

# Assoziationsstudien

Prof. DI Dr. Andrea Berghold  
Institut für Med. Informatik, Statistik und Dokumentation

- ◆ Ziele von Assoziationsstudien
- ◆ Design von Assoziationsstudien
- ◆ Statistische Analysemethoden
- ◆ Interpretation der Ergebnisse

## *Ziele von Assoziationsstudien*

---

- ◆ Beurteilung, ob ein Allel eines Kandidatengens mit einer Erkrankung assoziiert ist
- ◆ Abschätzung des Erkrankungsrisikos für Träger

## *Definition*

- ◆ Assoziation ist vorhanden, wenn das Erkrankungsrisiko für Träger und Nicht-Träger des Allels nicht gleich ist.
- ◆  $P(\text{Erkrankung} \ \& \ A) \neq P(\text{Erkrankung}) \times P(A)$ 
  - Beispiel: Magengeschwür ist assoziiert mit Blutgruppe A

- ◆ **Fall-Kontrollstudie**

Allelhäufigkeit bei Fällen (erkrankte Personen)  
vs. Kontrollen (nicht erkrankte Personen)  
Auswahl der Personen nach Krankheitsstatus

- ◆ **Kohortenstudie (Follow-up Studie)**

Häufigkeit der Erkrankung bei Trägern vs.  
Nicht-Trägern  
Auswahl der Personen nach Exposition

# *Kohortenstudie - Relatives Risiko*

Kandidatengen Mutation	Krankheit		Summe
	ja	nein	
vorhanden	a	b	a+b
nicht vorhanden	c	d	c+d

$$RR = \frac{a / (a+b)}{c / (c+d)}$$

$$RR = \frac{\text{Inzidenz der Exponierten}}{\text{Inzidenz der Nicht Exponierten}}$$

# *Relatives Risiko*

Es gibt den multiplikativen Faktor an, um den sich die Erkrankungswahrscheinlichkeit erhöht, wenn man einer definierten Exposition ausgesetzt ist –

Stärke des Zusammenhangs zwischen Krankheit und Exposition

**RR=1**      Risiken identisch

**RR > 1**      Risiko der Exponierten ist größer als Risiko der Nicht-Exponierten (Risikofaktor)

**RR < 1**      Risiko der Exponierten ist kleiner als Risiko der Nicht-Exponierten (protektiv)

## ◆ Vorteile

- direkte Schätzbarkeit der Risiken
- natürlicher Verlauf einer Erkrankung beobachtbar
- Wirkung der Exposition auf verschiedene Erkrankungen untersuchen
- weniger anfällig für Bias

## ◆ Nachteile

- zeit- und kostenintensiv
- nicht geeignet für seltene Erkrankungen
- Follow-up Verluste

# *Fall-Kontrollstudie*

## ◆ Vorteile:

- kurze Studiendauer, kostengünstig
- Untersuchung seltener Krankheiten, Krankheiten mit langer Latenzzeit
- Untersuchung mehrerer Risikofaktoren für eine einzige Krankheit

## ◆ Nachteile:

- anfällig für systematische Fehler

## *Fall-Kontrollstudie - Odds Ratio*

Allel	Krankheit	
	ja (Fälle)	nein (Kontrollen)
A1	a	b
A2	c	d

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

$$OR = \frac{\text{Chance, dass ein Fall Allel A1 hat}}{\text{Chance, dass eine Kontrolle Allel A1 hat}}$$

## *Relatives Risiko - Odds Ratio*

---

- ◆ Im Gegensatz zu Kohortenstudien kann bei Fall-Kontroll Studien das relative Risiko *nicht direkt* berechnet werden
- ◆ Das Odds Ratio einer Fall-Kontroll Studie kann als **Schätzwert (Näherung)** für das relative Risiko verwendet werden, wenn das absolute Krankheitsrisiko *gering* ist (dh. die untersuchte Krankheit *selten* auftritt).

# *Beispiele*

Krankheit	Relatives Risiko bei einem bestimmten HLA Typ
Diabetes Typ-1	10
Rheumatische Athritis	4
Multiple Sklerose	3

## Beispiel – Fall Kontrollstudie

Glucagon Rezeptor Gen - Mutation	Krankheit		Summe
	NIDDM	Kontrollen	
vorhanden	10 = a	10 = b	20 = (a+b)
nicht vorhanden	206 = c	1112 = d	1318 = (c+d)
Summe	216 = (a+c)	1122 = (b+d)	1338
Trägerhäufigkeit	4,63%	0,89%	

$$OR = \frac{10 / 206}{10 / 1112} = 5,4 \text{ (95\% CI: 2,2 – 13,1)}$$

## *Konfidenzintervall*

---

Um Anhaltspunkte für die **Genauigkeit** des Schätzers zu gewinnen, konstruiert man aus den Daten der Stichprobe ein sogenanntes  **$(1-\alpha)$ -Konfidenzintervall**, welches mit einer vorgegebenen Wahrscheinlichkeit (z.B. 95%) den wahren (unbekannten) Wert überdeckt.

# *Analysemethode - Chi-Quadrat Test*

Kann der Zusammenhang rein zufällig entstanden sein  
(d.h. Mutation erhöht das Risiko zu erkranken nicht)

oder

können wir davon ausgehen, dass es einen tatsächlichen  
Zusammenhang gibt.

Kontrollmaßnahme:      Statistischer Test / Konfidenzintervall

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

O: observed (beobachtete Werte); E: expected (erwartete Werte)

## Beispiel: Chi-Quadrat Test

Mutation	Krankheit		Summe
	NIDDM	Kontrollen	
vorhanden	3,2		20
nicht vorhanden			1218
Summe	216	1122	1338

Erwartete Werte unter der Nullhypothese (es gibt keinen Zusammenhang, es besteht Unabhängigkeit):

$$1. \text{ Zelle: } 216 * 20 / 1338 = 3,2$$

## Beispiel: Chi-Quadrat Test

Mutation	Krankheit		Summe
	NIIDM	Kontrollen	
vorhanden	a	b	a+b
nicht vorhanden	c	d	c+d
Summe	a+c	b+d	n

Erwartete Werte:

1. Zelle (a):  $(a+c) * (a+b) / n$
2. Zelle (b):  $(b+d) * (a+b) / n$
3. Zelle (c):  $(a+c) * (c+d) / n$
4. Zelle (d):  $(b+d) * (c+d) / n$

## Beispiel: Chi-Quadrat Test

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

O: observed (beobachtete Werte); E: expected (erwartete Werte)

In unserem Fall:

$$\chi^2 = \frac{(10 - 3,2)^2}{3,2} + \frac{(10 - 16,8)^2}{16,8} + \frac{(206 - 212,8)^2}{212,8} + \frac{(1112 - 1105,2)^2}{1105,2} = 17,46$$

$$p=0,0004$$

Kritischer Wert von  $\chi^2$  bei 2x2 Tafeln: 3,84 bei  $\alpha=0,05$

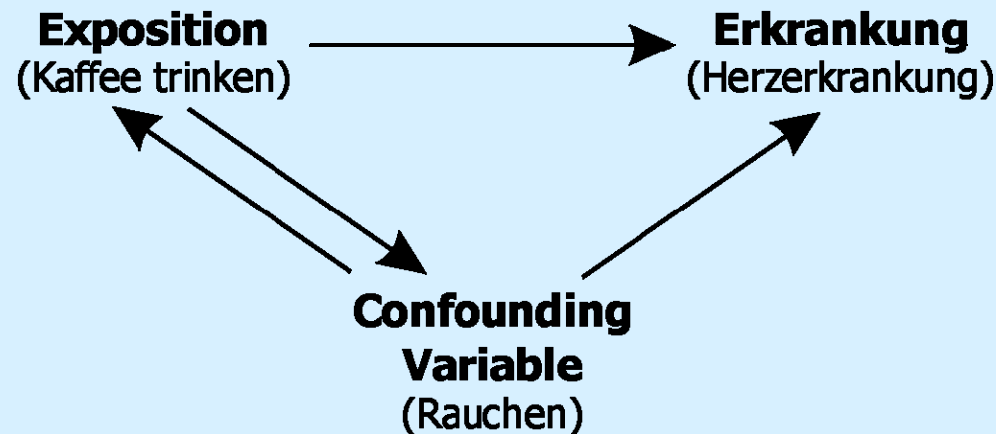
# Analyse der Daten

Allel	Fälle	Kontrollen
A1	$a=2n_{11}+n_{21}$	b
A2	c	d

Genotyp	Fälle	Kontrollen	Summe
A1A1	$n_{11}$	$n_{12}$	$n_{1*}$
A1A2	$n_{21}$	$n_{22}$	$n_{2*}$
A2A2	$n_{31}$	$n_{32}$	$n_{3*}$
Summe	$n_{*1}$	$n_{*2}$	$n_{**}$

# Assoziation

- ◆ In einer Studie gefundene Assoziation muss nicht darauf beruhen, dass das Allel einen Risikofaktor für die Erkrankung darstellt, sondern kann z.B. durch unterschiedliche ethnische Herkunft von Fällen und Kontrollen oder Selektionsdrift entstehen (Confounding!)



# *Interpretation der Ergebnisse*

---

- ◆ Kausale Beziehung
- ◆ Linkage Disequilibrium
- ◆ Populationsstratifikation
- ◆ Zufall (Typ I Fehler)

- ◆ Man benötigt biologische Evidenz:
  - Morphologische Studien
  - Funktionelle Studien (Tiere, Zellkulturen)
- ◆ Weitere genetisch-epidemiologische Studien
  - Replikation der Resultate in anderen Populationen

## *Beispiel – UK Sample*

Glucagon Rezeptor Gen – Mutation	Krankheit		Summe
	NIDDM	Kontrollen	
vorhanden	15 = a	1 = b	16 = (a+b)
nicht vorhanden	676 = c	424 = d	1100 = (c+d)
Summe	691 = (a+c)	425 = (b+d)	1116
Trägerhäufigkeit	2,2%	0,2%	

## Beispiel – UK Sample

Glucagon Rezeptor Gen – Mutation	Krankheit		Summe
	NIDDM	Kontrollen	
vorhanden	15 = a	1 = b	16 = (a+b)
nicht vorhanden	676 = c	424 = d	1100 = (c+d)
Summe	691 = (a+c)	425 = (b+d)	1116
Trägerhäufigkeit	2,2%	0,2%	

$$OR = \frac{15 / 676}{1 / 424} = 9,4 \text{ (95\% CI: 1,2 – 71,5)}$$

## *Beispiel – UK sample*

$$\chi^2 = \frac{(15 - 9,9)^2}{9,9} + \frac{(1 - 6,1)^2}{6,1} + \frac{(676 - 681,1)^2}{681,1} + \frac{(424 - 418,9)^2}{418,9} = 6,98$$

$$p=0,008$$

Kritischer Wert von  $\chi^2$  bei 2x2 Tafeln: 3,84 bei  $\alpha=0,05$

## *Linkage Disequilibrium*

---

- ◆ Eine Assoziation liegt auch vor, wenn Allel im Kopplungsungleichgewicht mit dem eigentlich krankheitsbeitragenden Mutation steht.

# Populationsstratifikation

Annahme:

- 2 gleich große Strata - Subpopulationen
- keine Krankheits-Allel Assoziation

Stratum	1	2
P (Erkrankung)	0,9	0,1
P (A)	0,8	0,3
P (Erkrankung & A)	0,72	0,03

Gesamtbevölkerung:

$$P (\text{Erkrankung}) = 0,5$$

$$P (A) = 0,55$$

$$P (\text{Erkrankung} \& A) = 0,375 \neq P (\text{Erkrankung}) \times P (A)$$

# Beispiel: Diab. Mellitus und Gm Haplotyp

## Gesamt

		DM	
		+	-
Gm	+	23	270
	-	1343	3284

Rohe OR = 0.21 (0.14 – 0.32)

## Halindianische Abstammung

		DM	
		+	-
Gm	+	7	8
	-	138	109

OR = 0.69 (0.24 – 1.96)

## Keine indianische Abstammung

		DM	
		+	-
Gm	+	3	2
	-	18	9

OR = 0.75 (0.11 – 5.32)

## Vollindianische Abstammung

		DM	
		+	-
Gm	+	12	57
	-	1048	3147

OR = 0.63 (0.34 – 1.18)

# *Assoziations-Analyse*

---

Populationsstratifikation kann nur auftreten, wenn Populationen sich sowohl hinsichtlich

- der Allelhäufigkeit
- des Erkrankungsrisikos

unterscheiden

◆ Testen auf "Hardy Weinberg Gleichgewicht"

Pritchard & Rosenberg (1999)

# *Hardy Weinberg-Gleichgewicht*

Bei 2 Allelen mit den Frequenzen  $p$  für das Allel  $A_1$  und  $q$  für das Allel  $A_2$  beträgt die Summe der Allelfrequenzen  $p + q = 1$ .

◆ Daher sind:

- Homozygote  $A_1A_1 = p^2$
- Heterozygote  $A_1A_2 = 2pq$
- Homozygote  $A_2A_2 = q^2$

◆ Daraus ergibt sich:

$$p^2 + 2pq + q^2 = 1$$

# *Assoziations-Analyse*

---

## ◆ Umgang mit Confounding

- Beim Studiendesign

Sorgfältige Auswahl der Kontrollen, matching

Familienbasierte Kontrollen

- Bei der Datenanalyse

Stratifikation

Multivariate Analyse (z.B. log. Regression)

# *Assoziations-Analyse*

---

## ◆ Familienbasierte Kontrollen

- Eltern-Kind: zur erkrankten Person werden die gesunden Eltern als Kontrolle herangezogen
- Diskordante Zwillingspaare: Erkranktes und gesundes Geschwister

Problem: schwierigere Probandenrekrutierung

## *Stat. Konzepte bei Planung*

---

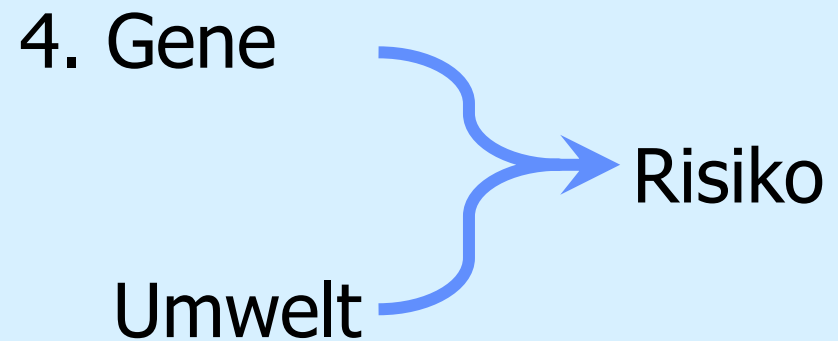
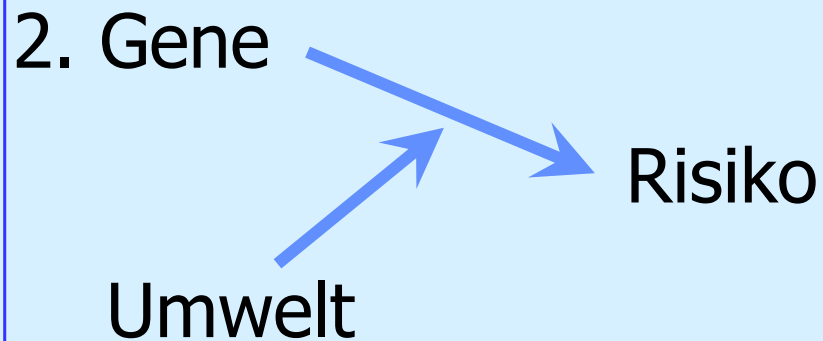
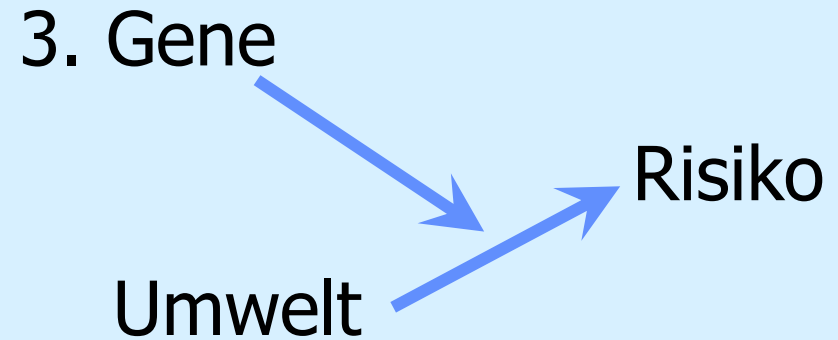
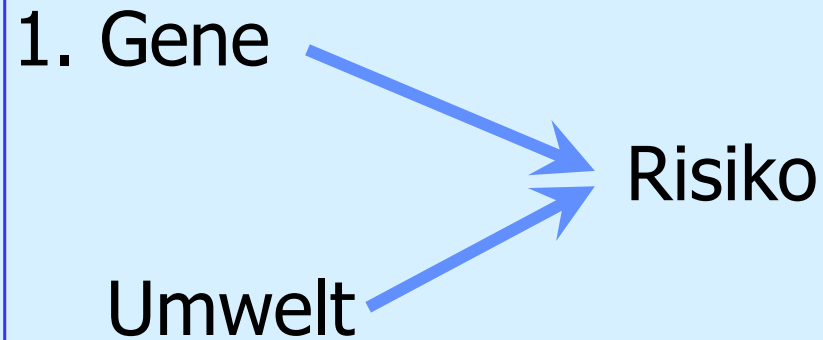
- ◆ Power der Studie  
Erkennen eines bedeutsamen Effektes  
d.h. Wahrscheinlichkeit für korrektes  
Verwerfen der Nullhypothese soll groß sein

## *Erweiterungen, Probleme*

---

- ◆ Jedes einzelne Gen hat nur einen kleinen Einfluss
- ◆ Mehrere Gene müssen zusammenwirken (schwer zu entdecken)
- ◆ Gen-Umwelt Interaktionen  
Stratifizierung, multivariate Analysen

# *Gen-Umwelt Interaktionen*



	<b>Allel-Assoziation</b>	<b>Linkage</b>
<b>Definition</b>	2 Allele, die häufiger gemeinsam auftreten als man durch Zufall erwarten würde	Die Loci zweier Gene liegen am selben Chromosom nahe beieinander
<b>Beziehung zwischen</b>	Allelen oder Allel und Krankheit	Loci
<b>Implikation</b>	Möglicherweise kausal	Keine Implikation für Kausalität
<b>Getestet in</b>	Population	Familien