend (Tendenz)

1. Eine Langzeitveränderung in einer Datenfolge, z.B. in einer Zeitserie, welche in eine bestimmte Richtung geht.
2. Das Wort Trend wird auch benützt, wenn Zusammenhänge zwischen verschiedenen Stichproben oder Einzelgruppen von Daten gefunden werden, welche in eine bestimmte Richtung weisen, aber nicht statistisch signifikant sind.

Vergleichsgruppe (Kontrollgruppe)

Eine Vergleichsgruppe, welche in allen wesentlichen Faktoren mit der Hauptuntersuchungsgruppe (Fallgruppe) übereinstimmt, sich daher von ihr nur durch einen kontrollierten, in der Studie untersuchten Einfluss oder Faktor unterscheidet.

Vertrauensintervall (Confidence limits)

bezeichnet die Grenzen, innerhalb welcher ein geschätzter Mittelwert mit einer Wahrscheinlichkeit von 95% liegt. Liegt der Wert "kein Unterschied" (z.B. 0 in einer rationalen Skala, 1 bei Verhältnissen, rel. Risiko) innerhalb der C.I., so besteht kein signifikanter Unterschied. Beispiel: Es wurde studiert, ob Reserpin Brustkrebs verursache: das relative Risiko war 1.7 C.I. 0.8 . R.; d.h. kein signifikanter Unterschied in den untersuchten Kohorten.

Literatur

Abramson J.M.	englisch	Making sense of data	Oxford University Press 1988
Bailar J.C. III Mosteller F.	englisch	Medical Use of Statistics	New England Journal of Medicine Boston, Massachusetts
Colton Th.	englisch	Statistics in Medicine	Little, Brown and Company, Massachusetts 02106
Dawson-Saunders B. Trapp R.G.	englisch	Basic and Clinical Biostatistics	Appleton and Lange Norwalk, CT
Egger M. Smith D.	<i>deutsch</i>	Meta-Analysen	Pharma-Kritik; 14, No 14, pp 53-6
Greenhalgh T.	englisch	How to read a paper	BMJ 1997
Harten H-U. et al.	<i>deutsch</i>	Statistik für Mediziner	VCH Verlagsgesellschaft, 1993
Hennekens Ch.H. Buring J.E	englisch	Epidemiology in Medicine	Little, Brown and Company, Boston, Massachusetts 02106
MaMahon B. Push Th.F	englisch	Epidemiology Principles and Methods	Little, Brown and Company, Boston, Massachusetts 02106
Sachs L.	<i>deutsch</i>	Angewandte Statistik	Springer Verlag Berlin, 1984
Sackett P.L. et al	englisch	Clinical epidemiology	Little, Brown and Company, Boston / Toronto
Schlesselmann J.J.	englisch	Case-Control Studies	Oxford University Press 1982