

UNIVERSITÄT LEIPZIG

Medizinische Fakultät

QSB1 Tag 4

8. Analyse des Zusammenhangs qualitativer Merkmale

Dr. rer. nat. Meinhard Mende
IMISE, Universität Leipzig

imise.

Lernziele

Am Ende dieser Vorlesung

- können Sie absolute Häufigkeiten zweier qualitativer Merkmale in einer Kreuztabelle darstellen und relative Häufigkeiten berechnen,
- können Sie das Odds Ratio berechnen und interpretieren,
- kennen Sie Bedingungen, unter welchem Sie aus einem statistischen Zusammenhang (Assoziation) auf eine kausale Wirkung schließen können.

V 8

8.1 Kontingenztafeln

Fall-Kontroll-Studie

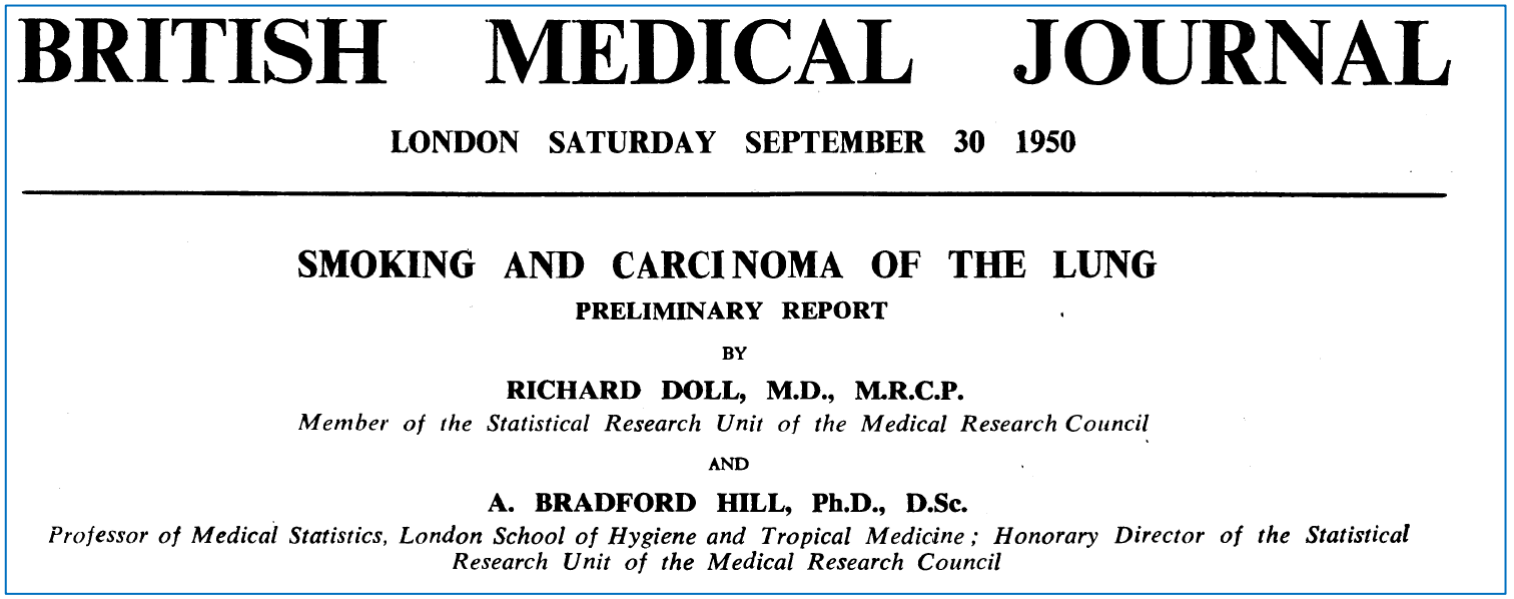
- untersucht Zusammenhänge zwischen Exposition und Erkrankung
- den Kranken werden vergleichbare Gesunde zugeordnet (matching) und retrospektiv die zu untersuchende Exposition in beiden Gruppen ermittelt.
- In Fall-Kontroll-Studien kann das Risiko-Verhältnis (RR) nicht bestimmt werden, nur das Odds Ratio (OR, → VL 14)

Beispiel: Zusammenhang von Rauchen und Lungenkarzinom



B 8.1 Fall-Kontroll-Studie

Doll R and Hill AB.
Br Med J 1950;4682: 739-48.



BRITISH MEDICAL JOURNAL

LONDON SATURDAY SEPTEMBER 30 1950

SMOKING AND CARCINOMA OF THE LUNG

PRELIMINARY REPORT

BY

RICHARD DOLL, M.D., M.R.C.P.

Member of the Statistical Research Unit of the Medical Research Council

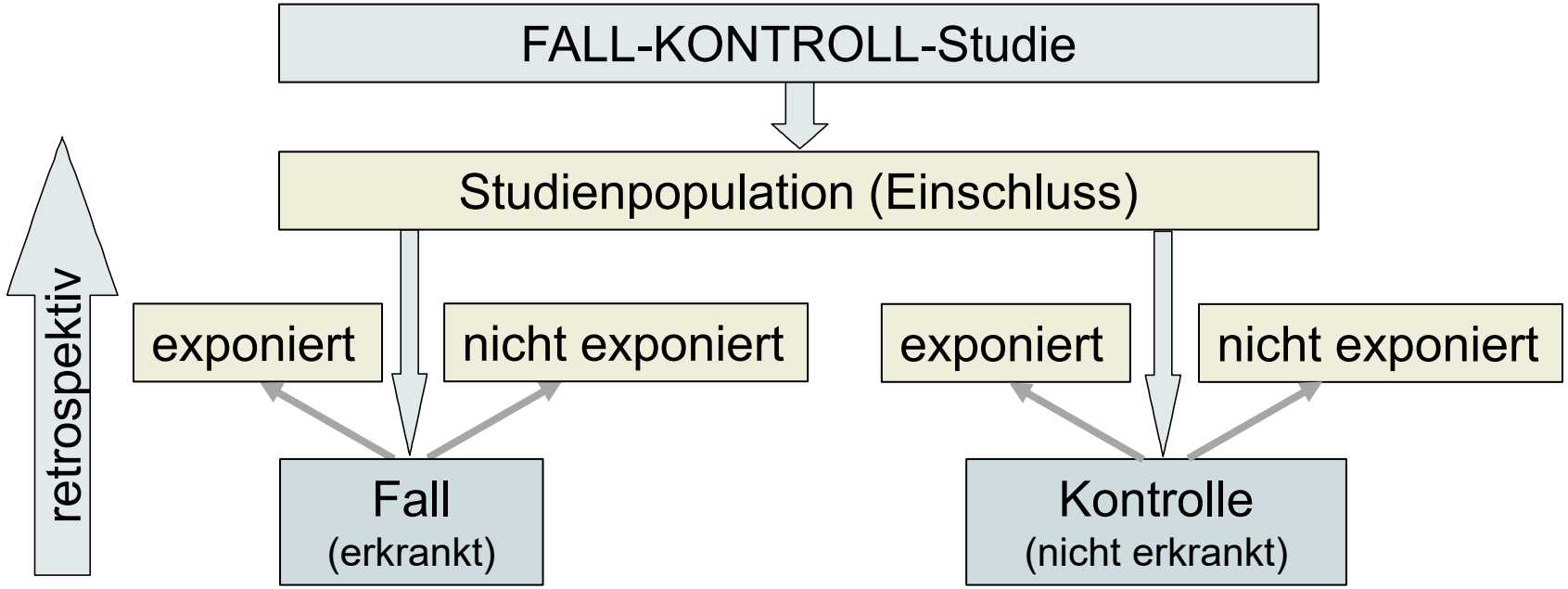
AND

A. BRADFORD HILL, Ph.D., D.Sc.

Professor of Medical Statistics, London School of Hygiene and Tropical Medicine; Honorary Director of the Statistical Research Unit of the Medical Research Council

- retrospektive Fall-Kontroll-Studie
- n = 649 männliche und 60 weibliche Patienten mit Lungenkarzinom
- ebenso viele Kontrollpatienten mit anderen Krankheiten
- Im Fokus: Zusammenhang mit Rauchen

Szenario: Fall-Kontroll-Studie



Kreuztabelle – mathematische Formalisierung

m Spalten

k x m - Kontingenztafel:		Merkmal 2				Randsumme
		1	2	…	m	
Merkmal 1 k Zeilen	1	b_{11}	b_{12}	…	b_{1m}	$b_{1.}$
	2	b_{21}	b_{22}	…	b_{2m}	$b_{2.}$
	…	…	…	…	…	…
	k	b_{k1}	b_{k2}	…	b_{km}	$b_{k.}$
Randsumme		$b_{.1}$	$b_{.2}$	…	$b_{.m}$	n

b_{ij} beobachtete Häufigkeit

Zeilenindex $i \in \{1, 2, \dots, k\}$

Spaltenindex $j \in \{1, 2, \dots, m\}$

Spezialfall: Vierfeldertafel (syn. 2 x 2 –Kontingenztabelle)

		Merkmal 2		Randsumme
		1	2	
Merkmal 1	1	b_{11}	b_{12}	$b_{1.}$
	2	b_{21}	b_{22}	$b_{2.}$
Randsumme		$b_{.1}$	$b_{.2}$	n

b_{ij} beobachtete Häufigkeit,

Zeilenindex $i \in \{1, 2\}$

Spaltenindex $j \in \{1, 2\}$

vereinfachte Darstellung
(nur für Vierfeldertafeln:

		Merkmal 2		Randsumme
		1	2	
	1	a	b	a + b
	2	c	d	c + d
Randsumme		a + c	b + d	n

Modellierung

Exposition (E):
- z. B. ein Umwelteinfluss,
- eine Intervention usw.

Erkrankung (D)
- erkrankt / nicht erkrankt

		Exposition (E)		Summe
		ja	nein	
Disease (D)	ja	a	b	a + b
	nein	c	d	c + d
Summe		a + c	b + d	n

Ziele:
- die Wahrscheinlichkeiten für Erkrankung in beiden Gruppen vergleichen
- das (z. B. erhöhte) Risiko der Erkrankung quantifizieren

B 8.1 Lungenkarzinome

binäre Variable Exposition

		Raucher		Summe
		ja	nein	
Gruppe	Lungenkarzinom	688 (51 %)	21 (26 %)	709
	Kontrollgruppe	650 (49 %)	59 (74 %)	709
Summe		1 338	80	50

binäre Variable Erkrankung

Spaltenprozente

V 8

8.2 Assoziation binärer Variablen

Assoziation

		Raucher		Summe
		ja	nein	
Erkrankung	Lungenkarzinom	688 (97 %)	21 (3.0 %)	709
	Kontrollgruppe	650 (92 %)	59 (8.3 %)	709
Summe		1 338	80	50

Beobachtung:

- Unter Nichtrauchern ist der Anteil der Kontrollgruppe deutlich höher als unter Rauchern: Assoziation
- Wie kann dieser Zusammenhang gemessen werden?

Unabhängigkeit kategorialer Merkmale

		Raucher		Summe
		ja	nein	
Erkrankung	Lungenkarzinom	669 (50 %)	40 (50 %)	709
	Kontrollgruppe	669 (50 %)	40 (50 %)	709
Summe		1 338	80	50

Gedankenexperiment

– Wären Rauchen und Lungenkarzinom unabhängig voneinander, müsste der Raucheranteil in beiden Gruppen gleich sein.

Modellierung 1

		Raucher (E)		Summe
		ja	nein	
Erkrankung (D)	Lungenkarzinom	688 (97 %)	21 (3.0 %)	709
	Kontrollgruppe	650 (92 %)	59 (8.3 %)	709
Summe		1 338	80	50

- zwei binäre Zufallsvariablen E und D:
  - Exposition: Raucher j / n
  - Erkrankung: Lungenkarzinom j / n
- Frage: Wahrscheinlichkeit für Erkrankung unter Rauchern häufiger als unter Nichtrauchern: $P(D^+|E^+) > P(D^+|E^-)$

Odds und Odds Ratio

– die Odds: $O = \frac{p}{1-p}$ (Wort gibt es nur in der Mehrzahl: dt. Chancen)

A: Ich wette 4 : 1, dass es morgen regnet.

(D. h., Wsk. für Regen 80%, trockenes Wetter 20%: $O_1 = \frac{0.8}{1-0.8} = 4 \mid 1$)

– Rater B schätzt die Situation anders ein:

B: Ich wette 9 : 1, dass es morgen regnet.

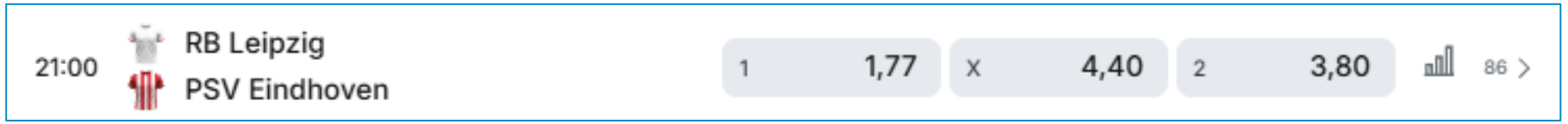
(D. h., die Odds sind hier $O_2 = \frac{0.9}{1-0.9} = 9 \mid 1$)

– Dann definieren wir das Odds Ratio (den Chancenquotienten):

– $OR = \frac{O_1}{O_2} = \frac{p_1 \cdot (1-p_2)}{(1-p_1) \cdot p_2}$

Beispiel für Odds

- Bei Sportwetten werden Odds (dt. *Chancen*, das Wort gibt es nur in der Mehrzahl) angegeben, z. B. hier Champions League 2025:



https://sports.tipico.de/de/alle/fussball/champions-league/ (2026-10-11)

- Wir blenden Unentschieden mal aus.
- So wurden die Chancen, dass RB Leipzig gegen PSV Eindhoven gewinnt, 3,80 zu 1,77 eingeschätzt, also etwa 2,1 zu 1.

Odds Ratio für Lungenkarzinom

		Raucher (E)		Summe
		ja	nein	
D	Lungenkarzinom	688	21	709
	Kontrollgruppe	650	59	709
Summe		1 338	80	50

		E		Summe
		ja	nein	
D	ja	a	b	a + b
	nein	c	d	c + d
Summe		a + c	b + d	n

Schätzung für das Odds Ratio:

$$O_1 = \frac{a/(a+c)}{c/(a+c)} \qquad O_2 = \frac{b/(b+d)}{d/(b+d)}$$

$$\widehat{OR} = \frac{a/c}{b/d} = \frac{a \cdot d}{b \cdot c},$$

also für das Beispiel Lungen-Ca.

$$\widehat{OR} = \frac{688 \cdot 59}{650 \cdot 21} = 2{,}97 \approx 3$$

D. h., die *Chancen* (Odds) für Lungenkarzinom sind in dieser Kohorte unter Rauchern ca. dreimal so groß wie unter Nichtrauchern.

Eigenschaften des Odds Ratios

- Assoziationsmaß mit Wertebereich: $0 < OR < \infty$
- gleiche Chance: OR = 1 also keine Assoziation
  (aus statistischer Unabhängigkeit folgt immer OR = 1)
- erhöhte Chance: OR > 1
- erniedrigte Chance (ggf. protektiver Effekt): OR < 1
- Interpretation:
  Das OR ist der Faktor, um den sich die Chancen (Odds) für das Merkmal 1 (z. B. zu erkranken) erhöhen bzw. verringern, wenn das Merkmal 2 (z. B. eine Exposition) vorliegt.

B 8.1.2 Odds Ratio für Lungenkarzinom

Männer		Raucher (E)		Summe
		ja	nein	
D	Lungenkarzinom	647	2	649
	Kontrollgruppe	622	27	649
Summe		1 269	29	1298

Frauen		E		Summe
		ja	nein	
D	ja	41	19	60
	nein	28	32	60
Summe		69	51	120

Geschätztes Odds Ratio:

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

- Frauen: $\widehat{OR} = \frac{41 \cdot 32}{28 \cdot 19} = 2{,}47$
- Männer: $\widehat{OR} = \frac{647 \cdot 27}{622 \cdot 2} = 14{,}0$

D. h., Odds Ratios weit über 1 zeigen einen mittleren bis starken Zusammenhang zwischen Lungen-Karzinom und Rauchen.

Odds Ratio für das Gegenereignis

		Raucher (E)		Summe
		ja	nein	
D	Lungenkarzinom	688	21	709
	Kontrollgruppe	650	59	709
Summe		1 338	80	50

		E		Summe
		ja	nein	
D	ja	a	b	a + b
	nein	c	d	c + d
Summe		a + c	b + d	n

Ist das OR gesucht, als Raucher nicht am Lungen-Ca. zu erkranken,

kehrt sich der Quotient um: $OR_{Nicht\ erkranken} = \frac{c/a}{d/b} = \frac{b \cdot c}{a \cdot d}$

- Also in diesem Beispiel:
- $OR = \frac{650 \cdot 21}{688 \cdot 59} = 0{,}34$

Das Odds Ratio für das komplementäre Ereignis D^- (auch geschrieben $\overline{D}$) ist das Reziproke des OR für das Ereignis D^+ (D).

Odds Ratio bei Wechsel der Referenzgruppe

		Raucher (E)		Summe
		ja	nein	
D	Lungenkarzinom	688	21	709
	Kontrollgruppe	650	59	709
Summe		1 338	80	50

		E		Summe
		ja	nein	
D	ja	a	b	a + b
	nein	c	d	c + d
Summe		a + c	b + d	n

Frage:
Wieviel geringer ist das Risiko für Lungen-Ca. unter Nichtrauchern?

$$OR_{NR} = \frac{b/d}{a/c} = \frac{b \cdot c}{a \cdot d}$$

hier:

– $OR = \frac{21 \cdot 650}{59 \cdot 688} = 0{,}34$

Auch wenn man die Referenzgruppe wechselt (also E^- anstelle von E^+) verkehrt sich das Odds Ratio in das Reziproke.

Beispiel 8.1

TABLE IV.—*Proportion of Smokers and Non-smokers in Lung-carcinoma Patients and in Control Patients with Diseases Other Than Cancer*

Disease Group	No. of Non-smokers	No. of Smokers	Probability Test
Males: Lung-carcinoma patients (649)	2 (0·3%)	647	P (exact method) = 0·00000064
Control patients with diseases other than cancer (649) . .	27 (4·2%)	622	
Females: Lung-carcinoma patients (60)	19 (31·7%)	41	$\chi^2 = 5{\cdot}76$; n = 1 0·01 < P < 0·02
Control patients with diseases other than cancer (60) . .	32 (53·3%)	28	

Fisher's exact test

Chi²-Test vgl. V 11

Doll R., Hill B., Smoking and carcinoma of the lung, BMJ 1950

Beispiel 8.1: SPSS-Ausgabe

Disease Group	No. of Non-smokers	No. of Smokers	Probability Test
Males: Lung-carcinoma patients (649)	2 (0·3%)	647	P (exact method) = 0·00000064
Control patients with diseases other than cancer (649) . .	27 (4·2%)	622	
Females: Lung-carcinoma patients (60)	19 (31·7%)	41	$\chi^2 = 5{\cdot}76$; n = 1 0·01 < P < 0·02
Control patients with diseases other than cancer (60) . .	32 (53·3%)	28	

Doll R., Hill B., Smoking and carcinoma of the lung, BMJ 1950

Chi-Quadrat-Tests

Geschlecht		Wert	df	Asymptotische Signifikanz (zweiseitig)	Exakte Sig. (zweiseitig)
male	Pearson-Chi-Quadrat	22,044[d]	1	<,001	<,001
	Kontinuitätskorrektur[b]	20,316	1	<,001	
	Likelihood-Quotient	26,140	1	<,001	<,001
	Exakter Test nach Fisher				<,001
	Zusammenhang linear-mit-linear	22,027[e]	1	<,001	<,001
	Anzahl der gültigen Fälle	1298			
female	Pearson-Chi-Quadrat	5,763[f]	1	,016	,026
	Kontinuitätskorrektur[b]	4,910	1	,027	
	Likelihood-Quotient	5,815	1	,016	,026
	Exakter Test nach Fisher				,026
	Zusammenhang linear-mit-linear	5,715[g]	1	,017	,026
	Anzahl der gültigen Fälle	120			

Beispiel 8.1: SPSS-Ausgabe

Disease Group	No. of Non-smokers	No. of Smokers	Probability Test
Males: Lung-carcinoma patients (649)	2 (0·3%)	647	P (exact method) = 0·00000064
Control patients with diseases other than cancer (649) . .	27 (4·2%)	622	
Females: Lung-carcinoma patients (60)	19 (31·7%)	41	$\chi^2 = 5{\cdot}76$; n = 1 0·01 < P < 0·02
Control patients with diseases other than cancer (60) . .	32 (53·3%)	28	

Doll R., Hill B., Smoking and carcinoma of the lung, BMJ 1950

Odds ratios (→ Folie 18)

Frauen: $\widehat{OR} = 2{,}47\ [1{,}17; 5{,}19]$

Männer: $\widehat{OR} = 14{,}0\ [3{,}32; 59{,}3]$

Risikoschätzung

Geschlecht		Wert	95%-Konfidenzintervall Unterer Wert	Oberer Wert
male	Quotenverhältnis für Lungen-Karzinom (no / yes)	14,043	3,325	59,301
	Für Kohorten-Analyse Raucher = no	13,500	3,224	56,536
	Für Kohorten-Analyse Raucher = yes	,961	,946	,977
	Anzahl der gültigen Fälle	1298		
female	Quotenverhältnis für Lungen-Karzinom (no / yes)	2,466	1,172	5,188
	Für Kohorten-Analyse Raucher = no	1,684	1,084	2,617
	Für Kohorten-Analyse Raucher = yes	,683	,496	,941
	Anzahl der gültigen Fälle	120		
Gesamt	Quotenverhältnis für Lungen-Karzinom (no / yes)	2,974	1,787	4,949
	Für Kohorten-Analyse Raucher = no	2,810	1,726	4,572
	Für Kohorten-Analyse Raucher = yes	,945	,921	,969
	Anzahl der gültigen Fälle	1418		

Konfidenzintervall für das Odds Ratio

– Der Logarithmus des OR ist approximativ normalverteilt. Damit kann man ein Konfidenzintervall konstruieren:

$$OR \cdot e^{\pm z_{1-\frac{\alpha}{2}}\sqrt{Var(ln(OR))}} \quad \text{mit} \quad \widehat{Var}\big(ln(OR)\big) = \frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}$$

i

zum Beispiel: Lungen-Ca. unter Rauchern $\widehat{OR} = 2{,}97$

$\widehat{OR} \cdot e^{\pm 1{,}96\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}}$, also $2{,}97 \cdot e^{\pm 1{,}96\sqrt{\frac{1}{688}+\frac{1}{650}+\frac{1}{21}+\frac{1}{59}}}$

$95\%\ \text{C. I.}: [1{,}78; 4{,}94]$

Weil die 1 nicht im C.I. enthalten ist, könnte man das als zum 5%-Niveau signifikant erhöhte Chance für Lungen-Ca. unter Rauchern vs. Nichtrauchern interpretieren.

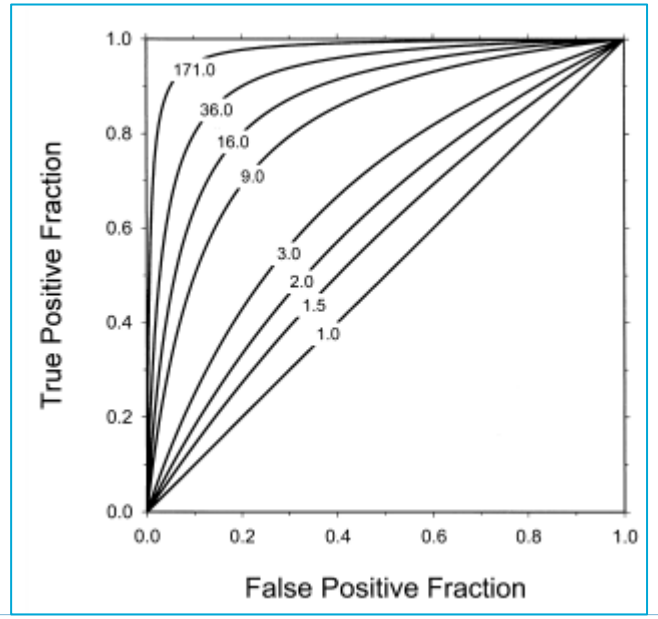
Zwei Anmerkungen

1. Enthält das $(1-\alpha)$-Konfidenzintervall für das Odds Ratio nicht die 1, kann das als signifikanter Test zum Niveau α interpretiert werden (→ V 9).
   - In der Tat sind Hypothesentest zum Signifikanzniveau α und $(1-\alpha)$- Konfidenzintervall zwei Seiten einer Medaille.
2. Nur bei ziemlich hohen Odds Ratios kann man bei exponierten Probanden ein Ereignis mit hoher Wahrscheinlichkeit vorhersagen (OR ≥ 36!).

Pepe MS et al. Am J Epidemiol 2004; 159:882-890



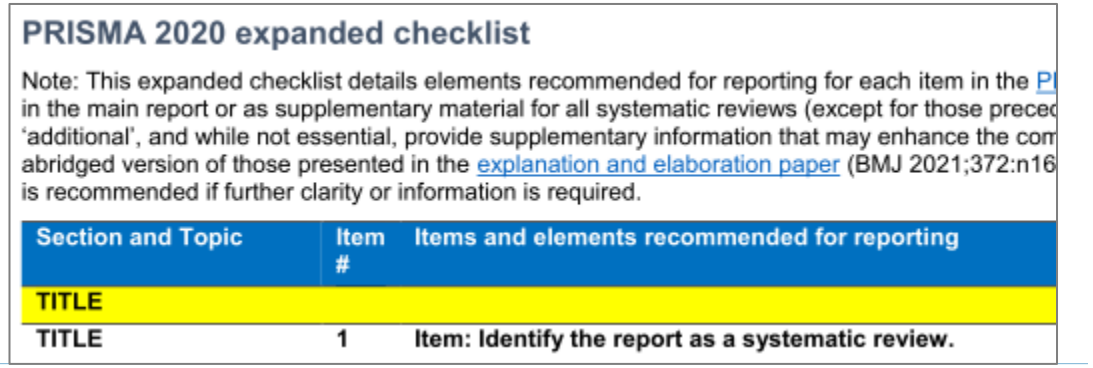
Jotquadrat: Grafitto in Bonn. commons. wikipedia.org



Systematisches Review und Metaanalyse

höchste Bündelung von Evidenz: systematisches Review & Metaanalyse

- Publikationen nach genauen Kriterien recherchieren (PICO …)
- Studien sorgfältig auswählen (z. B. nur RCTs)
- Risiko für Bias jeder Studie bewerten
- Effekte kombinieren
  - Odds Ratio (oder Risk Ratio) bei binären Endpunkten
- Komprimierte Darstellung von OR oder RR inkl. 95%-Konfidenzintervalle in Forest Plots.

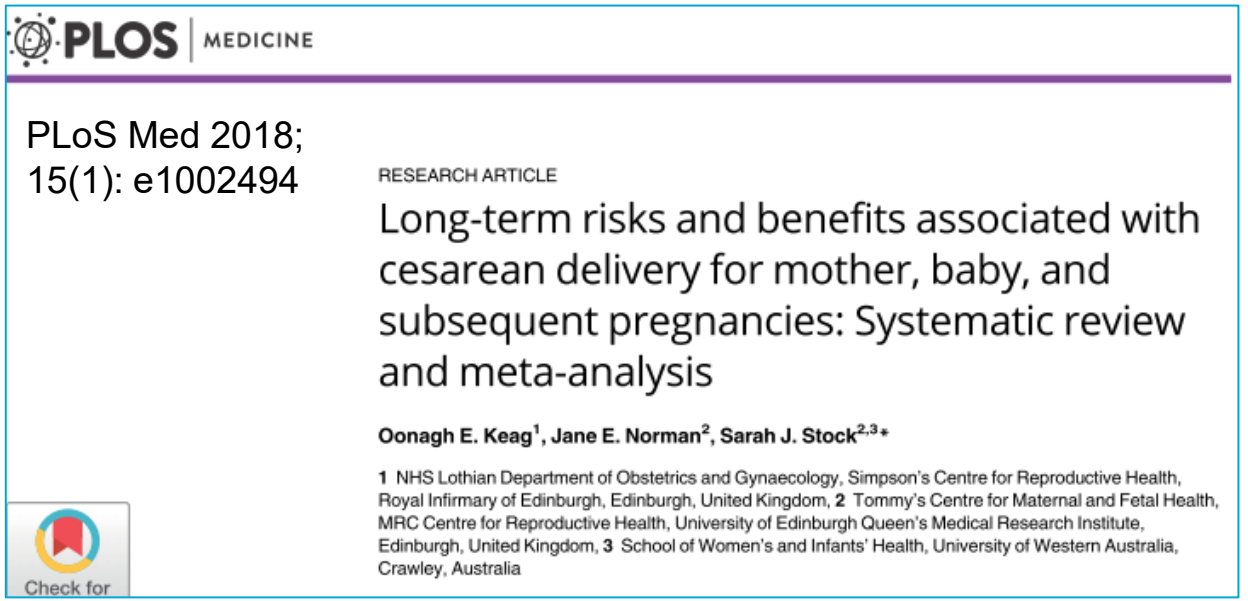


PRISMA 2020 expanded checklist

Note: This expanded checklist details elements recommended for reporting for each item in the PI in the main report or as supplementary material for all systematic reviews (except for those preced 'additional', and while not essential, provide supplementary information that may enhance the com abridged version of those presented in the explanation and elaboration paper (BMJ 2021;372:n16 is recommended if further clarity or information is required.

Section and Topic	Item #	Items and elements recommended for reporting
TITLE		
TITLE	1	Item: Identify the report as a systematic review.

Literaturbeispiel 8.5: Kaiserschnitt



PLOS | MEDICINE

PLoS Med 2018; 15(1): e1002494

RESEARCH ARTICLE

Long-term risks and benefits associated with cesarean delivery for mother, baby, and subsequent pregnancies: Systematic review and meta-analysis

Oonagh E. Keag[1], Jane E. Norman[2], Sarah J. Stock[2,3]*

1 NHS Lothian Department of Obstetrics and Gynaecology, Simpson's Centre for Reproductive Health, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom, **2** Tommy's Centre for Maternal and Fetal Health, MRC Centre for Reproductive Health, University of Edinburgh Queen's Medical Research Institute, Edinburgh, United Kingdom, **3** School of Women's and Infants' Health, University of Western Australia, Crawley, Australia

Check for

- Vergleich der Kurz- und Langzeitfolgen von Kaiserschnitt ohne medizinische Indikation vs. spontane Geburt
- nicht-indizierte Kaiserschnitte zum errechneten Geburtstermin
- eine RCT und 79 Kohorten-Studien: 29 928 274 Probandinnen
- Endpunkte: unmittelbare und späte Folgen für Mutter und Kind

B 8.5 – Ergebnisse der Metaanalysen

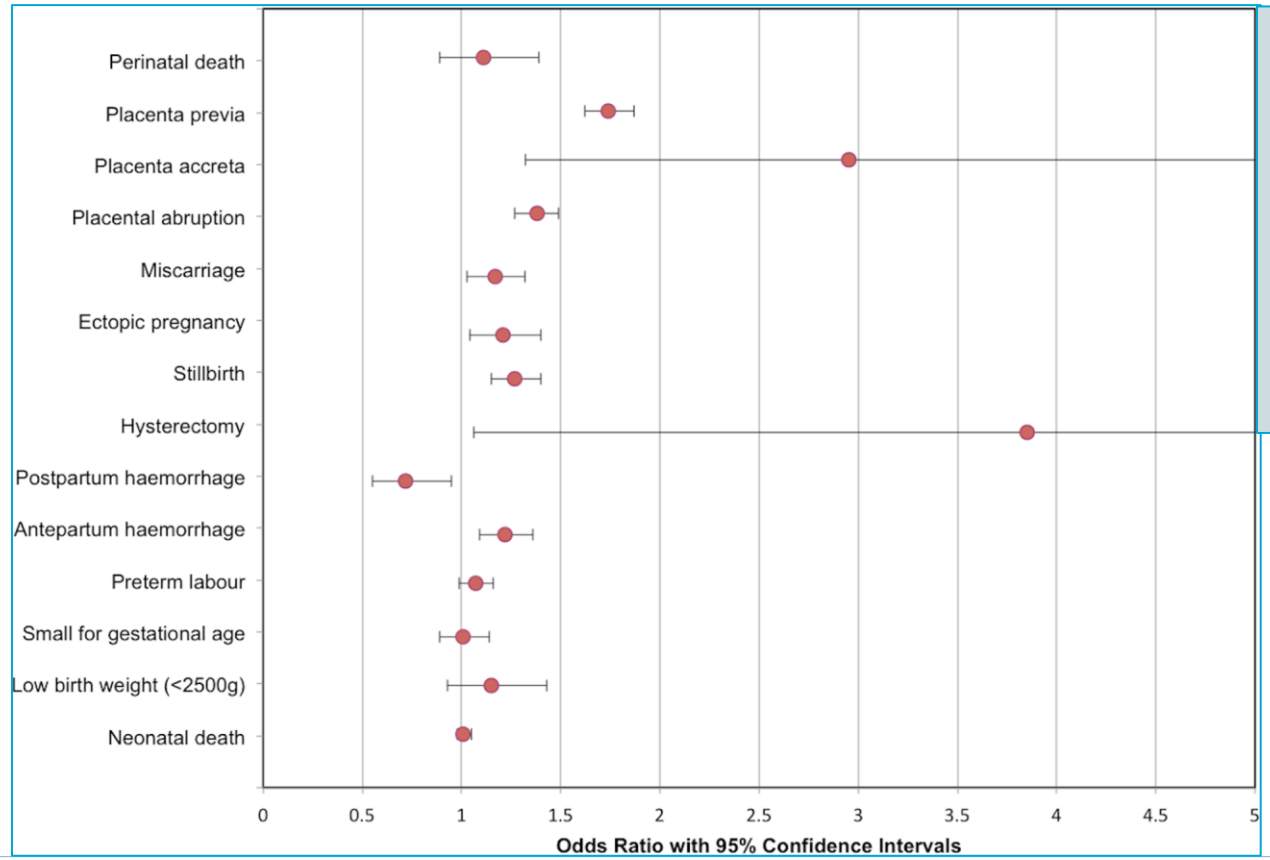
Table 2. Summary of meta-analyses. Table summarizing the meta-analyses performed detailing the number of studies, number of participants, effect estimate of each outcome and statistical method used. As studies had multiple cohorts and different follow-up periods, meta-analyses were divided according to age or duration of follow-up.

Outcome	Studies	Participants	OR [95%CI]	Statistical Method
Maternal Outcomes				
Primary Outcomes				
Urinary incontinence	8	58900	0.56 [0.47, 0.66]	Mantel-Haenszel Random Effects
Pelvic organ prolapse	2	39208	0.29 [0.17, 0.51]	Mantel-Haenszel Random Effects
Fecal incontinence	5	43260	1.04 [0.73, 1.48]	Mantel-Haenszel Random Effects
Secondary Outcomes				
Pelvic pain	2	18308	0.74 [0.54, 1.00]	Mantel-Haenszel Fixed Effects
Subfertility	11	3692014	1.60 [1.45, 1.76]	Mantel-Haenszel Random Effects
Childhood Outcomes				
Primary Outcomes				
Asthma up to 12 years	13	887960	1.21 [1.11, 1.32]	Mantel-Haenszel Random Effects
Adulthood asthma from 15 years	2	9072	1.87 [0.65, 5.32]	Mantel-Haenszel Random Effects

- Assoziationsmaß: Odds Ratio (OR)
- OR > 1 → höhere Chancen (Odds) für Risiko bei Kaiserschnitt-Geburten
- OR < 1 → geringere Chancen bei Kaiserschnitt-Geburten

8.5 Forest-Plot 1

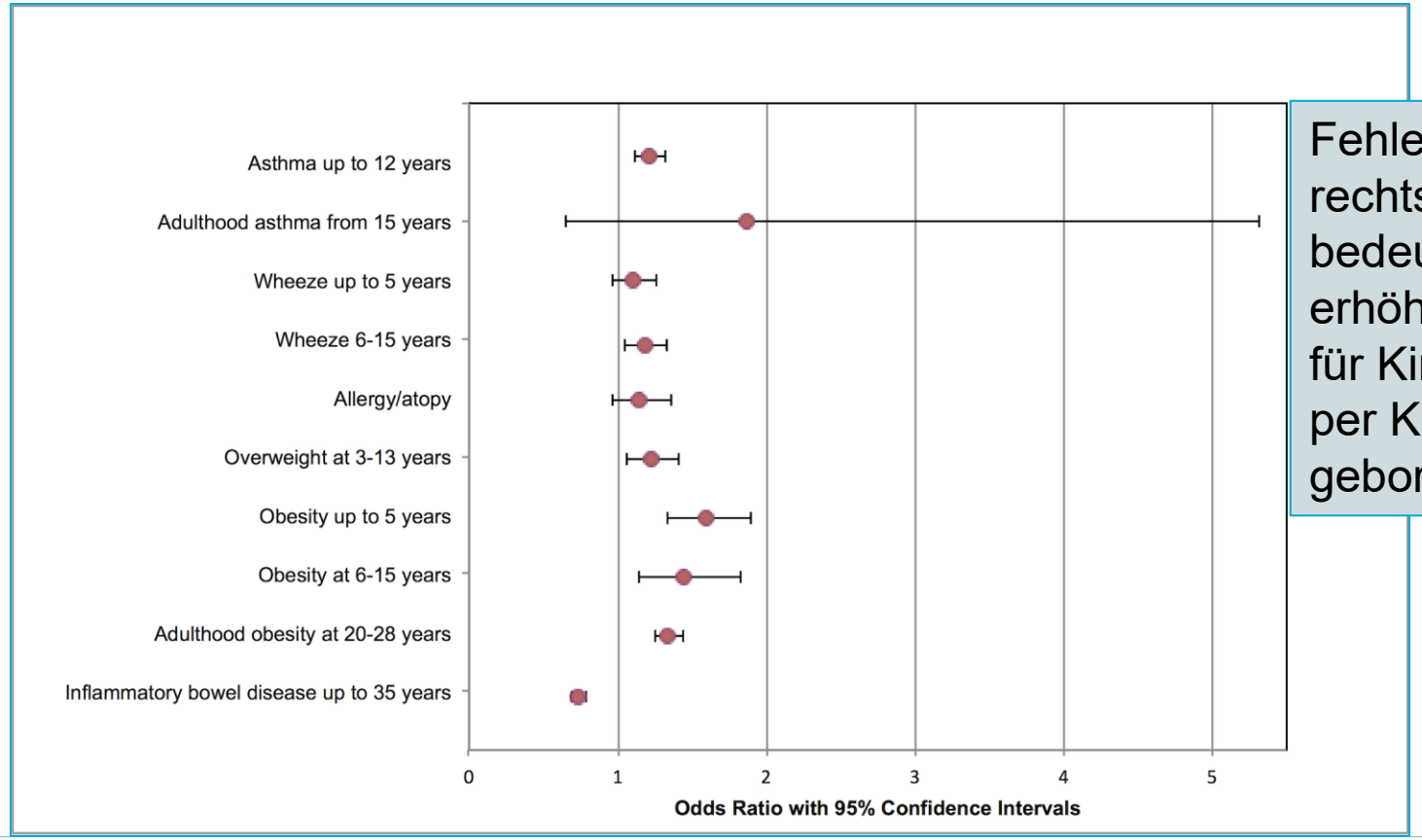
– z. B. Langzeitfolgen für die Mutter



Fehlerbalken rechts von 1 bedeuten erhöhtes Risiko für Mütter, welche mit Kaiserschnitt entbunden haben

8.5 Forest-Plot 2

– z. B. Langzeitfolgen für das Kind



Fehlerbalken rechts von 1 bedeuten erhöhtes Risiko für Kinder, welche per Kaiserschnitt geboren wurden.

Risk of bias assessment

- Alle eingeschlossenen Studien sollten nach diesen (oder ähnlichen) Kriterien hinsichtlich des Risikos für Verzerrung eingeschätzt werden.

S5 Table: The Risk of Bias Assessment tool for Non-randomized Studies (RoBANS)

Maternal outcomes studies

Study	Selection of participants	Confounding variables	Measurement of exposure	Blinding of outcome assessments	Incomplete outcome data	Selective outcome reporting
Abdel-Fattah (2011)[1]	Low	Low	Low	Unclear	Unclear	Low
Bjelland (2016) [2]	Low	Low	High	Unclear	Low	Low
Brown (2012) [3]	Low	High	High	High	High	Low
Elvander (2015) [4]	Low	Low	Low	Unclear	Unclear	Low
Fussing-Clausen (2014) [5]	Low	High	Low	Unclear	Unclear	Low
Gartland (2012) [6]	Low	Low	High	High	Unclear	Low
Gyhagen (2013) [7]	Low	Low	High	High	Low	Low
Gyhagen (2013) [8]	Low	Low	High	High	Low	Low
Gyhagen (2015) [9]	Low	Low	High	High	Low	Low

Zusammenfassung

- Zusammenhänge zwischen kategorialen Merkmalen äußern sich darin, dass einzelne Merkmals-Kombinationen unter- / überproportional häufig vorkommen.
- Zusammenhänge (Assoziationen) binärer Größen können z. B. durch das Odds Ratio gemessen werden
- OR > 1 bedeutet höhere Chancen (odds) für die Kategorie gegenüber der Referenzgruppe, OR < 1 geringere Chancen.
- Das OR ist ein *ziemlich künstliches* Assoziationsmaß, das aber in Fall-Kontroll-Studien angewandt werden kann und in vielen Publikationen (z. B. Metaanalysen) angegeben wird.
  (→ vgl. V 14)

UNIVERSITÄT LEIPZIG

Medizinische Fakultät

VIELEN DANK!

Dr. rer. nat. Meinhard Mende

Senior-Wissenschaftler, Institut für Medizinische Informatik, Statistik und Epidemiologie und Zentrum für Klinische Studien

Härtelstraße 16-18, 04107 Leipzig

Meinhard.mende@zks.uni-leipzig.de