

# Statistische Probleme in epidemiologischen Studien

## Krebs und Mobilfunk – Was sagt die Epidemiologie?

Informationsveranstaltung FGF Deutschland und FMK Österreich, Wien, 2. Dezember, 2008

Hagen Scherb

Institute of Biomathematics and Biometry

# Inhalt – Statistische Probleme in epidemiologischen Studien

---

- Spektrum statistischer Probleme in epidemiologischen Studien
  - Theoretische Grundlagen
  - Daten-Analyse – Techniken
  - Elemente des statistischen Schließens (Inferenz)
  
- Hypothesen in der Epidemiologie
  - Hypothesenart: qualitativ vs. quantitativ
  - Null- vs. Alternativhypothese
  - Problemfall** nichtsignifikantes Resultat
  
- **INTERPHONE** Fazit aus statistischer Sicht

# Das Spektrum statistischer Probleme in epidemiologischen Studien

- **Theoretische Grundlagen** Erkenntnistheorie, Erkenntnislogik
- **Daten-Analyse – Techniken** deskriptiv, explorativ, analytisch
- **Elemente der statistischen Inferenz** Definition und konkrete Berechnung

# Theoretische Grundlagen

---

- **Induktive Logik**      100 Lungenkrebspatienten      *Fälle*  
100 vergleichbare Gesunde      *Kontrollen*  
Mehr Raucher unter den Erkrankten  
**Hypothese** *Rauchen erhöht Risiko für Lungenkrebs*
- **Iteration**      Hypothesen-Evolution mit Vielzahl von  
Beobachtungsstudien → *„sicheres Wissen“*
- **Hypothesen-Logik**      Hypothesentests: Nullhypothese, Alternativhypothese  
(J Neyman, E Pearson, ab 1920er )
- **Spezialthema**      *Beweis der Nullhypothese* (ID Bross, Biometrics 1985,  
*Why proof of safety is much more difficult than proof  
of hazard*)
- **Frequentist / Bayes**      Relative Häufigkeit / Update von Wahrscheinlichkeiten  
für Hypothesen

# Daten-Analyse – Techniken

---

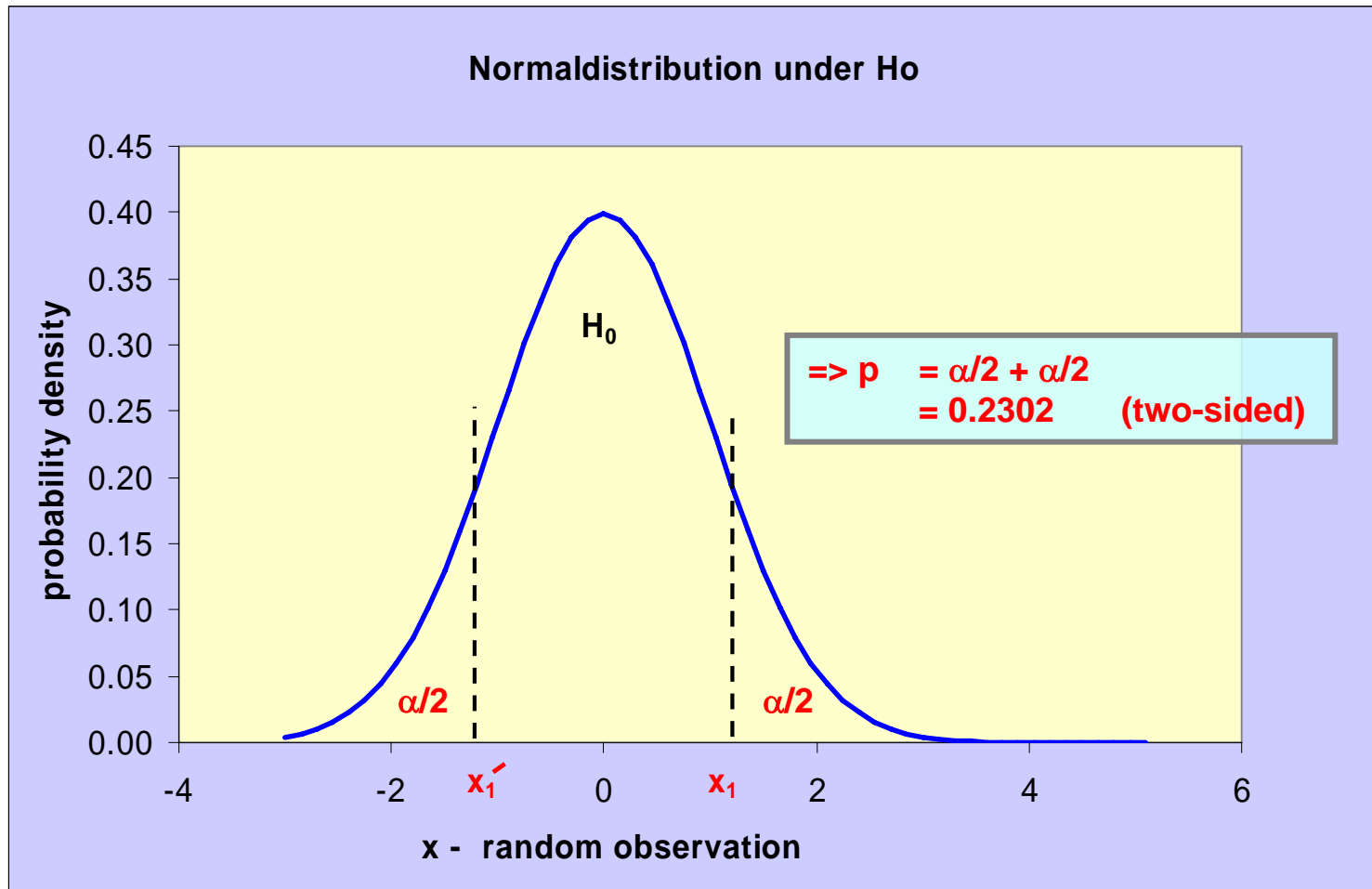
- **V**arianzanalyse
- **R**egression (multiple)
- **V+R** → Kovarianzanalyse
- Diskriminanzanalyse
- Faktorenanalyse
- **Logistische Regression – bedingte logistische Regression**
- **Modellierung in der Regression: Interaktion und Confounding**

# Elemente der statistischen Inferenz

---

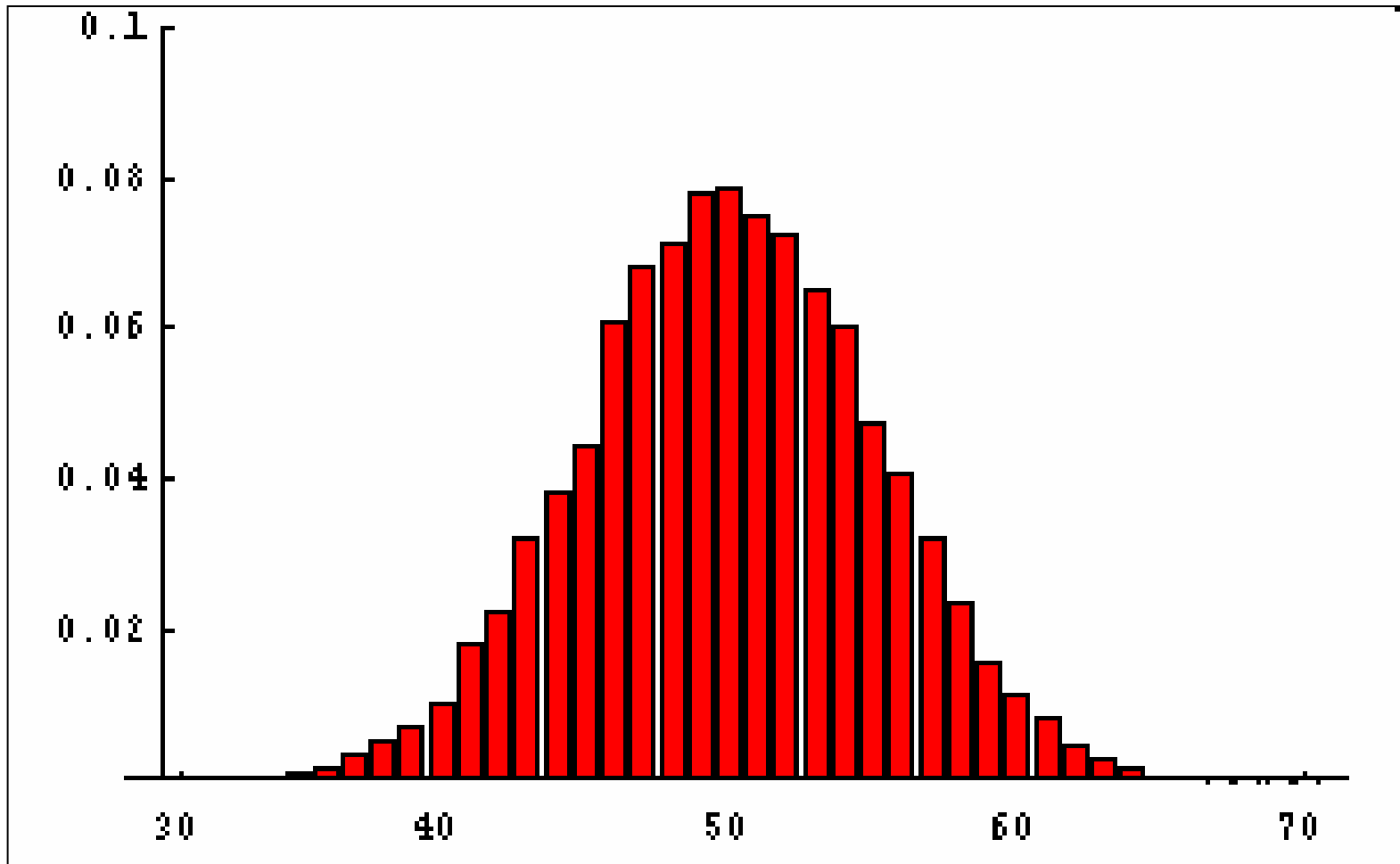
- **Zufall** fairer Zufall muss stets unterstellt werden
- **Effektschätzer** Risiko, relatives Risiko, odds ratio
- **Statistische Tests** Fehler 1. Art,  $\alpha$ -Niveau, meist  $\alpha=0.05$  (=5%)
- **p-Werte** *empirisches Signifikanzniveau* ( $\alpha_{\text{observed}}$ )
- **Konfidenzintervalle** zum Niveau  $\alpha$ ; Bereich plausibler Parameter
- **Konfidenzkurven** Omnibusdarstellung für Schätzer, Tests, p-Werte  
Konfidenzintervalle
- **Statistische Power** Fehler 2. Art,  $\beta$ -Niveau, Testgüte/-Stärke

# Elemente der statistischen Inferenz p-Wert



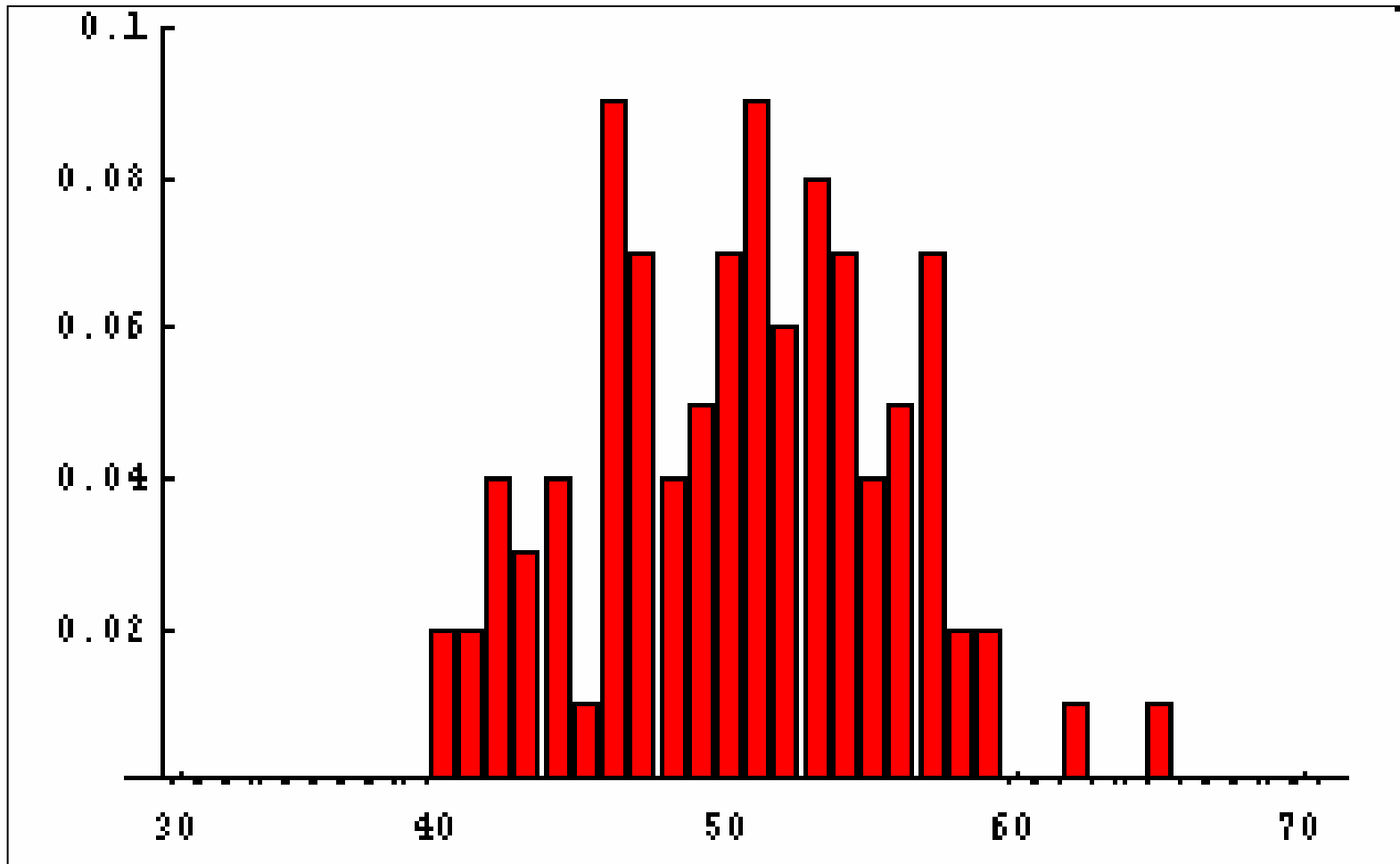
# Elemente der statistischen Inferenz **p-Wert**

100-facher Münzwurf: Anzahl Kopf (n=100, Anzahl=10000)



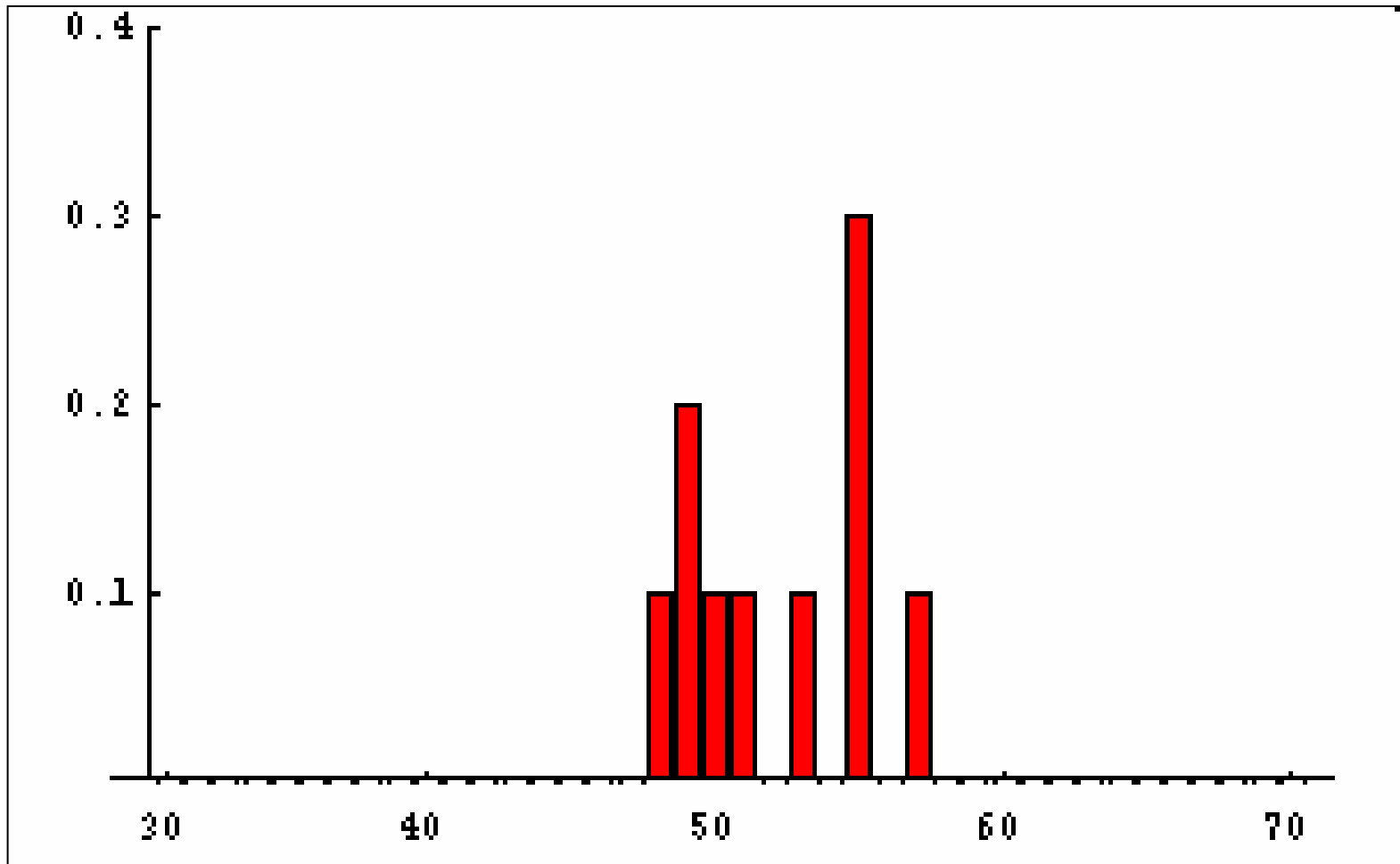
# Elemente der statistischen Inferenz **p-Wert**

100-facher Münzwurf: Anzahl Kopf (n=100, Anzahl=100)

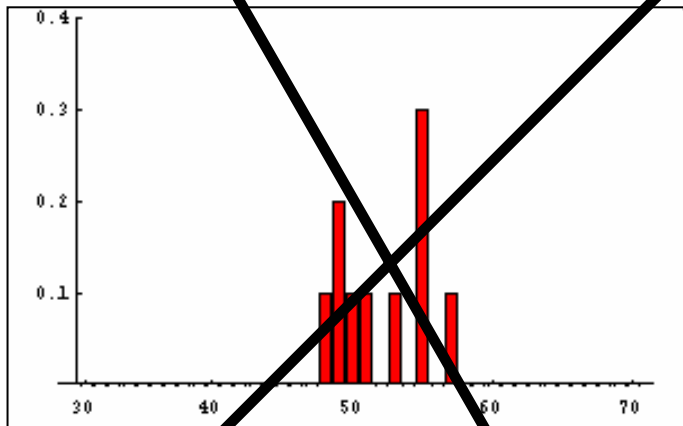


# Elemente der statistischen Inferenz **p-Wert**

100-facher Münzwurf: Anzahl Kopf (n=100, Anzahl=10)



## 100-facher Münzwurf: Anzahl Kopf (n=100, Anzahl=1)



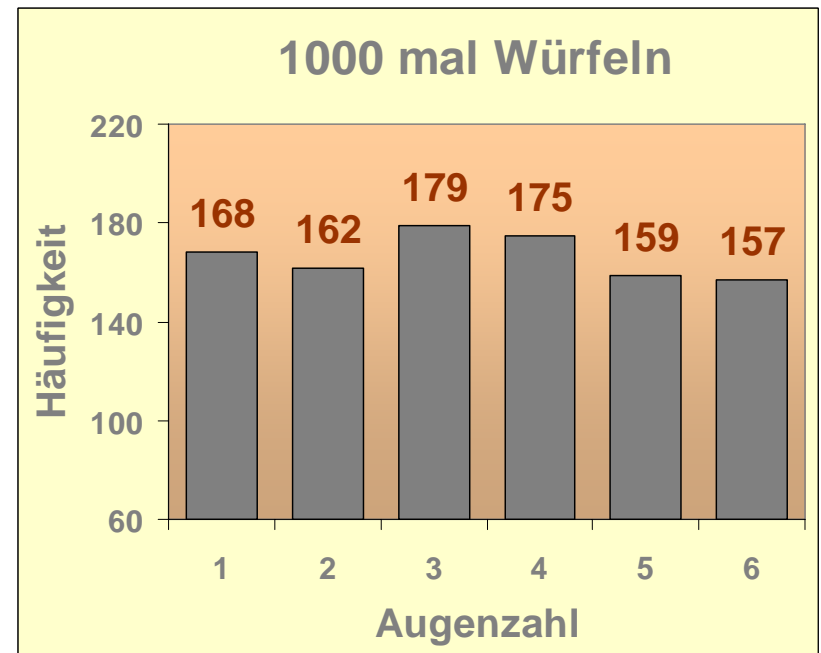
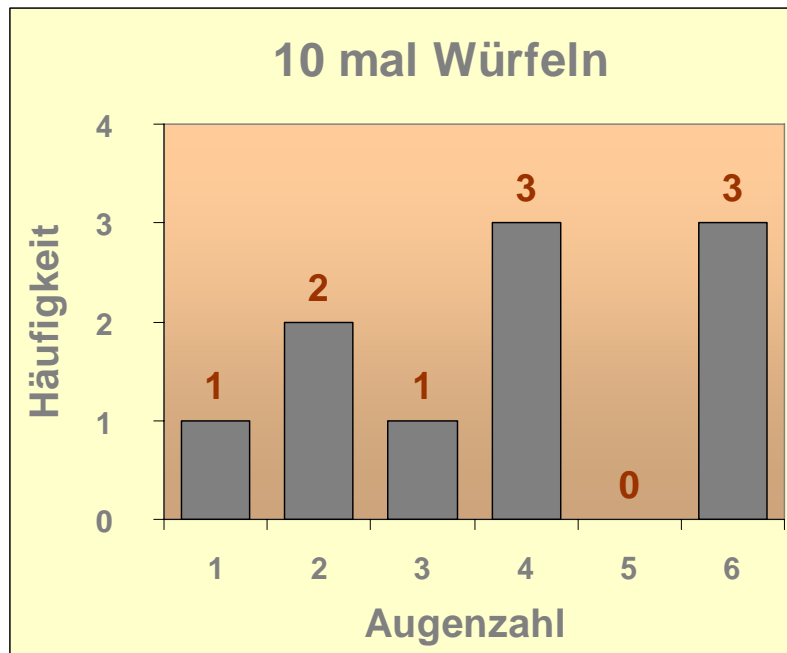
➤ In der Epidemiologie liefert jede Studie i. a. nur **1 Ergebnis**

➤ Beispiel: Anzahl Kopf = 63  
=>  $p = 0.0093$  (UMPUT)

➤ **Münze unfair**

# Anwendung in der Epidemiologie **Fallzahlschätzung**

- Das **Gesetz der großen Zahl** besagt im Prinzip, daß mit zunehmender Stichprobengröße die prozentuale Differenz zwischen tatsächlichen und erwarteten Werten sich verringert
- Beispiel: Würfel



# Anwendung in der Epidemiologie **Fallzahlschätzung**

---

- **Gute Nachricht** Es gibt eine gut entwickelte Methodik und brauchbare Software zur Fallzahlschätzung in der Epidemiologie
- **Schlechte Nachricht** In aller Regel werden in der Epidemiologie aber keine Fallzahlschätzungen durchgeführt, z.B. INTERPHONE Einzelländerstudien und „regular use“. Trotzdem werden negative Resultate ohne Power-Angabe publiziert

INTERPHONE weist die Power nur für die kombinierten Analysen und nur für 5+ Jahre Exposition mit OR=1.5 als ausreichend aus. Für den liberal definierten „regular use“ wird deshalb die Power nicht ausreichen. (Study protocol p. 2, siehe: <http://www.iarc.fr/en/content/download/2212/18242/file/INTERPHONEStudyProtocol.pdf>).

# Anwendung in der Epidemiologie **Fallzahlschätzung**

## ➤ Beispiel **INTERPHONE**

➤ Kein spezifisches Vorwissen => Power-Schätzung spekulativ – „politisch“

Tumour	Case-control ratio	Number of study subjects		Estimated power (%) Mobile phone use	
		Cases	Controls	5 years	8+ years
Brain tumours	1:1	7420	7420	100	100
Acoustic neurinoma	1:2	1080	2415	99.3	87.2
Parotid gland tumour	1:3	796	2175	98.7	74.8
Acute Leukaemia	1:3	1298	3294	99.7	85.1

➤ Power für „regulären Gebrauch“ und für einzelne Länder unspezifiziert

➤ Falls Power=100% => Problem: „**Messinstrument am Anschlag**“

# Anwendung in der Epidemiologie Fallzahlschätzung

- **Beispiel INTERPHONE** Tumor der Ohrspeicheldrüse  
Power im study protocol kann verifiziert werden

The screenshot shows the 'Power and Sample Size Program: Main Window' with the 'Dichotomous' tab selected. The 'Output' section displays 'Power' as 0.9662. The 'Design' section is set to 'Independent', 'Retrospective', 'Odds ratio', and 'Uncorrected chi-square or Fisher's exact test'. The 'Input' section shows  $\alpha = 0.05$ ,  $p_0 = 0.15$ ,  $n = 796$ ,  $m = 3$ , and  $\psi = 1.5$ . Buttons for 'Calculate', 'Graphs', and 'Exit' are visible.

The screenshot shows the 'Power and Sample Size Program: Main Window' with the 'Dichotomous' tab selected. The 'Output' section displays 'Power' as 0.7983. The 'Design' section is set to 'Independent', 'Retrospective', 'Odds ratio', and 'Uncorrected chi-square or Fisher's exact test'. The 'Input' section shows  $\alpha = 0.05$ ,  $p_0 = 0.07$ ,  $n = 796$ ,  $m = 3$ , and  $\psi = 1.5$ . Buttons for 'Calculate', 'Graphs', and 'Exit' are visible.

Dupont WD, Plummer WD, Power and Sample Size Calculations: A Review and Computer Program, Controlled Clinical Trials 1990; 11:116-28

# Anwendung in der Epidemiologie **Fallzahlschätzung**

- **Beispiel INTERPHONE** **Power-Reduktion** durch Korrelation der Matching Variablen mit der Exposition (Ohrspeicheldrüse)

Power and Sample Size Program: Main Window

File Log Help

Survival t-test Regression 1 Regression 2 **Dichotomous** Log

Output [Studies that are analysed by chi-square or Fisher's exact test](#)

[What do you want to know?](#) Power

[Power](#) 0.5535

Design

[Matched or Independent?](#) Matched or paired

[Case control?](#) Retrospective

Input

$\alpha$  0.05  $p_0$  0.07

$n$  796  $m$  3

$\phi$  0.5  $\psi$  1.5

Calculate

Graphs

Logging is enabled.

Exit

- **Problem beim „Matching“**  
Die Exposition (z.B. *regular use*) ist in gepaarten (Gruppen von) Studienteilnehmern korreliert (*patient responses within pairs or matched groups are correlated*)
- **Beispiel**  
Matching nach Geschlecht und Handynutzung überwiegend bei Männern

# Anwendung in der Epidemiologie **Fallzahlschätzung**

- **Beispiel INTERPHONE** Weitere Power-Reduktion durch kleine Fallzahl auf Einzelländer-Ebene (Ohrspeicheldrüse)

Power and Sample Size Program: Main Window

File Log Help

Survival t-test Regression 1 Regression 2 **Dichotomous** Log

Output [Studies that are analysed by chi-square or Fisher's exact test](#)

[What do you want to know?](#) Power

[Power](#) .4463

Design

[Matched or Independent?](#) Matched or paired

[Case control?](#) Retrospective

Input

$\alpha$  0.05  $p_0$  0.07

$n$  600  $m$  3

$\phi$  0.5

$\psi$  1.5

Calculate

Graphs

Logging is enabled.

Exit

- Power kann z.B. auf unter 50% sinken; bei nur ca. 600 Fällen in Israel+Dänemark+Schweden
- Bei unter 50% Power ist eine (epidemiologische) Studie nicht mehr sinnvoll!

# Anwendung in der Epidemiologie **spezifischere Fallzahlschätzung**

- **Beispiel** Kinderkrebs um Kernkraftwerke (KiKK-Studie) in Deutschland, mehr und genauere Information war vorhanden

## Statistical power for the German childhood leukemia study (1400 cases and 4200 controls)

Method	Uniform population density	Reduced population density near NPPs
2x2-table distance " $\leq$ or $>$ 5 km"	0.69	0.51
t-Test Mean case/control distances	0.65	0.48
<b>Logistic regression</b> <b>Reciprocal distance law</b>	0.91	0.77
Logistic regression Step function (robust)	0.75	0.58

# Hypothesen in der Epidemiologie

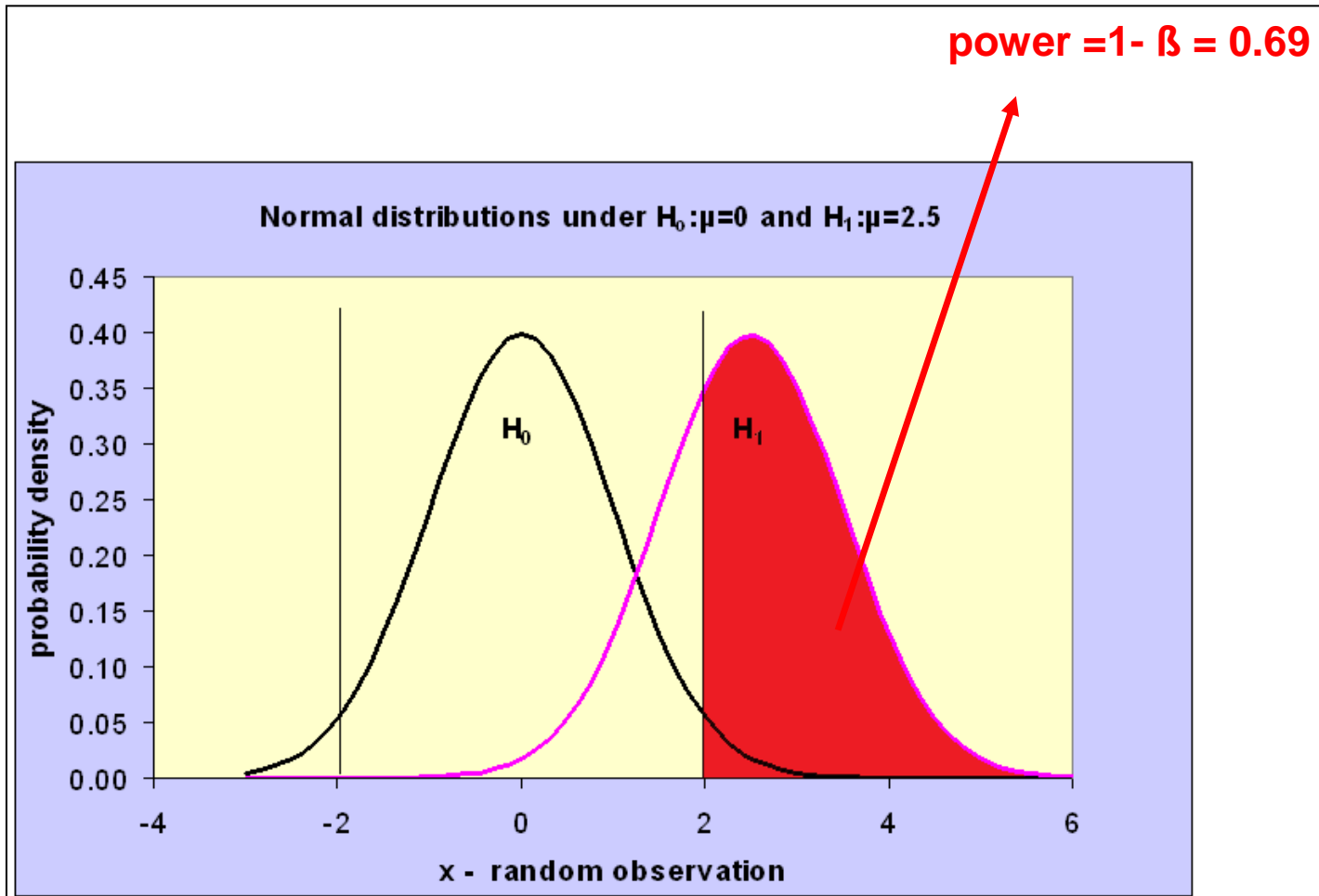
---

- **qualitativ**  
Handy *ja – nein*  
KKW *< 5 km – > 5 km Wohn-Abstand*
- **quantitativ**  
Handy *Gebrauchintensität*  
KKW *Wohnabstand vom KKW*
- **Power**  
i.a. quantitativ > qualitativ (mehr Information)  
z.B. KiKK-Studie
- **Hypothesentausch**  
Möchte man „Sicherheit nachweisen“, ist das mit dem Rekurs auf „nicht-signifikante“ Resultate i.a. nicht sinnvoll bzw. möglich. Man müsste Null- und Alternativhypothesen vertauschen, was meistens unrealistisch bzw. unmöglich wäre.

# Hypothesen in der Epidemiologie Null- vs. Alternativhypothese

## ➤ Vorstellung

Wahrscheinlichkeitsdichten



# Hypothesen in der Epidemiologie **Beweis der Nullhypothese**

---

- **Historisch** Eine Alternativhypothese wurde in den Anfängen der schließenden bzw. mathematischen Statistik nicht formuliert
- **Regel** Die Aussage, die Alternativhypothese  $H_1$  sei statistisch gesichert zum Niveau  $\alpha$ , ist analog für  $H_0$  ohne Berechnung der Power nicht möglich  
  
Will man vermöge eines Signifikanztests die Gültigkeit einer Hypothese statistisch sichern, so hat man sie stets als  $H_1$  zu formulieren (z. B. H. Witting, 1974, Teubner)  
Stichwort: Bioäquivalenz-Studien
- **Hinweis** *Absence of evidence is not evidence of absence* s. z.B. <http://www.bmj.com/cgi/content/full/311/7003/485>

➤ **Beispiel: „Entwarnung“**

*Die bereits vorliegenden epidemiologischen Studien zeigen, dass bei weniger als 10 Jahren Nutzung kein Risiko für Tumorerkrankungen, insbesondere von Hirntumoren, bei Erwachsenen durch die Nutzung von Mobiltelefonen besteht.*

➤ Diese Aussage ist auf Basis der vorliegenden epidemiologischen Studien inhaltlich nicht sinnvoll und formal nicht korrekt

➤ **Weitere Beispiele**      SSK-Stellungnahme zur KiKK-Studie  
(Kommentar in Strahlentelex November 2008  
[http://www.strahlentelex.de/Stx\\_08\\_524\\_S04-05.pdf](http://www.strahlentelex.de/Stx_08_524_S04-05.pdf) )

## INTERPHONE Teilnahme-Verzerrung (*participation bias*)

### Hirntumoren (Meningiome+Gliome) Nordische Länder kombiniert

Nordic combined (Lahkola et al, 2007)		2000-2004	1,522	3,301	0.78 (0.68, 0.91)	867
--	--	-----------	-------	-------	-------------------	-----

Nordic combined (Lahkola et al, 2008)		2000-2004	1,209	3,299	0.76, (0.65, 0.89)	573
--	--	-----------	-------	-------	--------------------	-----

	ever ru	never ru	total
<b>cases</b>	1440	1291	<b>2731</b>
<b>cntrls</b>	3905	2695	<b>6600</b>
<b>total</b>	<b>5345</b>	<b>3986</b>	<b>9331</b>

**OR = 0.77, 95%- KI=[0.69, 0.85], p<E-6**

**Regulärer Gebrauch hängt hochsignifikant negativ mit Hirntumoren zusammen Bias**

## Gliome – Nordische Länder kombiniert

### regular use

	ever	never	total
<b>cases</b>	<b>867</b>	655	<b>1522</b>
<b>controls</b>	2077	1224	<b>3301</b>
<b>total</b>	<b>2944</b>	<b>1879</b>	<b>4823</b>

**OR = 0.78 (0.68, 0.91)**

### long time user

	exposed	non-exposed	total
<b>cases</b>	<b>143</b>	1379	<b>1522</b>
<b>controls</b>	325	2976	<b>3301</b>
<b>total</b>	<b>468</b>	<b>4355</b>	<b>4823</b>

**OR = 0.95 (0.74, 1.23)**

- Falls die “regular use OR” „baseline“ wäre, dann hätte der Langzeitgebrauch ein Risiko von ca. 20% oder darüber
- Nutzertypen trennen, weil „Langzeit  $\subset$  Regulär“
- Vielleicht ist es möglich und sinnvoll, diese „Meta-Dosis-Wirkungs-Analyse“ im Rahmen der INTERPHONE Gesamtauswertung durchzuführen

## Gliome – Lahkola 2007, Table III

Years since first use <sup>1</sup>					
Never/nonregular use	629	1.0	330	1.0	1,281
1.5-4 <sup>2</sup>	384	0.77 (0.65, 0.92)	165	0.82 (0.65, 1.04)	895
5-9	342	0.75 (0.62, 0.90)	141	0.69 (0.54, 0.88)	739
≥10	143	0.95 (0.74, 1.23)	64	0.86 (0.62, 1.21)	220

	ltu	never or ru	total
<b>cases</b>	143	1355	<b>1498</b>
<b>controls</b>	220	2915	<b>3135</b>
<b>total</b>	<b>363</b>	<b>4270</b>	<b>4633</b>

**OR = 1.40, 95%- KI=[1.12, 1.74], p<0.0028**

- Relativ große Diskrepanz zwischen den ORs von Lahkola 2007 und den rohen ORs. Wie ist das zu erklären?

# INTERPHONE Fazit aus statistischer Sicht

---

- **INTERPHONE-Studie** weist statistische / methodische Defizite auf
- **Hypothesenbildung** vorwiegend qualitativ
- **Power / Fallzahl** unspezifisch – methodenabhängig  
Korrelation des Matching mit Exposition unklar  
Einzelländer zu kleine Fallzahlen  
„Regular use“ und OR=1.5 ist unrealistisch  
weil ursprünglich OR=1.5 mit 5+ Jahre Exposition veranschlagt worden war (siehe study protocol, p. 2)
- **Einzelländerreports** nicht-signifikante Resultate ohne Power
- **Partizipationsbias** keine Korrektur
- **Modellierung** nicht transparent, keine methodologischen Sensitivitätsanalysen
- **Meta-Analyse** Dosis-Wirkungs-Beziehung möglich, wenn *base-line-Adjustierung* und kein *recall-bias*