

Anwendungen der Vierfeldertafel

2x2 - Tafel, Berechnungen per Hand (pocket calculator)

Die 2x2-Tafel ist eine zweidimensionale Darstellung von Häufigkeiten der vier verschiedenen Ausprägungskombinationen zweier dichotomer Variablen. Eine übliche Nomenklatur zeigt nebenstehende Abbildung mit den Variablen Var 1 und Var 2 und der Bezeichnung (1,0) = (krank, gesund) oder (ja, nein) oder (Raucher, Nichtraucher) oder (weiblich, männlich) oder (exponiert, nicht exponiert) oder (testpositiv, testnegativ) usw.

		Var 1		
		1	0	
Var 2	1	a	b	a+b
	0	c	d	c+d
		a+c	b+d	n

Solche Tafeln können für verschiedene Berechnungen mit zwei dichotomen Variablen verwendet werden. Beispiele sind: Test auf Unabhängigkeit, Test auf Homogenität, McNemar - Test, Maße der Übereinstimmung und des Zusammenhangs, diagnostischer Test, epidemiologische Maßzahlen oder Fehlklassifikationen.

1. Chi² - Unabhängigkeitstest

Mit **einer Zufallsstichprobe** werden im Rahmen einer Studie **zwei** dichotome Merkmale simultan erfasst, z.B. Geschlecht (W,M) und Rauchgewohnheit (ja = 1, nein = 0). Bei solchen Studien kann nur die Fallzahl n frei gewählt werden (Fallzahlplanung !), die Randsummen (a+b), (a+c), (c+d) und (b+d) ergeben sich zufällig. Man möchte die Frage beantworten, ob die Rauchgewohnheit mit dem Geschlecht assoziiert ist oder ob beide Merkmale voneinander unabhängig sind. Hierzu dient der **Chi²-Unabhängigkeitstest**.

		Rauchen		
		1	0	
Geschl	W	41	72	113
	M	60	77	137
		101	149	250

$$\chi^2 = \frac{n \cdot (a \cdot d - b \cdot c)^2}{(a+b) \cdot (a+c) \cdot (b+d) \cdot (c+d)}$$

$$\chi_Y^2 = \frac{n \cdot \left(|a \cdot d - b \cdot c| - \frac{n}{2} \right)^2}{(a+b) \cdot (a+c) \cdot (b+d) \cdot (c+d)}$$

Yates-Korrektur

(Daten: Hartung, Statistik, 11. Aufl.)

Per Handrechnung erhält man mit n = 250 für Chi² = 1,45142 (**p = 0,228**). Für Fallzahlen kleiner n < 60 sollte man die **Yates-Korrektur** des Test verwenden. Dieses nichtsignifikante Ergebnis bedeutet, dass kein Zusammenhang besteht, beide Merkmale demnach in dieser Stichprobe voneinander unabhängig sind.

2. Chi² - Homogenitätstest

Vergleicht man **zwei** Zufallsstichproben **A** und **B** (z.B. zwei Therapiegruppen) hinsichtlich **einer** dichotomen Charakteristik (z.B. krank = 1, gesund = 0), so ergibt sich eine andere Fragestellung. Man möchte prüfen, ob die Anteile der z.B. Kranken in beiden Gruppen gleich sind. Hierfür gibt es zwei Möglichkeiten: **a. Chi² - Homogenitätstest** und **b. 2-Stichprobentest für Anteile**.

		Var 1		
		1	0	
A		a	b	a+b
B		c	d	c+d
		a+c	b+d	n

Als **Beispiel** soll die bekannte APSAC - Studie dienen, bei der für Patienten mit akutem Herzinfarkt ein neuer Blutgerinnungshemmer APSAC (*Anisoylated Plasminogen Streptokinase Activator Complex*) mit der damaligen Heparin - Standardtherapie verglichen wurde. Hauptzielkriterium war die Verminderung der Hospitalmortalität (Sterblichkeit) innerhalb der ersten 28 Tage nach Infarkt. Die Studiendaten sind in der folgenden 2x2 - Tafel zusammengefasst.

	gestorben	überlebt	
APSAC	9	153	162
Heparin	19	132	151
	28	285	313

$$\chi^2 = \frac{313 \cdot (9 \cdot 132 - 19 \cdot 153)^2}{28 \cdot 285 \cdot 162 \cdot 151} = 4,738$$

$$p = 0,0295$$

a. Chi² - Homogenitätstest

Ergebnis: Unter APSAC versterben 9 von 162 Patienten (5,6%), unter Heparin 19 von 151 (12,6%). Es kann wieder die Chi²-Teststatistik (wie bei Punkt 1) angewendet werden, die den Wert Chi² = 4,738 liefert. Der Unterschied ist gerade signifikant (p = 0,0295) und deutet darauf hin, dass die Mortalität unter APSAC geringer ist, als unter Heparin.

b. 2-Stichprobentest für Anteile

Die Teststatistik lautet: $z = (p_A - p_H) / SE(p_A - p_H)$ mit

$$SE(p_A - p_H) = \sqrt{p \cdot (1-p) \cdot \left(\frac{1}{n_A} + \frac{1}{n_H} \right)} \quad \text{und} \quad p = \frac{n_A}{n} \cdot p_A + \frac{n_H}{n} \cdot p_H$$

$$n_A = 162 ; n_H = 151 ; n = 313$$

Wie oben unter a. angegeben ist der Anteil unter APSAC Verstorbener $p_A = 0,0556$ und entsprechend der Anteil unter Heparin $p_H = 0,1258$. Den Standardfehler SE unter H_0 : „beide Anteile sind gleich“, erhält man aus obiger Formel zu $SE = 0,0322837$.

Damit ergibt sich für die Teststatistik $z = -0,0702722 / 0,0322837 = -2,1767$ mit einem p - Wert von $p = 0,0295$.

Ergebnis:

Beide Methoden ergeben ein signifikantes Ergebnis und man erkennt außerdem wegen $\chi^2 = z^2$ also $\{4,738 = (-2,1767)^2\}$ die Äquivalenz der beiden Teststatistiken.

3. Chi²_{McN} - McNemar-Test

Werden zwei Beobachtungen eines dichotomen Merkmals (1,0) an den gleichen Individuen oder Objekten durchgeführt, so sind sie nicht unabhängig (sie sind **verbunden**). Einige Beispiele:

- Das Vorhandensein eines bestimmten Symptoms (1,0) vor und nach einer Behandlung,
- Ein Eignungstest (ja,nein) zur Beurteilung von Personen durch zwei Beurteiler (z.B. Personalchef und externer Psychologe),
- Beurteilung von Röntgenbildern (krank, gesund) durch zwei Ärzte.

		Beob B		
		1	0	
Beob A	1	a	b	a+b
	0	c	d	c+d
		a+c	b+d	n

Beurteilen zwei Beobachter A und B (jeder für sich) die gleiche Stichprobe von n Individuen / Objekten hinsichtlich eines dichotomen Merkmals, so erwartet man unter der Nullhypothese H_0 gleiche Anteile von „1“ (oder „0“), somit $(a+c)/n = (a+b)/n$, d.h., $b = c$. Daher basiert die Teststatistik nur auf Häufigkeiten der unterschiedlichen Beurteilungen

$$\chi^2_{McN} = (b - c)^2 / (b + c)$$

für $8 \leq (b+c) < 30$ mit Stetigkeitskorrektur

$$\chi^2_{McN} = (|b - c| - 1)^2 / (b + c)$$

Beispiel:

40 Schüler werden hinsichtlich ihrer Aufmerksamkeit im Unterricht (+, -) von 2 Personen (A und B) beurteilt. Es ergaben sich die in der Tabelle angegebenen Häufigkeiten: (Daten: Hedderich / Sachs)
Der Anteil „+“ beträgt bei A 17/40 und bei B 19/40.

		B		
		+	-	
A	+	14	3	17
	-	5	18	23
		19	21	40

Wegen $(b+c) = 8$ verwendet man χ^2_{McN} mit Stetigkeitskorrektur und erhält $\chi^2_{McN} = 0,125$. Mit $DF = 1$ resultiert damit ein $p = 0,724$ (nicht signifikant).

Ergebnis: In der Grundgesamtheit sind die Beurteilungen gleich.

Der McNemar - Test prüft nur, ob es einen Unterschied gibt zwischen den Beurteilungen. Zum Grad der Übereinstimmung gibt er keine Auskunft. Hierfür wird z.B. Kappa verwendet.

4. Zufallskorrigierte Übereinstimmung - Kappa κ

Cohen's Kappa κ ist ein Maß für die zufallskorrigierte Übereinstimmung zweier Beobachter (rater, observer) bei der Bewertung eines dichotomen Merkmals der gleichen Stichprobe. Es quantifiziert die Fähigkeit beider Beobachter zur übereinstimmenden Klassifikation der Stichprobenelemente. Von der beobachteten Übereinstimmung p_0 wird die erwartete Übereinstimmung p_e für eine zufällig gleiche Klassifizierung abgezogen, so daß nur die „echte Leistung“ der Beobachter einfließt. Die maximale zufallskorrigierte Übereinstimmung ist $(1 - p_e)$. Damit kann κ Werte zwischen 0 (Übereinstimmung nur aufgrund des Zufalls) und 1 (vollständige Übereinstimmung) annehmen. Negative Werte (schlechter als der Zufall) sollten nicht interpretiert werden. Eine Interpretation der Werte für κ liefert die nachfolgende Tabelle.

Berechnung: $\kappa = (p_0 - p_e) / (1 - p_e)$

Werte von κ	Stärke der Übereinstimmung	
< 0,20	sehr schwach	Approximales Konfidenzintervall: $\{\kappa - 1,96 \cdot SE(\kappa) ; \kappa + 1,96 \cdot SE(\kappa)\}$ mit $SE^2(\kappa) = \{p_0 \cdot (1 - p_0) / n \cdot (1 - p_e)^2\}$
0,21 - 0,40	schwach	
0,41 - 0,60	moderat	
0,61 - 0,80	gut	
0,81 - 1,00	sehr gut	

Beispiel:

Für die Beurteilung der 40 Schüler in Punkt 3. (McNemar) erhält man für $p_0 = (a+d) / n = (14+18) / 40 = 0,8$ und $p_e = \{(a+b) \cdot (a+c) + (b+d) \cdot (c+d)\} / n^2 = (19 \cdot 17 + 21 \cdot 23) / 1600 = 0,504$.
 $\kappa = (0,8 - 0,504) / (1 - 0,504) = 0,597$. Die von Hand berechnete Größe des Konfidenzintervalls ergibt sich mit $SE = 0,127447$ zu $(0,347 ; 0,847)$.

Ergebnis: Die Stärke der Übereinstimmung liegt aufgrund der geringen Fallzahl zwischen schwach und sehr gut und kann vielleicht als moderat interpretiert werden.

5. Maße des Zusammenhangs

Der χ^2 -Test von Punkt 1 ergibt ein nicht signifikantes Ergebnis. Die dichotomen Merkmale Rauchen und Geschlecht sind bei diesen Daten voneinander unabhängig, es besteht also kein Zusammenhang zwischen beiden Merkmalen (Variablen).

Beispiel für Zusammenhang:

Betrachtet man dagegen aus einer anderen Studie mit 82 Kindern die Merkmale Zahnpflege (1=gut, 0=schlecht) und Karies (1=vorhanden, 0=nicht vorhanden), so erhält man mit dem χ^2 - Test ein signifikantes Ergebnis.

$\chi^2 = 68,366$ und damit ein p - Wert von $p = 0,000$. Somit besteht ein deutlicher Zusammenhang zwischen den beiden Variablen, der z.B. mit dem Wert für **Cramers V**, aber auch mit **Pearsons Korrelationskoeffizienten CC_{corr}** (korrigiert) quantifiziert wird.

Die unten angegebene Excel - Tabelle ermöglicht zudem eine approximative Berechnung der entsprechenden Konfidenzintervalle (Quelle: Hartung: Statistik, 11. Aufl.).

		Karies		
		1	0	
Zpfl	1	1	23	24
	0	56	2	58
		57	25	82

Berechnung der Konfidenzintervalle für Cramers V und Pearsons CCcorr in 2x2 Tafeln der Form							
Chi² =	68,366				a	b	a+b
					c	d	c+d
					a+c	b+d	n
Einfügen der Werte bei a bis d							82
	a	b	c	d	a+b	c+d	a+c
	1	23	56	2	24	58	57
							b+d
Cramers V =		0,9131			K.I. =	0,8170	1,0092
Pearson CCcorr =		0,9536			K.I. =	0,8989	1,0083

$$\text{Cramers V} = \sqrt{\frac{\chi^2}{n}}$$

$$\text{CC}_{\text{corr}} = \sqrt{\frac{2 \cdot \chi^2}{\chi^2 + n}}$$

Werte größer als 1 (obere Konfidenzgrenze in der Tabelle) werden auf 1 gesetzt. Beide Werte (Cramers V und CC_{corr}) verändern sich nicht mit der Fallzahl, solange die Anteile gleich bleiben. Für eine 2x2 - Tafel gelten die obigen Formeln. Werte von V oder CC_{corr} kleiner 0,2 deuten auf keinen oder einen sehr schwachen Zusammenhang hin. Im aufsteigenden Bereich von 0,2 bis 0,8 spricht man von „eher schwach“, „mittel“ und „eher stark“. Werte größer als 0,8 deuten auf einen starken Zusammenhang hin.

6. Diagnostischer Test

Ein diagnostischer Test ist eine klinische Entscheidung anhand von Informationen, wie z.B. Laborwerten, Symptomen, Röntgen-Bildern sowie diagnostischer Kriterien, ob eine Erkrankung bei einem Individuum wahrscheinlich vorliegt oder nicht. Dabei sind zwei richtige und zwei falsche Entscheidungen möglich, die man in einer Vierfeldertafel zusammenfassen kann. T+ und T- sind die Testentscheidungen (testpositiv; testnegativ), K+ und K- stehen für „krank“ bzw. „nicht krank“. Als „Goldstandard“ bezeichnet man die wahre Diagnose, also ob eine Krankheit tatsächlich vorliegt (z.B. mit Biopsie). Die Häufigkeiten in der 2x2-Tafel haben folgende Bedeutung:

	Goldstandard		
	K+	K-	
T+	a	b	a+b
T-	c	d	c+d
	a+c	b+d	n

Häufigkeit der Diagnosen: a richtig positiv
 b falsch positiv
 c falsch negativ
 d richtig negativ

Zeilen- und Spaltensummen: a+b alle Testpositiven
 c+d alle Testnegativen
 a+c alle wirklich Kranken
 b+d alle wirklich Gesunden
 n = a+b+c+d Gesamtzahl

Außerdem sind:
 beobachtete Prävalenz: $P^* = (a+b) / n$
 geschätzte wahre Prävalenz: $P = (a+c) / n$ (bei repräsentativer Stichprobe)

Zusammenhang zwischen beobachteter und wahrer Prävalenz:

$$P^* = P \times \text{SENS} + (1-P) \times (1-\text{SPEZ})$$

$$= \frac{a}{n} + \frac{b}{n}$$
richtig positiv Rate falsch positiv Rate

Die Validität (Güte) eines diagnostischen Tests wird quantifiziert durch 2 Größen: **SENSITIVITÄT** (SENS) = $W(T+|K+)$ und **SPEZIFITÄT** (SPEZ) = $W(T-|K-)$

Dabei ist $W(T+|K+)$ die bedingte WSK, dass der Test positiv ist (T+) bei vorhandener Krankheit und entsprechend $W(T-|K-)$ die bedingte WSK für einen negativen Test bei Gesunden (WSK = Wahrscheinlichkeit). Somit berechnet man beide Größen aus der Vierfeldertafel nach:

SENS = $a / (a+c)$ und SPEZ = $d / (b+d)$

SENS und SPEZ sind nicht prävalenzabhängig.

Die Validität des Tests hängt z.B. ab von der:

- Fähigkeit des Laborwertes oder Teststreifens, Krank von Gesund zu trennen (Grenzwert)
- Fähigkeit des Arztes, Kranke als krank und Gesunde als gesund richtig zu identifizieren (dichotom).

Die Patienten möchten in der Regel aber nicht wissen, wie gut der Test bei Gesunden und Kranken die richtige Auskunft gibt. Sie möchten wissen, ob sie erkrankt sind, wenn der Test positiv ist oder nicht erkrankt bei negativem Test. Die Antwort geben die **prävalenzabhängigen (!) prädiktiven Werte PW+ und PW-**.

Prädiktive Werte: **PW+ = $W(K+|T+)$ und PW- = $W(K-|T-)$** (Vorhersagewerte)

Man erhält sie durch Anwendung der Bayes - Formeln:

$$PW+ = \frac{P \cdot \text{SENS}}{(P \cdot \text{SENS}) + (1 - P) \cdot (1 - \text{SPEZ})} \quad PW- = \frac{(1 - P) \cdot \text{SPEZ}}{(1 - P) \cdot \text{SPEZ} + P \cdot (1 - \text{SENS})}$$

Dabei ist P die Prävalenz in der Population, die vor Anwendung des Tests geschätzt werden muss.

Zur Ermittlung von SENS und SPEZ für einen Test auf Vorliegen einer bestimmten Krankheit rekrutiert man eine Anzahl wirklich daran Erkrankter und eine Anzahl wirklich nicht daran Erkrankter (Goldstandard). An diesen Personen wird der Test angewendet und gemessen, wie groß der Anteil der Testpositiven an den Kranken (SENS) und der Anteil der Testnegativen an den Gesunden (SPEZ) ist. Entsprechen die Werte den Anforderungen, kann der Test in der Bevölkerung verwendet werden und liefert die prädiktiven Werte z.B. mit der Bayes - Formel. Als Vorkennntnis ist hierzu eine Schätzung der Prävalenz der fraglichen Krankheit in der Bevölkerung erforderlich. Statt Bayes - Formel kann auch eine Vierfeldertafel aufgestellt und PW+ und PW- hieraus gewonnen werden, sofern $(a+c)/n$ der Prävalenz in der Bevölkerung entspricht.

Beispiel: Die jährliche Erkrankungsrate an Darmkrebs in Deutschland ist abhängig von Alter und Geschlecht. Für 70 - 74 Jährige Männer beträgt sie etwa 0,3% pro Jahr, d.h., 30 Fälle pro 10.000 Männer (*Quelle: RKI*). Zur Früherkennung wird ein immunologischer Test auf Blut im Stuhl angeboten (iFOBT), der typischerweise eine SENS von 75% und eine SPEZ von 92% hat. Welche Ergebnisse liefert dieser Test bei einem Screening von $n = 10.000$ Männern dieser Altersgruppe ?

	krank	gesund	
T+	22	798	820
T-	8	9172	9180
	30	9970	10000

Ausgehend von der Definition der bedingten WSK: $\text{SENS} = W(T+|K+) = a / (a+c)$ erhält man für $a = 30 \cdot 0,75 = 22,5$ (gerundet 22). Analoge Rechnung führt zu $d = 9172,4$ (gerundet 9172). Hieraus ergeben sich b und c und die Randhäufigkeiten (gerundet). Da es sich hier um die Prävalenz in der gesamten Population der 70 - 74 Jahre alten

Männer handelt, kann $PW+ = W(K+|T+) = 22 / 820 = 0,0268$ (2,68%) direkt aus der 2x2-Tafel abgelesen werden. Mit einer Wahrscheinlichkeit von nur 2,68% hat ein Proband aus dieser Männergruppe mit positivem Testergebnis tatsächlich Darmkrebs. Somit ist die WSK, dass er gesund ist trotz eines positiven Tests (falsch positiv) gleich $W(K-|T+) = 97,32\%$. Alternativ kann diese Berechnung mittels Bayes - Formel erfolgen.

Mit welcher Wahrscheinlichkeit ein Proband mit negativem Testergebnis keinen Darmkrebs hat ergibt sich aus $PW- = W(K-|T-) = 9172 / 9180 = 0,9991$ (99,91%) Die WSK für Darmkrebs trotz negativem Test (falsch negativ) erhält man aus $W(K+|T-) = 1 - W(K-|T-) = 0,0009$ (0,09%). Alternativ kann diese Berechnung mittels Bayes - Formel erfolgen.

Die beobachtete Erkrankungsrate P^* pro Jahr ergibt sich aus $P^* = 0,003 \cdot 0,75 + 0,997 \cdot 0,08 = 0,082$ (8,2%) und ist damit deutlich größer, als die tatsächliche Erkrankungsrate von 0,3%.

7. Epidemiologische Maßzahlen

a. Risiko und relatives Risiko

Beobachtet werden Personen, die zu einem Startzeitpunkt gesund sind und von denen (im einfachsten Fall) ein Teil mit einem Risikofaktor für eine Krankheit exponiert ist, der andere Teil nicht. In solchen **Kohortenstudien** werden einige Exponierte und auch einige Nicht-exponierte während der Studie erkranken. Die entsprechenden Häufigkeiten lassen sich nach Abschluss der Studie in einer 2x2 - Tafel zusammenfassen.

Das **Risiko** zu erkranken beträgt für Exponierte $R_E = a/(a+b)$ und für Nichtexponierte $R_N = c/(c+d)$ und ist somit jeweils der Anteil der Erkrankten unter den Exponierten bzw. Nichtexponierten. Diese Anteile quantifizieren die jeweilige Wahrscheinlichkeit (WSK). Der Quotient dieser beiden kumulativen Inzidenzmaße CI_E und CI_N ist

		Krankheit		
		1	0	
Exposition	1	a	b	a+b
	0	c	d	c+d
		a+c	b+d	n

das **relative Risiko** $RR = R_E / R_N$. Es quantifiziert die Stärke des Zusammenhangs zwischen Exposition und Krankheit. Ist $RR > 1$ fördert die Exposition die Erkrankung (z.B. Rauchen), bei $RR < 1$ wird sie gehemmt (z.B. Impfen) und bei $RR = 1$ ist das Auftreten der Erkrankung unabhängig von der Exposition (hierbei Konfidenzintervalle beachten).

b. Attributable Risiken ($RR > 1$)

Die **Risikodifferenz** RD beschreibt die WSK, dass eine exponierte Person erkrankt aufgrund des zusätzlichen, über das Grundrisiko der Nichtexponierten hinausgehenden Risikos. $RD = \text{attributables Risiko}$ (excess risk). Ist $W(K|E)$ die WSK zu erkranken bei Exposition (Bezeichnung ähnlich Punkt 6.) dann erhält man für RD :

$$RD = W(1|1) - W(1|0) = [a/(a+b)] - [c/(c+d)] = CI_E - CI_N \text{ (auch Angabe in \% möglich)}$$

$RD =$ Anteil der zusätzlichen Fälle unter den Exponierten, die man vermeiden könnte, wenn unter ihnen nur das Grundrisiko bestünde.

Beispiel: Herzinfarktstudie (Kohortenstudie), Exposition fördert die Erkrankung $RR > 1$ In einer Studie über 5 Jahre sollte untersucht werden, ob Raucher nach einem Herzinfarkt ihr Risiko zu versterben reduzieren können, wenn sie ab Herzinfarkt das Rauchen aufgeben. Zum Zeitpunkt t_0 werden 156 Raucher nach Herzinfarkt in die Studie aufgenommen. Nach 5 Jahren ergeben sich folgende Ergebnisse:

- 75 Raucher rauchten weiter (W-Raucher),
- 81 gaben auf (N-Raucher),

(Daten: ActivEpi)

		Versterben		
		1	0	
Rauchen weiter	1	27	48	75
	0	14	67	81
		41	115	156

- 41 insgesamt starben, 115 überlebten,
die WSK zu sterben für W-Raucher beträgt $27/75 = 0,36$ (36%),
die WSK zu sterben für N-Raucher beträgt $14/81 = 0,17$ (17%).
Daraus ergibt sich ein $RR = 0,36/0,17 = 2,08$ (WSK zu Versterben ist somit bei W-Rauchern doppelt so hoch, wie bei N-Rauchern),
Das attributable Risiko $RD = 0,36 - 0,17 = 0,19$ und somit beträgt die WSK zu sterben nur aufgrund des W-Rauchens 19%, d.h., rund 20 von 100 oder 1 von 5 stirbt aufgrund des W-Rauchens. Eine Größe hierfür ist **NNH** (Number needed to harm) = $1/RD = 1 / 0,19 \approx 5$ (von 5 W-Rauchern stirbt einer durch das Rauchen).
Anmerkung: RD wird auch als **ARR** (Absolute Risiko-Reduktion) bezeichnet.

Das **Populationsattributable Risiko PAR** ist der Anteil der Fälle in der Population, der nur auf die Exposition zurückgeführt werden kann. $PAR = (CI - CI_0) / CI$. Durch Vermeidung der Exposition kann dieser Anteil vermieden werden (WHO - „Global Burden of Disease“).
Im Beispiel Herzinfarktstudie ist $PAR = (41 - 156 \cdot 14/81)/41 = 14/41 = 0,3424$ (34,24%).
Über 34% der Todesfälle in der gesamten Kohorte (nicht in der gesamten Bevölkerung) könnten verhindert werden, wenn alle nach dem Herzinfarkt aufgehört hätten zu Rauchen.

Das **Attributable Risiko der Exponierten ARE** ist der Anteil der Fälle unter den Exponierten, der nur auf die Exposition zurückgeführt werden kann. $ARE = (CI_1 - CI_0) / CI_1$. Durch Vermeidung der Exposition würde dieser Anteil nicht auftreten.
Im Beispiel Herzinfarktstudie ist $ARE = (27 - 75 \cdot 14/81)/27 = 14/27 = 0,5199$ (51,99%). Fast 52% der Todesfälle unter den Exponierten W-Rauchern könnten verhindert werden, wenn sie nach dem Herzinfarkt aufgehört hätten zu Rauchen.

c. Attributable Risiken ($RR < 1$)

Das Risiko zu erkranken kann sich durch eine Exposition (z.B. Impfung) auch verringern. Dann ist $RR < 1$ und RD negativ. Das Risiko der Exponierten (Geimpften) ist jetzt gleich dem Grundrisiko (ohne Impfung) vermindert um das attributable Risiko.

Beispiel: Impfung mit AstraZeneca gegen Covid-19 Erkrankung.
Laut einer Studie von AstraZeneca (publiziert im Januar 2021, *Quelle: www.unstatistik.de*) waren insgesamt 8 Monate nach der ersten und 3 Monate nach der zweiten Impfung

- 5807 Personen mit AstraZeneca und
- 5829 mit Placebo geimpft
- 30 Geimpfte und 101 nicht Geimpfte erkrankten an Covid-19 (positiver Test + mindestens ein Symptom)

		Corona		
		1	0	
AstraZeneca	1	30	5777	5807
	0	101	5728	5829
		131	11505	11636

Die Ergebnisse lassen sich in einer 2x2 - Tafel darstellen.

Danach beträgt das Risiko zu erkranken für Geimpfte $30/5807 = 0,0052$ und für nicht Geimpfte $101/5829 = 0,0173$. Das $RR = 0,0052/0,0173 = 0,3$ und damit ist für Geimpfte das Risiko nur etwa 1/3 so groß im Vergleich zu nicht Geimpften.

Von 1000 nicht Geimpften erkranken rund 17, von 1000 Geimpften nur etwa 5. Somit sind 12 von 1000 Geimpften aufgrund der Impfung nicht erkrankt (1,2%). Oder anders ausgedrückt:

Das attributable Risiko $RD = 0,0052 - 0,0173 = - 0,0122$ und somit reduziert sich die WSK zu erkranken bei Geimpften aufgrund der Impfung um 1,2% d.h., rund 1,2 von 100 oder 1 von 82 wird aufgrund der Impfung geschützt. Eine Größe hierfür ist **NNT** (Number needed to treat) = $1/RD = 1 / 0,0122 \approx 82$ (rund 82 Personen müssen geimpft werden, um 1 Covid-19 Erkrankung durch Impfung zu verhindern).

Die Wirksamkeit einer Impfung kann durch die **Relative Risikoreduktion unter Exponierten** RRR_E beschrieben werden (auch „Prevented fraction in exposed“). Es ist der Anteil potentiell neuer Fälle unter den Exponierten, der durch die Exposition verhindert werden kann. $RRR_E = (CI_0 - CI_1) / CI_0 = 1 - RR$

Im Beispiel AstraZeneca erhält man $RRR_E = 1 - 0,3 = 0,7$. Etwa 70% der potentiell neuen Covid-19 Erkrankungen unter den Geimpften sind nicht aufgetreten infolge der Impfung.

d. Odds Ratio (OR)

Die **Odds Ratio OR** ist neben dem relativen Risiko ein weiteres Vergleichsmaß in der Epidemiologie, das auf der so genannten **Odds** basiert. Es kann in prospektiven (Kohorte) und retrospektiven Studien (Fall-Kontroll-Studie) verwendet werden. Die Odds (Chance), dass ein Ereignis (z.B. eine Krankheit K) in einem bestimmten Zeitraum eintritt, ist:

Odds (K) =: $W(K) / (1-W(K))$. Für bedingte WSK gilt analog **Odds =: $W(K|E) / (1-W(K|E))$**

Für obiges Impfbeispiel ist $Odds(1|1) = 0,0052 / (1 - 0,0052) = 0,00523$ und

$Odds(1|0) = 0,0173 / (1 - 0,0173) = 0,0176$. Damit erhält man

die **OR = $Odds(1|1) / Odds(1|0) = 0,00523 / 0,0176 = 0,297 \approx 0,3$**

In Vierfeldertafeln wird die OR als Kreuzprodukt berechnet: $OR = a \cdot d / b \cdot c$, im

Impfbeispiel: $OR = 30 \cdot 5728 / 5777 \cdot 101 = 0,2945 \approx 0,3$, wie oben.

Da es sich bei Covid-19 um eine seltene Krankheit handelt ist $RR \approx OR$ (Rare Disease Annahme).

e. Confounder und Simpson-Paradoxon

(Daten: Deutsches Ärzteblatt, 2009)

Zur Behandlung von Nierensteinen wurden zwei Therapien hinsichtlich ihres Erfolges miteinander verglichen (siehe Punkt 2.). Insgesamt erhielt man folgendes Resultat

Erfolg von Therapie A: $W(1|A) = 273/350 = 0,78$ (78%)

Erfolg von Therapie B: $W(1|B) = 289/350 = 0,826$ (82,6%)

Therapie B scheint überlegen zu sein:

- χ^2 -Homogenitätstest: $\chi^2 = 2,311$; $p = 0,1286$ n.s.
- 2-Stipro-Test für Anteile: $z = -1,5201$; $p = 0,1285$ n.s.

Es besteht kein Unterschied zwischen den Therapieerfolgen.

		Erfolg		
		1	0	
Therapie	A	273	77	350
	B	289	61	350
		562	138	700

Für die OR findet man: $OR = (273 \cdot 61) / (289 \cdot 77) = 0,75$ (0,51 ; 1,09). Der Therapieerfolg von A ist zwar zahlenmäßig kleiner als der von B, doch das C.I. enthält die „1“, daher kann ein Unterschied, in Übereinstimmung mit den Tests, nicht belegt werden.

Wird nun die Größe der Nierensteine (**Confounder**) bei der Auswertung berücksichtigt, ergibt sich folgendes Bild:

Stratum

- (1) **kleine** Nierensteine: $OR = 2,08$ (0,84 ; 5,11)
Ein Unterschied zwischen den Therapien kann bei kleinen Nierensteinen nicht belegt werden.

		Erfolg		
		1	0	
Therapie	A	81	6	87
	B	234	36	270

Stratum

- (2) **große** Nierensteine: $OR = 1,23$ (0,71 ; 2,12)
Ein Unterschied zwischen den Therapien kann auch bei großen Nierensteinen nicht belegt werden.

		Erfolg		
		1	0	
Therapie	A	192	71	263
	B	55	25	80

Wird die Größe der Nierensteine als Schichtungsvariable verwendet, kann man die OR adjustieren und so ein Confounding kontrollieren. Ein gebräuchlicher Schätzer für die adjustierte OR ist der Mantel-Haenszel-Schätzer OR_{MH} mit k = Anzahl der Schichten.

$$OR_{MH} = \frac{\sum_k \frac{a_k \cdot d_k}{n_k}}{\sum_k \frac{b_k \cdot c_k}{n_k}}$$

Im Beispiel sind zwei Schichten gegeben und die Rechnung lautet:

$$OR_{MH} = \{(81 \cdot 36 / 357) + (192 \cdot 25 / 343)\} / \{(6 \cdot 234 / 357) + (55 \cdot 71 / 343)\} = 1,447$$

Voraussetzung für diese Adjustierung ist ein nichtsignifikanter Test (Breslow-Day-Test) auf Homogenität zwischen den Schichten. Diese Rechnungen werden jetzt dem Programm OpenEpi überlassen, das hierzu den unten stehenden Ausdruck liefert. Für die adjustierte OR erhält man $OR_{MH} = 1,45$ (0,92 ; 2,29). Auch mit Adjustierung zeigt sich hier kein Unterschied zwischen den Therapieerfolgen.

Ohne Berücksichtigung der Größe der Nierensteine findet man $OR = 0,75 < 1,0$ und in den Schichten $OR = 2,08 > 1,0$ und $1,23 > 1,0$. Diesen Effekt nennt man **Simpson-Paradoxon**.

Odds-Based Estimates and Confidence Limits

Stratum	Point Estimates		Confidence Limits	
	Type	Value	Lower	Upper
1	Odds Ratio	2.077	0.844	5.110
2	Odds Ratio	1.229	0.712	2.121
Crude	Odds Ratio	0.748	0.515	1.088
Adjusted	ORMantel-Haenszel	1.447	0.916	2.286

Breslow-Day test for interaction of Odds Ratio over strata:
 chi square = 0.9535 p = 0.3288
 p greater than 0.05 does not suggest interaction. Adjusted OR can be used.

f. Fehlklassifikationen (Misclassifications)

Fehlklassifikationen entstehen durch fehlerhafte Informationen über Exposition oder Krankheit oder beides. Sie beeinflussen (verzerren) die Größe des untersuchten Effektes (z.B. RR). Wir beschränken uns hier auf folgende Rahmenbedingungen:

- dichotome Variable für Exposition und Krankheit,
- Exposition korrekt klassifiziert (typisch für Kohortenstudien),
- Fehlklassifikationen der Krankheit bei Exponierten und Nichtexponierten gleich.

Einige Gründe für Fehlklassifikationen des Krankheitsstatus sind:

- Inkorrekte Diagnose:
 - limitiertes Wissen über die Krankheit
 - komplexer diagnostischer Prozess
 - Laborfehler
 - Krankheit noch nicht erkennbar (subklinisch)
- Unvollständige Erinnerung an die Erkrankung (Fragebogen)
- Unehrlichkeit
- Codierungsfehler im Datensatz

Beispiel einer Kohortenstudie mit der Frage:

Ist Bronchitis bis zum Alter von 5 Jahren (Exposition **husten5**) ein Risikofaktor für Husten am Tag oder in der Nacht im Alter von 14 Jahren (Krankheit **husten14**)? (Daten M. Bland)

Die Daten dieser Kohortenstudie sind in der folgenden 2x2 - Tafel aufgeführt.

$W(1|1) = 26 / 273 = 0,09524$, $W(1|0) = 44 / 1046 = 0,04207$. Damit sind:

$RR = 0,09524 / 0,04207 = 2,26$ (1,42 ; 3,61)

$OR = (26 \cdot 1002) / (44 \cdot 247) = 2,40$ (1,45 ; 3,97)

		husten14		
		1	0	
bronch5	1	26	247	273
	0	44	1002	1046
		70	1249	1319

Ergebnis der Kohortenstudie (Effekt):
Das Risiko von Husten am Tag oder in der Nacht im Alter von 14 Jahren ist mehr als doppelt so hoch für Kinder mit Bronchitis bis zum Alter von 5 Jahren im Vergleich zu Kindern ohne Bronchitis in dieser Zeit.
(Effekt aufgrund beobachteter Daten)

Bei den Werten in der 2x2 - Tafel seien die beobachteten Klassifikationen für die dichotome Variable **husten14** mit Fehlklassifikationen behaftet. Fälle mit Husten wurden nicht erkannt (falsch negativ) und Fälle ohne Husten falsch positiv dokumentiert, so dass die korrekte Variable von der beobachteten Klassifikation abweicht. Bezugnehmend auf die Bezeichnungen unter 7.a seien a^* , b^* , c^* , d^* die beobachteten Werte in der 2x2-Tafel oben mit Fehlklassifikation und a , b , c , d die wahren Werte ohne Fehlklassifikation. Bei Kenntnis der WSK für "richtig-positive" und "richtig-negative" Klassifikation, Sens und Spez, lässt sich die Fehlklassifikation und damit die Verzerrung der Effektmaße (RR, OR) korrigieren.

Für Sens und Spez der Diagnose Husten bei 14-Jährigen sei aus Pilotstudien bekannt: **Sens = 0,8** und **Spez = 0,97** (hypothetische Werte). Der beobachtete Wert a^* setzt sich zusammen aus der Anzahl richtig-positiver und falsch-positiver Klassifizierungen, d.h., $a^* = a \cdot \text{Sens} + b \cdot (1 - \text{Spez})$. Da die Expositionen korrekt klassifiziert wurden gilt zudem $a^* + b^* = a + b$. Daraus folgt für $b^* = (a+b) - a^* = a \cdot (1 - \text{Sens}) + b \cdot \text{Spez}$. Aus diesen zwei Gleichungen mit zwei Unbekannten können a und b berechnet werden:
 $26 = a \cdot 0,8 + b \cdot 0,03$ und $247 = a \cdot 0,2 + b \cdot 0,97$. Auflösung der Gleichungen ergibt $a = 23,13$ und $b = 249,87$

Der beobachtete Wert d^* setzt sich zusammen aus der Anzahl richtig-negativer und falsch-negativer Klassifizierungen, d.h., $d^* = d \cdot \text{Spez} + c \cdot (1 - \text{Sens})$. Da die Expositionen korrekt klassifiziert wurden gilt zudem $c^* + d^* = c + d$.

Daraus folgt für $c^* = (c+d) - d^* = d \cdot (1 - \text{Spez}) + c \cdot \text{Sens}$. Hieraus berechnet man die zwei Unbekannten $c = 16,39$ und $d = 1029,61$

Mit den korrigierten Werten erhält man folgende 2x2 - Tafel:

		husten14		
		1	0	
bronch5	1	23	250	273
	0	16	1030	1046
		39	1280	1319

Für die Effektmaße RR und OR erhält man jetzt:
 $RR = 5,51$ (2,95 ; 10,28) und $OR = 5,92$ (3,08 ; 11,38).
Sie sind beide im Vergleich zu den beobachteten Werten mehr als doppelt so groß.

Manchmal sind die beobachteten Häufigkeiten nicht vereinbar mit den Sensitivitäts- und Spezifitätswerten. Dann sind die Ergebnisse aus den Gleichungen nicht realistisch (negative Häufigkeiten). Es kann es sich um Stichprobenfehler oder Auswirkungen von weiteren Verzerrungen handeln.

g. Korrektur von Fehlklassifikationen mit WinPepi

Nach Aufrufen von WinPepi / Compare klickt man in der Kopfleiste auf „Misclass“ und erhält diese Dialogbox.

Compare2

Compare Misclass Sample size Power Note View Saving Help Manual Finder F9
WinPepi Quit

Appraisal of effect of misclassification on 2x2 table

Back to "Comparison of..." menu

FIRST CLICK ON THE TYPE OF STUDY

- Study of total population, or representative sample
- Comparison of cases and controls
- Comparison of exposed and nonexposed

USE COLUMN A (ONLY), OR BOTH B AND C

	A	B	C
	In both	In exposed	In nonexposed
EASY ENTRY: Sensitivity (%) of disease measure	80		
Specificity (%) of disease measure	97		
	In both	In diseased	In nondiseased
Sensitivity (%) of exposure measure	100		
Specificity (%) of exposure measure	100		

2x2 TABLE

	Diseased	Not diseased
Exposed	26	247
Not exposed	44	1002

Clear all Clear sensitivity and specificity Run
Clear 2x2 table

Fehlklassifizierungen betreffen nur die Krankheit. Daher Sens und Spez für Exposition auf 100 setzen. Die Krankheit wird bei Exponierten und nicht Exponierten gleichartig klassifiziert. Daher nur Spalte „A“ ausfüllen. Die beobachteten Werte in die 2x2-Tabelle eintragen. Die Ergebnistabelle von WinPepi zeigt beobachtete und adjustierte (korrigierte) Werte der 2x2-Tafel in diesem Bronchitisbeispiel sowie beobachtete und adjustierte (korrigierte) Effektmaße.

Compare2

Compare Misclass Sample size Power Note View Saving Help Manual Finder F9
WinPepi Quit

Appraisal of effect of misclassification on 2x2 table

Back to "Comparison of..." menu

The following "adjusted" values represent the "true values" that would be expected if there were no misclassification

	Observed	Adjusted
Diseased, exposed	26	23.13
Not diseased, exposed	247	249.87
Diseased, not exposed	44	16.39
Not diseased, not exposed	1002	1029.61

Odds ratio: Observed: 2.397
Adjusted: 5.815 (or 5.923*)
Approx. 95% CI* = 3.083 to 11.377

Risk ratio: Observed: 2.264
Adjusted: 5.407 (or 5.508*)
Approx. 95% CI* = 2.951 to 10.280

* If adjusted frequencies are rounded off.

Repeat Run Print or save