

Some Contributions of Statistics to Environmental Epidemiology

Seokwoo Park

[Epidemiologie]

- Epidemiology :

“ the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems.”

“ Epidemiologie handelt von der Verteilung und den Determinanten von Gesundheit und Krankheiten, Morbidität, Verletzung, Beeinträchtigung, usw.. in Bevölkerungen.”

- Umweltepidemiologie :

befäßt sich mit der Verteilung von Krankheiten, physiol. Variablen und sozialen Krankheitsfolgen in menschl. Bevölkerungsgruppen, die durch Umweltfaktoren beeinflusst bzw. hervorgerufen werden.

[Studiendesign]

▪ Prospektive Studie :

- Bedeutet im Sinne von „vorausschauend“
- Erhebung neuer Daten
- Hohe Datenqualität
- Zeit- und kostenintensiv

▪ Retrospektiv Studie :

- Im Gegensatz zu prospektive Studie
- Benutzung vorhandener Daten
- Schnell und billig
- Fehlende Werte vorhanden möglich

▪ Korrelationsstudie (ökologische Studie)

Ökologische Studien sind spezielle epidemiologische Studien, die charakterisiert sind durch den Versuch, Exposition und Krankheit auf der Ebene von Regionen und Bevölkerungsgruppen in Beziehung zu setzen, und nicht auf der Stufe einzelner Individuen.

▪ Querschnittstudie

In einer Querschnittstudie werden Bestrahlung (zum Beispiel die gemessene Exposition gegenüber den elektromagnetischen Feldern von Mobilfunkbasisstationen) und die zu untersuchende Erkrankung (zum Beispiel Schlafstörungen oder Kopfschmerzen) zeitgleich in Beziehung zueinander gesetzt und ausgewertet.

▪ **Randomisierte Kontrollierte Studie**

Die randomisierte kontrollierte Studie (RCT englisch: randomized controlled trial) ist in der medizinischen Forschung das nachgewiesene beste Studiendesign, um bei einer eindeutigen Fragestellung eine eindeutige Aussage zu erhalten und die Kausalität zu belegen.

▪ **Interventionsstudie**

Mit Hilfe von Interventionsstudien werden Präventivmaßnahmen, z. B. durch Änderung von Ernährungsgewohnheiten, Zu- oder Abnahme von körperlicher Aktivität oder Medikamente, auf ihre Wirksamkeit geprüft. Interventionsstudien erfordern in der Regel eine Randomisierung, d. h. eine Zuordnung der Teilnehmer zur Placebo- oder Verumgruppe nach dem Zufallsprinzip. Die Interventionsstudie kann offen, blind oder doppelblind durchgeführt werden.

[Analytische Studie]

▪ **Exposition :**

Die Exposition bezeichnet die Summe aller Umgebungseinflüsse, die auf einen Gegenstand oder ein Lebewesen einwirken, wie zum Beispiel Krankheitserregern, toxischen chemischen Elementen oder Verbindungen oder physikalischen Einflüssen wie Hitze, Lärm oder Strahlung

1. Kohortenstudie

▪ Kohorte :

Gruppe von Personen, die zur Definitionszeit Eigenschaften gemeinsam haben und die eine Zeit lang beobachtet werden.
z.B. Alterskohorten, Geschlechtskohorte, Geburtskohorte, usw..

▪ Kohortenstudie

(Längsschnittstudie, Follow-up Studie) :

Für eine Kohortenstudie werden Probanden in die Studie eingeschlossen und zum Zeitpunkt der Rekrutierung werden ein oder mehrere Risikofaktoren erhoben. Risikofaktoren sind z.B. Blutdruck, Rauchen oder Essgewohnheiten. Die Teilnehmer werden über die Zeit beobachtet und es wird erfasst, ob ein vorher definierter Endpunkt eintritt.

Der Endpunkt ist meist eine Erkrankung oder zumindest der Vorläufer einer Erkrankung, es kann natürlich auch der Tod sein.

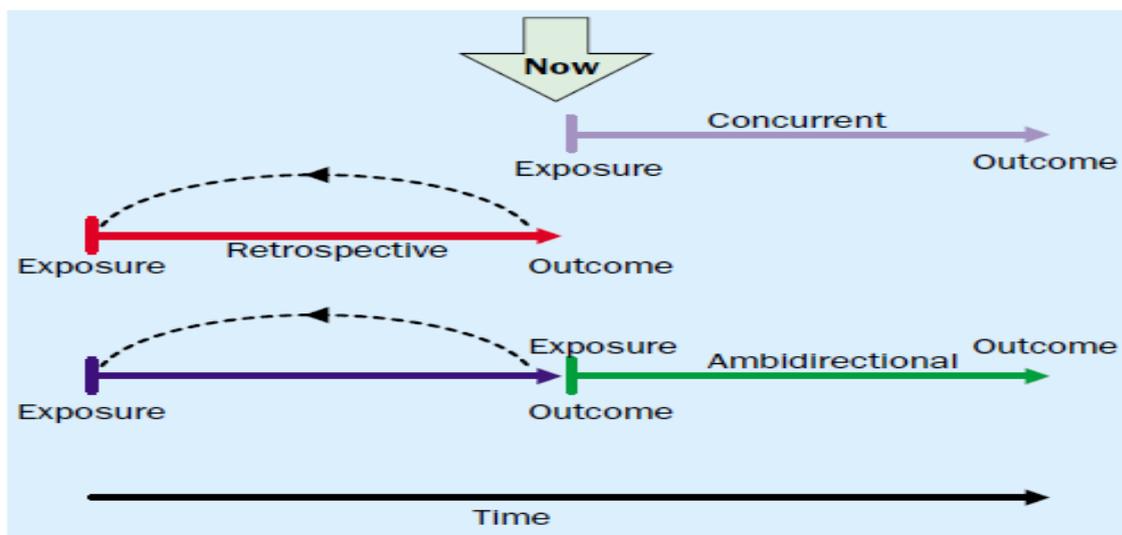


Figure 2: Schematic diagram of concurrent, retrospective, and ambidirectional cohort studies

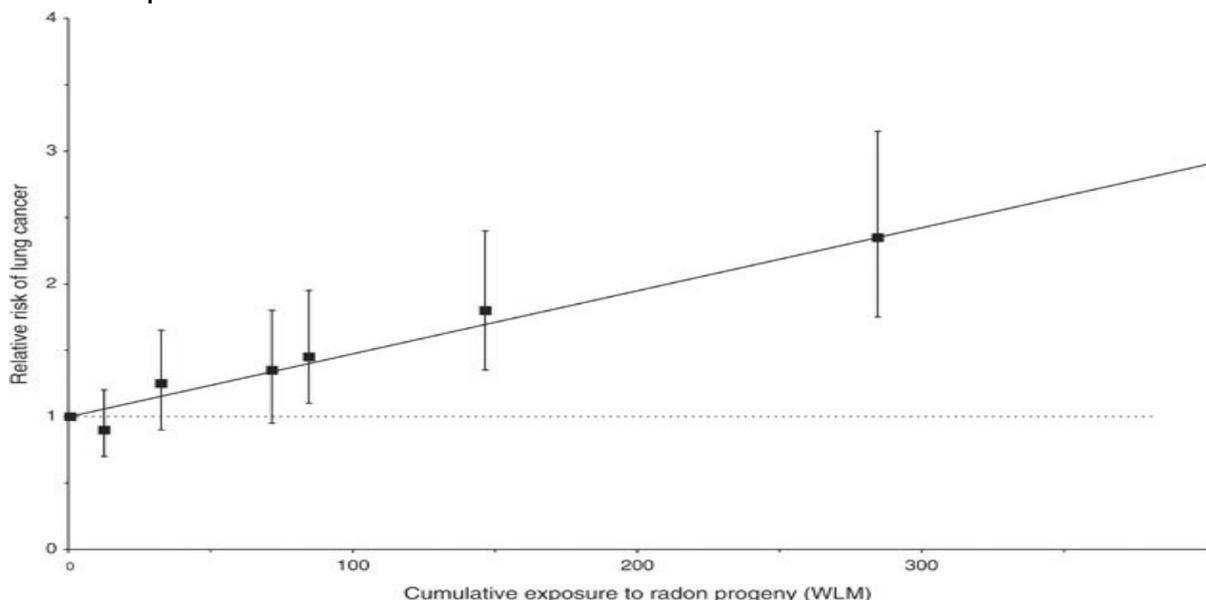
- Prospektive Kohortenstudie :

Eine Kohortenstudie kann prospektiv oder retrospektiv sein. Die Kohortenstudie ist eindeutig prospektiv, wenn der Risikofaktor jetzt erfasst wird und die Probanden dann weiterbeobachtet werden, um den in der Zukunft eintretenden Endpunkt zu erfassen.

- Retrospektive Kohortenstudie :

Das Design ist retrospektiv, wenn sowohl der Risikofaktor, als auch der Endpunkt noch vor der Planung der Studie erfasst bzw. eingetreten ist. Es ist aber interessant, dass unter den Spezialisten keine einheitliche Meinung darüber herrscht, was es bedeutet, wenn der Risikofaktor in der Vergangenheit gemessen wurde (zum Beispiel Geburtsgewicht) und man nun eine Kohorte von Erwachsenen über Jahre (prospektiv) hinsichtlich des Auftretens einer koronaren Herzerkrankung beobachtet

Ein Beispiel zu den Kohortenstudien :



Hier wurde die Kohortenstudie mit 11 Kohorten durchgeführt und diese Grafik zeigt das Ergebnis des Zusammenhangs zwischen der Radonbelastung und Lungenkrebs.

Je mehr Exposition die Kohorten haben, desto größer wird das Relative Risiko des Lungenkrebs.

(Schwarze Punkte : Kohorte und obere, untere Grenze 95% Konfidenzintervall für RR und lineares Modell eingezeichnet)

▪ Relatives Risiko :

Das relative Risiko (relative risk) gibt an, wie hoch die Wahrscheinlichkeit ist, dass ein Merkmal von zwei Alternativmerkmalen für eine Gruppe von zwei Gruppen vorliegt. Dazu werden die Häufigkeiten in einer Vierfeldertafel angegeben :

	Anzahl Person mit Risikofaktor	Anzahl Person ohne Risikofaktor
Anzahl Erkrankter	a	b
Anzahl Nichterkrankter	c	d

Aus dieser Tabelle kann das relative Risiko berechnet werden :

$$RR = \frac{P(\text{Erkrankung} \mid \text{mit Risikofaktor})}{P(\text{Erkrankung} \mid \text{ohne Risikofaktor})} = \frac{a/a+c}{b/b+d}$$

Es gibt ein Beispiel dafür :

	Anzahl Person, die rauchen	Anzahl Person, die nicht rauchen
Anzahl Person mit Herzinfarkt	130	70
Anzahl Person ohne Herzinfarkt	1870	7930

$$RR = \frac{a/a+c}{b/b+d} = \frac{130/130+1870}{70/70+7930} = 7,4$$

D.h, die Wahrscheinlichkeit für Herzinfarkt bei Rauchern ist 7,4-mal so groß wie bei Nichtrauchern.

▪ Poisson Regression :

Wiederholung :

- Dichte der Poisson Verteilung :

$$f(y_i) = \begin{cases} \frac{\lambda^{y_i} \exp(-\lambda)}{y_i!} & , y_i = 0, 1, 2, \dots \\ 0 & , \text{sonst} \end{cases}$$

- Das generalisierte lineare Modell :

$$f(y_i | \theta_i, \phi_i) = \exp \left(\frac{y_i \theta_i - b(\theta_i)}{\phi_i} + \underbrace{c(y_i, \phi_i)}_{\text{Normierung}} \right)$$

θ_i : natürlicher/kanonischer Parameter

ϕ_i : Dispersionsparameter

$b(), c()$: Funktionen

Bsp.: Normalverteilung, Binomialverteilung,...

$\theta_i = \theta(\mu_i), \mu = \mathbb{E}(y_i | x_i)$

$$f(y_i) = \frac{\lambda_i^{y_i}}{y_i!} \exp(-\lambda_i) = \exp \left(\frac{y_i \log(\lambda_i) - \lambda_i}{1} - \log(y_i!) \right)$$

D.h, die Poisson Verteilung gehört zu einer Exponentialfamilie.

- Poisson Regression

Die Poisson Regression ist ein generalisiertes Regressionsmodell mit Zielgröße y_i . y_i soll hier poisson-verteilt mit dem Parameter λ_i sein. Außerdem ist der Erwartungswert von y_i gegeben den Einflussgrößen x_i $E(y_i|x_i) = \lambda_i$

Annahme : $h(x_i'\beta) = \lambda_i$

z.B Zielgröße wäre cases im folgendes Beispiel und Einflussgröße : personyears, Cigarette-day und years of smoking.

Beispiel :

A cohort of subjects, some non-smokers and others smokers, was observed for several years. The number of cases of cancer of the lung diagnosed among the different categories was recorded. Data regarding the number of years of smoking were also obtained from each individual. For each category the person-years of observation were calculated. The investigators wish to address the question of the relative risks of smoking.

Log cases = $\alpha + \beta_1$ Personyears + β_2 Cigarettes/day + β_3 Years of Smoking

= $-4.669 + 0.00041$ Personyears + 0.0559 Cigarettes/day + 0.0888 Years of Smoking

In diesem Beispiel hat man mit log-lineare poisson Modell verwendet, d.h $h=\exp$.

1. Beispiel : Log-lineare Poisson Regression

Eine gängige Wahl für $h(\cdot)$ ist Exponentialverteilung. Dann gilt :

$$\lambda_i = \exp(x'_i \beta)$$

Also wirken die Kovariablen multiplikativ und exponentiell auf die durchschnittliche Anzahl zu erwartender Ereignisse λ .

2. Beispiel : Lineare Poisson Regression

Eine andere Wahl für $h(\cdot)$ ist die Identität als Responsefunktion.

$$E(y_i | x_i) = \lambda_i = x'_i \beta$$

Dann wirken die Kovariablen additiv auf den bedingten Erwartungswert λ_i .

▪ Cox Regression :

$$h(t) = h_0(t) \cdot \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)$$

Es wird zur Modellierung von Überlebenszeiten in der Survival Analysis benutzt und basiert auf dem Konzept der Hazardrate. $h_0(t)$ bezeichnet dabei eine unbekannte baseline-Hazardfunktion, die im Ausgangsfall ohne Einflüsse (also $x_i = 0$) die zugehörige Hazardfunktion darstellt. X_i ist Einflussvariablen und β ein unbekannter Regressionskoeffizient.

Beispielweise ist Zielgröße Überlebenszeit ab Beginn der Beobachtung und Einflussgröße bestimmte Treatments.

Beispiel : Das Lampenexperiment könnte man noch einmal mit höherer oder niedrigerer Spannung als den bei uns üblichen 220 V durchführen und würde finden, dass die Lebensdauern tendenziell bei höheren Spannungen, also bei Überlastung, kürzer sind und bei niedrigeren Spannungen größer.

▪ Vor- und Nachteil :

Vorteil wäre :

durch die Versuchsanordnung ist die zeitliche Abfolge der Ereignisse klar festgelegt, was Hinweise auf eine mögliche Kausalität geben kann. Und dieser Studientyp ist auch für seltene Exposition geeignet.

Aber nachteilig :

Um einen Zusammenhang zwischen den Risikofaktoren und einem Endpunkt zu finden, benötigt man bei prospektiven Kohortenstudien oft viele Teilnehmer. Das wird irgendwie aufwendig sein. Oft muss man die Studienteilnehmer über lange Zeiträume beobachten. Eine große Anzahl von Studienteilnehmern eventuell über lange Beobachtungszeiträume zu verfolgen, ist nicht nur sehr aufwendig, sondern kann auch sehr teuer werden.

Bei seltenen Outcomes sind Kohortenstudien ineffizient, weil für eine vernünftige statistische Auswertung sehr grosse Kohorten nötig werden.

2. Fall-Kontroll-Studie

▪ Fall-Kontroll-Studie :

Eine Fallkontrollstudie ist eine Form der epidemiologischen Studien. Es handelt sich um eine retrospektive Untersuchung einer Stichprobe, die aus erkrankten Personen besteht (Fall), und einer Stichprobe, die aus gesunden Personen besteht (Kontrolle). Bei beiden Gruppen wird nun ermittelt, ob in der Vergangenheit eine Exposition gegenüber potentiellen Risikofaktoren vorlag. Ein signifikanter Unterschied zwischen beiden Gruppen bedeutet eine Korrelation zwischen Risikofaktor und Erkrankung.

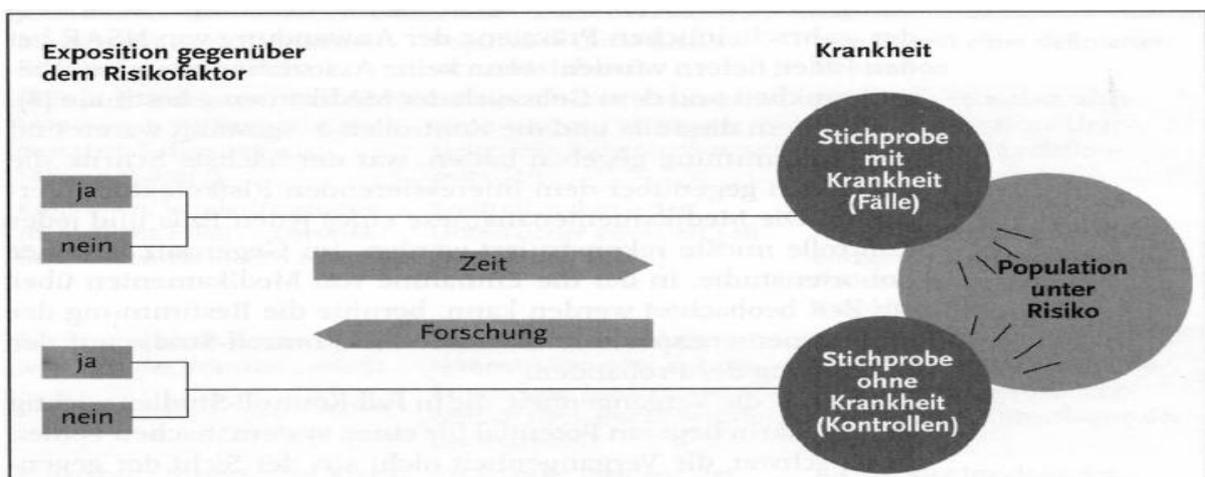


Abb. 10.1: Das Studiendesign von Fall-Kontroll-Studien

- **Fall :**

Für die Untersuchungen wird eine Fallgruppe ausgewählt, für die die interessierenden Merkmale zutreffen, d.h beispielweise wählen wir die Person aus, die Krankheit Tumor haben. (interessierende Merkmal : Tumor) : Einfach Person mit Krankheit.

- **Kontrolle :**

Kontrollen können bekannten oder unbekanntem Studienpopulationen entstammen. Eine bekannte Gruppe besteht aus einer definierten Grundgesamtheit, die über einen bestimmten Zeitraum beobachtet wird, beispielsweise den Passagieren eines Kreuzfahrtschiffes. Wenn die Studiengruppe bekannt ist, so kann als Kontrolle eine aus dieser Population gezogene Stichprobe dienen. Wenn für die Grundgesamtheit keine Namensliste vorliegt, bieten sich Verfahren wie die telefonische Zufallsziffernanwahl (Random Digit Dialling) an. Wenn die Studiengruppe nicht bekannt ist (z.B. Bei Verkehrsunfallopfern in der Notaufnahme, die unter Umständen aus weiter entfernten Orten kommen), bieten sich Kontrollen aus dem Krankenhaus oder aus der Nachbarschaft oder Freunde, Begleitpersonen bzw. Angehörige an.

Folgende sind die Eigenschaften der Kontrolle :

1. Frei von der interessierenden Zielerkrankung
2. Repräsentativ für die Personen, die unter Risiko für die Zielerkrankung stehen
3. Unabhängig von der interessierenden Exposition ausgewählt

▪ Odds-Ratio :

Die Odds ratio (Chancenverhältnis) gibt an, wie hoch die Chance ist, dass ein Merkmal von zwei Alternativmerkmalen für eine Gruppe von zwei Gruppen vorliegt.

Dazu werden die Häufigkeiten, d. h. die Anzahl der Individuen in einer Vierfeldertafel angegeben :

	Anzahl Person mit Risikofaktor	Anzahl Person ohne Risikofaktor
Anzahl Erkrankter(Fall)	a	b
Anzahl Nichterkrankter (Kontrolle)	c	d

Aus dieser Tabelle kann die Odds Ratio berechnet werden :

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

Ein Beispiel :

	Anzahl Person, die rauchen	Anzahl Person, die nicht rauchen
Bronchial- karzinom	1350	7
Kein Bronchial- karzinom	1296	61

$$OR = \frac{a/c}{b/d} = \frac{a*d}{c*b} = \frac{1350/1296}{7/61} = 9,45$$

D.h., die Chance für Person mit Bronchialkarzinom bei Rauchern gegenüber Nichtrauchern ist 9,45-mal so groß wie Person ohne Bronchialkarzinom.

! Man sollte beachten, dass in Fall-Kontroll-Studien nur die Odds Ratio verwendet werden muss. Denn die Anzahl der Kontrollen kann beliebig gewählt werden und wir können somit nicht wissen, welchen Anteil an der Gesamtpopulation die Kontrolle tatsächlich darstellen und. Aus diesem Grund ist es nicht möglich, das relative Risiko zu berechnen.

▪ Das logistische Regression :

Wiederholung :

$$\pi_i = P(Y_i = 1 | x_i) = G(x'_i \beta) = \frac{\exp(x'_i \beta)}{1 + \exp(x'_i \beta)}$$

Unter logistischer Regression oder Logit-Modell versteht man Regressionsanalysen zur (meist multivariaten) Modellierung der Verteilung diskreter abhängiger Variablen.

Zielgröße ist binär und hier x_i ist Vektor der Einflussgröße

Beispielweise kann Zielgröße Krankheit ja/nein und x_i

Kaffeekonsum, Rauchen usw.. Exposition sein.

Beispiel :

Wir interessieren uns jetzt für den Zusammenhang zwischen den Kaffeekonsum – Pankreaskarzinom. Es wurde daher Fall-Kontroll-Studie durchgeführt und als Ergebnis :

		krank	nicht krank	Gesamt
Kaffee- Konsum (pro Tag)	0	20	88	108
	1-2	153	271	424
	Mehr	194	284	478
	Gesamt	367	643	1010

In R :

Coefficients:

```

      Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.4816     0.2477  -5.981 2.22e-09 ***
Kaffee_1-2   0.9099     0.2676   3.401 0.000672 ***
Kaffee_Mehr  1.1005     0.2646   4.158 3.21e-05 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Was man zunächst erkennen kann, sind alle hochsignifikant. Und wenn man den Koeffizienten schaut, kann man so interpretieren, dass sich die Chance für Erkrankten gegenüber Nichterkrankten bei Kaffeekonsum 1-2 um den Faktor 2,4841 erhöht. Analog wie bei Kaffee Mehr steigt die Chance um den Faktor 3,0057.

▪ Vor- und Nachteil :

- Vorteil : Fall-Kontroll-Studien sind nicht sehr aufwendiger und deshalb kostengünstiger als Kohortenstudie. Zudem benötigen sie wenig Zeit. In der gleichen Studie können mehrere Faktoren untersucht werden.
- Nachteil : Der Informationsgewinn aus Fall-Kontroll-Studien ist beschränkt. So können z.B keine Prävalenz- und Inzidenz-Schätzungen ermittelt werden, weil das Verhältnis von Fällen zu Kontrollen durch die Anzahl Personen, die in die Studie eingeschlossen werden, bestimmt wird und nicht durch die Häufigkeit der Fälle in der Bevölkerung. Die Auswahl der Kontrollgruppe ist oft problematisch und die zeitliche Abfolge von Exposition und Erkrankung unklar.

[Literaturverzeichnis]

1. Auswertung einer Brustkrebsstudie mit Hilfe von logistischen Regressionsmodellen -Diplomarbeit von Marco Grzegorzcyk- 13p-14p

2. Fall-Kontroll-Studie zur Bewertung von beruflichen Faktoren im Zusammenhang mit Gonarthrosen – die ArGon Studie
(A. Klußmann, Hj. Gebhardt, M. Nubling, L. V. von Engelhardt, E. Quiros Perea, F. Liebers, B. Bouillon, M. A. Rieger) 35p – 40p

3. Figur2 : rimes DA, Schulz KF: Cohort studies: marching towards outcomes Lancet 2002;359:341

4. Poisson-Verteilung : Aus dem Vorlesungsskript Lineare Modell bei Prof.Helmut Küchenhoff 55p~

5. Was ist eine «Odds Ratio» – und wann wird sie verwendet?
Ulrike Held Horten-Zentrum, UniversitätsSpital, Zürich 634p-635p

6. Abb.10.1 : Methodik der Klinischen Epidemiologie Fall-Kontroll-Studie Fletcher, Fletcher & Wagner 1999 4p

7. Methodik der Klinischen Epidemiologie Kohortenstudie Fletcher, Fletcher & Wagner 1999 4p

8. Verglichen womit? Kontrollen für Fall-Kontroll-Studien finden. Z ärztl Fortbild Qual Gesundh.wes 2006; 100:209p–215p

9. Case-control studies, research in reverse. Lancet 2002; 359(9304):431p-434p

10. Cox Modell -Marcel Noack- 3p-6p

11. LMU Vorlesungsskript, Epidemiologie 1 und Deskriptive Epidemiologie -E. Grill, R. von Kries, M Wildner, A. Crispin, A Peters, U. Siebert - 3p-5p 15p- 17p

12. Tilo Görl, Grundseminar: Grundlegende Modelle der multivariaten Datenanalyse 1p-4p

13. Einführung in die Poisson Regression von Margret-Ruth Oelker 11p-16p

14. Logistische Regression Diplomarbeit von Natali Gerschmann und Anke Wiegmann 3p-7p

15. Epidemiologie wikipedia auf Google

Umweltepide miologie Medizinlexikon auf Google

16. Vorlesungsskript : Statistische Methoden in der Epidemiologie (16468 / 16469) (Mansmann, Adrion, Jurinovic) WiSe 2011/2012
Folie 1 – 1p, 11p-13p, Folie 3 – 1p , Übungsblätter 7 – Aufgabe2
Ü4 – Aufgabe1

17. Wichtige epidemiologische Studientypen – Artikel Nr. 18 der Statistik-Serie in der DMW - 1p-4p

18. Projekt Neue Statistik 2003 - Lernmodul: Cox-Regression

19. Epidemiologie II Der Begriff Risiko 5p-9p

20. Simulation von Überlebenszeiten mit Hilfe von SAS® 1p-4p

21. **Kohortenstudien in Deutschland** *Michael Wagner* 2p-7p

