Thieme

Grundlegende Maßzahlen zur Analyse 2-stufiger Merkmalsausprägungen: Risiko, Odds, Logits, Relatives Risiko, Odds Ratio

Basic Statistical Measures for Dichotomous Data Formats: Risk, Odds, Logits, Relative Risk, Odds Ratio

Autor
Markus Antonius Wirtz

Institut
Institut für Psychologie, Pädagogische Hochschule Freiburg

Schlüsselwörter
dichotome Daten, Relatives Risiko, Odds, Logits, Odds Ratio

Key words
dichotomous data, relative risk, odds, logits, odds ratio

Bibliografie
DOI https://doi.org/10.1055/s-0043-109930
Rehabilitation 2017; 56: 264–271
© Georg Thieme Verlag KG Stuttgart · New York
ISSN 0034-3536

Korrespondenzadresse
Prof. Dr. Markus Antonius Wirtz
Institut für Psychologie, Pädagogische Hochschule Freiburg
Kunzenweg 21, 79117 Freiburg
markus.wirtz@ph-freiburg.de

ZUSAMMENFASSUNG

In den Rehabilitationswissenschaften sind Merkmale häufig dichotom ausgeprägt oder werden dichotom erfasst. Für die Kommunikation von Merkmalsausprägungen oder als Entscheidungsgrundlage erweisen sich 2-gestufte Formate wie ‚ja vs. nein', ‚unauffällig vs. auffällig' oder ‚arbeitsfähig vs. nicht arbeitsfähig' oft als günstig, auch wenn die unterliegende Information als feiner gestuft angenommen werden kann. Für die statistische Darstellung der Auftretenshäufigkeit eines Merkmals werden die Definitionen und Eigenschaften der in der Rehabilitationsforschung typischerweise angewendeten Maße Risiko, Odds und Logits behandelt. Relatives Risiko, Risiko-Reduktion, Logit-Reduktion, Odds Ratio und ln(Odds Ratio) bilden den Zusammenhang zweier Merkmale ab. Odds Ratio und ln(Odds Ratio) bleiben konstant, unabhängig davon, wie häufig die Einzelmerkmale auftreten (z. B. seltene vs. häufige Erkrankung). Diese Maße bilden die unterliegende Zusammenhangsstärke unabhängig von anderen Verteilungseigenschaften ab. Aufgrund dieser sehr günstigen Eigenschaft basieren wichtige korrelationsstatistische Maßzahlen auf Odds Ratio. Es wird gezeigt, dass sowohl statistische Eigenschaften als auch Aspekte der verständlichen und fehlerfreien Interpretierbarkeit der Analysebefunde bei der Auswahl und Darstellung der Maßzahlen berücksichtigt werden müssen.

ABSTRACT

In rehabilitation science dichotomous data formats are frequently used to indicate or analyze patient characteristics. Although underlying information may be more fine graded, dichotomous or dichotomized data formats like 'yes vs. no', ‚salient vs. not salient' or ‚capable to work vs. not capable to work' are often adequate information formats for communication or decision making purposes. Definitions and properties of risk, odds and logits - typically applied in rehabilitation research - are outlined as frequency measures of occurrence. The association of two variables can be characterized by relative risk, risk reduction, logit reduction, odds ratio and ln(odds ratio). As odds ratio is independent from base rates it reflects association strength unambiguously. Hence, important correlational measures are defined as functions of odds ratio. It will be illustrated that statistical aspects as well as aspects regarding comprehensibility of the represented information are decisive to select appropriate measures and present data information appropriately.

Einleitung

Der Analyse von Merkmalen mit 2 möglichen Ausprägungen kommt in der rehabilitationswissenschaftlichen Forschung eine hohe Bedeutung zu. Viele Merkmale wie z. B. Geschlecht, Frühberentung, Diagnosestellung oder Therapieteilnahme sind von Natur aus 2-gestuft oder es liegt nahe, die Merkmale in zwei Hauptkategorien zu erfassen. Häufig werden auch feiner skalierte Merkmale (z. B. Blutdruck, BMI, Depressivität) aus klinisch-diagnostischen Gründen vereinfachend in 2 Kategorien abgebildet (z. B. Bluthochdruck vs. kein Bluthochdruck; übergewichtig vs. nicht übergewichtig; auffällige vs. unauffällige Depressivitätswerte). In der klinischen

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

Praxis ermöglicht das dichotome oder dichotomisierte Antwortformat die ökonomische Kodierung klinisch bedeutsamer Informationen und es korrespondiert oftmals sehr gut mit Entscheidungsformaten: Die Notwendigkeit einer Behandlung oder Intensivdiagnostik wird in der Regel daran festgemacht, ob sich (z. B. bei Screeningverfahren) unauffällige vs. auffällige Werte in den Gesundheitsparametern ergeben haben.

Der guten Anschaulichkeit, Kommunizierbarkeit und Nutzbarkeit dichotomer Datenformate steht ein hoher Anspruch bei der statistischen Analyse und der angemessenen Interpretation der Ergebnisse gegenüber. So muss bspw. bei der Kategorisierung einer feiner gestuften Variablen in 2 Kategorien (Dichotomisierung) beachtet werden, dass durch die Kategorisierung der Daten ein Informationsverlust resultiert und dies insbesondere eine systematische Unterschätzung von Zusammenhängen mit anderen Variablen zur Folge hat [1, 2]. Entsprechend wurden die speziellen Eigenschaften dichotomer Datenformate schon mehrfach in dieser Methodenreihe behandelt [3–5].

In diesem Beitrag stehen Maßzahlen des Zusammenhangs zweier dichotomer Merkmale bzw. der Daten in Vierfeldertafeln im Mittelpunkt. Zunächst wird die Bedeutung der Maßzahlen Risiko, Odds und Logits erläutert, deren angemessenes Verständnis grundlegend für die korrekte Interpretation der Zusammenhangsmaße Relatives Risiko und Odds Ratio sind.

Maßzahlen der Verteilung eines Merkmals in einer Gruppe: Der Unterschied zwischen Risiko, Odds und Logits

Das geschätzte **Risiko** eines Ereignisses entspricht der relativen Häufigkeit eines Ereignisses in einer repräsentativen Stichprobe des Umfangs N:

$$\text{Risiko (Ereignis)} = n(\text{Ereignis})/N$$

Das geschätzte Risiko dient der Schätzung des wahren Risikos in der Population. Das wahre Risiko in der Population entspricht der Wahrscheinlichkeit des Ereignisses.[1] Der Wertebereich des Risikos ist auf das Intervall [0;1] beschränkt. Tritt bspw. eine Frühberentung bei 200 von 1000 Patienten einer Untersuchungsgruppe auf, so beträgt das Risiko 200/1000 = 0,2 (20 %). In ▶ **Tab. 1** sind für verschiedene Ereignishäufigkeiten (N(Ereignis)) in einer Stichprobe des Umfangs N = 1000 die entsprechenden Werte des Risikos in der dritten Spalte abgetragen.

Der Koeffizient **Odds** wird in der Regel mit ‚Chance' übersetzt. Diese Übersetzung ist nicht unkritisch, da im allgemeinen Sprachgebrauch ‚Risiko' und ‚Chance' im Wesentlichen synonym gebraucht werden, wobei ‚Risiko' (z. B. Erkrankungsrisiko) eher im Zusammenhang mit unerwünschten und ‚Chance' (z. B. Heilungschance) eher im Zusammenhang mit erwünschten Ereignissen verwendet wird. Diese sprachlich-semantische Differenz ist bei der Unterscheidung der statistischen Kennzahlen überhaupt nicht hilfreich und trägt häufig zu Problemen der angemessenen Interpretation der Maße bei.

Die Bezeichnung Odds [engl. Wettquotient, Gewinnchance] stammt aus dem Bereich der Glückswetten und ist wie folgt definiert:

$$\text{Odds (Ereignis)} = n(\text{Ereignis})/n\,(\text{Gegenereignis})$$

„Gegenereignis" bezeichnet den Fall, dass das Ereignis nicht eintritt. Der Wertebereich der Odds liegt zwischen $[0; +\infty[$. Wetten gleich viele Personen auf den Sieg von A wie auf den von B, so ist der Wert der Odds gleich 1. Wenn bspw. 700 Personen auf den Sieg von A und 300 Personen auf den Sieg von B wetten, so gilt: Odds(A) = 700/300 = 2,333; Odds(B) = 300/700 = 0,429. Dass Odds für A einen Wert größer 1 bzw. für B einen Wert kleiner 1 annimmt, zeigt an, dass A größere Siegchancen eingeräumt werden als B. Deswegen wird Odds in seiner ursprünglichen Verwendung als ‚Wettquotient' bezeichnet. Für die Anwendung in der klinischen Statistik ist die daran angelehnte Bezeichnung ‚Ereignisquotient' dem Begriff der ‚Chance' vorzuziehen, um eine klarere Abgrenzung zum Begriff ‚Risiko' zu gewährleisten.

Tritt in einer Untersuchungsgruppe von N = 1000 Personen das Merkmal Frühberentung in der Hälfte aller Fälle auf, so beträgt das Frühberentungsrisiko 500/1000 = 0,5, der Ereignisquotient Odds (Frühberentung) beträgt hingegen 500/500 = 1. Treten bspw. 600 Frühberentungsfälle auf, so ist der Wert der Odds(Frühberentung) mit 600/400 = 1,5 größer als 1. Treten 400 Frühberentungsfälle auf, so ist der Wert der Odds(Frühberentung) mit 400/600 = 0,667 kleiner als 1.

Grundsätzlich gilt, dass sich die Odds eines Ereignisses und dessen Gegenereignisses reziprok zu einander verhalten:

$$\text{Odds}(\text{Ereignis}) = 1/\,\text{Odds}(\text{Gegenereignis})$$

In ▶ **Tab. 1** sind die Werte der Odds für verschiedene Ereignishäufigkeiten in einer Stichprobe des Umfangs N = 1000 abgetragen. Hier kann ein wesentliches Problem der Maßzahl Odds nachvollzogen werden. Die Werteverteilung ist deutlich asymmetrisch um den neutralen Punkt 1: Ein Sprung von N(Ereignis) von 800 auf 900 entspricht bspw. 9−4 = 5 Odds Einheiten. Betrachtet man hingegen N(Gegenereignis) so entspräche dieser Sprung einer Differenz von 0,250–0,111 = 0,139 Odds-Einheiten. Dieselbe Veränderung in den Ereignisraten wird deutlich unterschiedlich abgebildet, je nachdem welches der komplementären Alternativereignisse analysiert bzw. als Zielereignis definiert wird. Es wäre völlig unzulässig, die Zunahme von 100 auf 200 Frühberentungen statistisch anders zu werten, als die entsprechende Abnahme von 900 auf 800 Nicht-Frühberentungen, da die Informationen identisch sind.

Die Odds sind also zum einen ein inhaltlich eher unanschauliches, nicht einfach zu interpretierendes Maß. Zum anderen sind z. B. die Differenzen zwischen Odds sowohl statistisch als auch inhaltlich nicht vergleichbar. Der besondere Nutzen der Odds liegt darin, dass diese die Basis für 2 andere Maße darstellen, die statistisch sehr nützlich sind: die Logits und Odds Ratio.

[1] Im Folgenden werden idealisierte Daten verwendet, die so unter idealen Bedingungen aus Stichproben stammen könnten und theoretisch angenommene Wahrscheinlichkeiten numerisch fehlerfrei abbilden. In empirischen Untersuchungen variieren die Ergebnisse zwischen Stichproben. Entsprechend müsste streng genommen stets zwischen wahrem bzw. stichprobenunabhängigen Kennwerten und geschätzten bzw. stichprobenabhängigen Kennwerten unterschieden werden. Aus Gründen der Klarheit der deskriptiven Darstellung wird auf diese Unterscheidung und auf Fragen der statistischen Signifikanz und Konfidenzintervallen in diesem Beitrag verzichtet (s. hierzu: [6]).

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

▶ **Tab. 1** Risiken, Odds und Logits für verschiedene Ereignishäufigkeiten in einer Stichprobe von N = 1000 Fällen.

N(Ereignis)	N(Gegnereignis)	Risiko	Odds	Logit
5	995	0,005	0,005	−5,293
10	990	0,01	0,010	−4,595
100	900	0,1	0,111	−2,197
150	850	0,15	0,176	−1,735
200	800	0,2	0,250	−1,386
250	750	0,25	0,333	−1,099
300	700	0,3	0,429	−0,847
350	650	0,35	0,538	−0,619
400	600	0,4	0,667	−0,405
450	550	0,45	0,818	−0,201
500	500	0,5	1	0
550	450	0,55	1,222	0,201
600	400	0,6	1,500	0,405
650	350	0,65	1,857	0,619
700	300	0,7	2,333	0,847
750	250	0,75	3	1,099
800	200	0,8	4	1,386
850	150	0,85	5,667	1,735
900	100	0,9	9	2,197
990	10	0,99	99	4,595
995	5	0,995	199	5,293

Ein **Logit** entspricht den logarithmierten Odds:

$$\text{Logit} = \ln(\text{Odds})$$

Die Logits der Risiken bzw. die logarithmierten Odds sind in ▶ **Tab. 1** in der fünften Spalte abgetragen. Diese Logarithmierung führt dazu, dass der neutrale Punkt (Chancengleichheit = 500/500; Risiko = 0,5) nicht mehr wie für die Odds bei 1, sondern bei 0 liegt.

Risiko < 0,5, Odds < 1 und Logit < 0: Das Ereignis tritt seltener als das Gegenereignis auf.
Risiko = 0,5, Odds = 1 und Logit = 0: Das Ereignis und das Gegenereignis treten gleich häufig auf.
Risiko > 0,5, Odds > 1 und Logit > 0: Das Ereignis tritt häufiger als das Gegenereignis auf.

Zudem verteilen sich die Werte nun symmetrisch um den neutralen Punkt, sodass sich Veränderungen im positiven Bereich in gleichen zahlenmäßigen Abständen widerspiegeln wie im negativen Bereich. Der Zunahme von 100 auf 200 und der Abnahme von 900 auf 800 Nicht-Frühberentungen entspricht dieselbe Logit-Differenz: $|-1{,}386-(-2{,}197)| = |2{,}197-1{,}386| = 0{,}816$.

Der wichtigste Vorteil besteht aber darin, dass die resultierenden Werte nicht nur im Unterschied zu den Odds sondern auch zu den Risiken als metrisch oder intervallskaliert gelten können. Eine angemessen Begründung würde den Rahmen dieses Tutorials sprengen, hierzu sei auf von Urban [7] verwiesen. Dass neben den Odds auch das Risiko nicht als metrisch oder intervallskaliert gelten kann, ist allein schon durch die Tatsache begründet, dass der Wertebereich des Risikos [0;1] beschränkt ist. Dies führt dazu, dass die Werteverteilungen statistisch im Bereich der Extremwerte als ‚gestaucht' betrachtet werden müssen. So ist dieselbe Differenz zweier Risikowerte im Bereich um 0,5 bzw. 50 % als statistisch weniger bedeutsam zu betrachten als in Bereichen, die näher an den Extremwerten liegen:

- Während z. B. eine Differenz in den Frühberentungshäufigkeiten zwischen 100 und 200 bzw. 900 und 800 einer Logit-Differenz von 0,816 entspricht, entspricht eine Differenz zwischen 400 und 500 bzw. 500 und 600 lediglich einer Differenz von 0,201.
- Angenommen in einer Hochrisikogruppe liegt das Frühberentungsrisiko bei 0,5 (50 %), in einer Vergleichsgruppe bei 0,1 (10 %). Dann ist eine Absenkung des Risikos in der Hochrisikogruppe von 0,5 (50 %) auf 0,45 (45 %) als statistisch weniger bedeutsam einzuschätzen als eine Absenkung von 0,1 (10 %) auf 0,05 (5 %). Die Risikodifferenz 0,1 − 0,05 ist mit einer Logitdifferenz $-2{,}197-(-2{,}944) = 0{,}747$ mehr als 3,5-mal so groß einzuschätzen wie die Risikodifferenz 0,5–0,45 mit einer Logitdifferenz $0-(-0{,}201) = 0{,}201$.

Dass die Analyse der Unterschiede oder -veränderungen auf Basis von Risiken problematisch ist, wird auch unmittelbar dadurch deutlich, dass für den Risikowert 0,1 (10 %) aufgrund des beschränkten Wertebereiches eine Verringerung um maximal 0,1 (10 %) möglich ist. Eine Maßnahme, die in einer Gruppe mit einem Grundrisiko für Frühberentung von 0,5 (5 %) eine Verringerung um 0,2 (20 %) auf 0,3 (30 %) bewirkt, würde in einer Vergleichsgruppe mit einem Grundrisiko von 0,1 (10 %) niemals eine Verringerung des Risikos um dieselbe Differenz von 0,2 (20 %) bewirken können [2].
Diese kurze Klärung des mathematischen Hintergrunds dient der exemplarischen Erläuterung der Systematik des Zusammenhangs von Risiken und Logits: Generell gilt, dass Risiken erst transformiert werden müssen, wenn Differenzen und statistische Effekte aussagekräftig analysiert oder beurteilt werden sollen [7]. Die Transformation von Risiken in Logits ist auch der zentrale Ansatz, um multivariate statistische Verfahren wie die logistische Regression (zur Anwendung dieses Verfahrens in der Rehabilitation: [5]) oder Skalierungsmethoden (zur Anwendung dieser Verfahren in der Rehabilitation: [4]) im Falle dichotomer oder polytomer (mehrstufiger) kategorialer Antwortformate anwenden zu können.

[2] Vertiefende Detailbetrachtung Um dies genauer nachvollziehen zu können, ist es an dieser Stelle notwendig kurz auf die mathematischen Berechnungsdetails einzugehen. Der Verringerung des Risikos von 0,5 auf 0,3 (Logitdifferenz $0-(-0{,}847) = 0{,}847$) entspräche für das Ausgangsrisiko von 0,1 diejenige Risikoverringerung, die den Logitwert von 0,1 um den Wert 0,847 absenkt. Es muss also die Frage beantwortet werden: Für welches x gilt: Logit(0,1) – Logit (x) = 0,847. Da aus ▶ **Tab. 1** bekannt ist, dass Logit(0,1) = −2,197 ist, muss gelten: Logit(x) = −2,197−0,847 = −3,044. Der Wert von x kann über die der Logit-Funktion inverse Funktion $(\exp(-3{,}044)/1+\exp(-3{,}044)) = 0{,}045$ bestimmt werden. Der Risikodifferenz von 0,5 und 0,3 entspricht also eine Risikodifferenz von 0,1 und 0,045, da sich für die beiden Differenzen dieselbe Logitdifferenz von 0,847 ergibt.

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

Das Risiko entspricht der relativen Auftretenshäufigkeit eines Merkmals. Der Ereignisquotient Odds gibt das Verhältnis der Auftretenshäufigkeit des Merkmals geteilt durch die Auftretenshäufigkeit des Gegenereignisses an. Logits sind logarithmisch transformierte Odds-Werte: Im Unterschied zu den Risiken und Odds sind Logits metrisch skaliert, sodass für diese auch Differenzen begründet interpretiert und verglichen werden können.

Maßzahlen des Zusammenhangs zweier dichotomer Merkmale: Relatives Risiko, Risikoreduktion, Logitreduktion und Odds Ratio

Maßzahlen des Unterschieds der Häufigkeit eines dichotomen Merkmales zwischen 2 Gruppen sind auch Maßzahlen des Zusammenhangs bzw. der Korrelation zweier dichotomer Merkmale [8, 9]. Besteht bspw. ein Unterschied in der Häufigkeit von Frühberentungen (Merkmal 1: Frühberentung ja vs. nein) in einer rehabilitativ behandelten vs. nicht behandelten Gruppe (Merkmal 2: Rehabilitation ja vs. nein), so besteht auch ein Zusammenhang zwischen Frühberentung und Teilnahme an einer Rehabilitation. Unterscheiden sich Männer und Frauen in der Häufigkeit einer Erkrankung, so besteht auch ein Zusammenhang zwischen Geschlecht und Erkrankungshäufigkeit. Um zu zeigen, weshalb die Begriffe Risiko und Odds für Maße des Zusammenhangs grundlegende Bedeutung zukommt, soll im Folgenden der Zusammenhang eines Risikofaktors und eines Zielkriteriums betrachtet werden. Es soll die diagnostische Frage beantwortet werden, in welchem Maße die Ausprägung eines untersuchten Risikofaktors mit der Ausprägung des Zielkriteriums korrespondiert. Beispiele für mögliche Paare von Risikofaktor und Zielkriterium könnten ‚Tätigkeitsfeld A vs. B' und ‚Frühberentung: ja vs. nein', ‚Behandlung- vs. Kontrollgruppe' und ‚Besserung: ja vs. nein' ‚Alkoholkonsum: normal vs. kritisch' und ‚Leberwerte: auffällig vs. unauffällig' sein. Bei diesen Beispielen wird jeweils ein systematischer Einfluss des Risikofaktors auf das Zielkriterium unterstellt. Im Folgenden wird jedoch nur der Zusammenhang fokussiert, für den Nachweis eines kausalen Zusammenhangs wäre ein experimentelles Design erforderlich [6]. ▶ **Tab. 2** zeigt 4 exemplarische Datenverteilungen.

- Beispiel A: Hier handelt es sich um ein Risikomerkmal, das in der Hälfte aller Fälle vorliegt, zudem tritt das Zielereignis in 50 % der Fälle auf. Wird eine Stichprobe von Busfahrern untersucht, so würde dies z. B. bedeuten, dass 50 % auf ergonomisch nicht angepassten Sitzen fahren müssen (Risikofaktor) und 50 % der Busfahrer frühberentet werden.
- Beispiel B: Im Vergleich zu Beispiel A liegt das Risikomerkmal „ergonomisch nicht angepasste Sitze" mit 18 % deutlich seltener vor. Es kommt mit 10 % auch deutlich seltener zu Frühberentungen.
- Beispiel C: Im Vergleich zu Beispiel A liegt das Risikomerkmal „ergonomisch nicht angepassten Sitze" mit 25 % deutlich seltener vor. Es werden aber auch hier 50 % der Busfahrer frühberentet.
- Beispiel D: Ähnlich wie in Beispiel B liegt hier das Risikomerkmal „ergonomisch nicht angepasste Sitze" mit 13 % selten vor und es kommt mit 10 % auch vergleichsweise selten zu Frühberentungen.

In allen Beispielen haben 90 % der Personen mit unauffälligen Werten im Zielkriterium ‚Frühberentung' auch unauffällige Werte im Risikofaktor (Spezifität: Anteil korrekt vorhergesagter Nicht-Merkmalsträger). In den Beispielen A und B haben 90 % der Personen und in den Beispielen C und D haben 40 % der Personen mit auffälligen Werten im Zielkriterium ‚Frühberentung' auffällige Werte im Risikofaktor (Sensitivität: Anteil korrekt vorhergesagter Nicht-Merkmalsträger). Zur genauen Bedeutung und den Eigenschaften von Sensitivität und Spezifität sei auf den entsprechenden Beitrag in dieser Methodenreihe verwiesen [3]. Aufgrund der geringeren

▶ **Tab. 2** In der oberen Tabelle treten auffällige Werte im Zielkriterium häufig auf (in 50 % der Fälle). In der unteren Tabelle liegt der Anteil lediglich bei 10 %. In den Beispielen A und C beträgt die Sensitivität der kritischen Ausprägung des Risikofaktors für die kritische Ausprägung des Zielkriteriums 90 % (stärkerer Zusammenhang). In den Beispielen B und D beträgt die Sensitivität 40 % (schwächerer Zusammenhang).

		Stärkerer Zusammenhang Spezifität: 90 % Sensitivität: 90 %		**Schwächerer Zusammenhang Spezifität: 90 % Sensitivität: 40 %**	
		Zielkriterium			
		Beispiel A		Beispiel C	
		nein bzw. unauffällig	ja bzw. auffällig	nein bzw. unauffällig	ja bzw. auffällig
Risikofaktor	nein bzw. unauffällig	a = 450	b = 50	a = 450	b = 300
	ja bzw. auffällig	c = 50	d = 450	c = 50	d = 200
		Beispiel B		Beispiel D	
Risikofaktor	nein bzw. unauffällig	a = 810	b = 10	a = 810	b = 60
	ja bzw. auffällig	c = 90	d = 90	c = 90	d = 40

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

▶ **Tab. 3** Berechnungsformeln, Aussage und Ausprägung der im Text erläuterten Kennwerte für die Beispiele A, B, C und D in ▶ **Tab. 2**.

Kennwert	Formel	Aussage	A	B	C	D
Risiko($RF_{unauffällig}$)	b/(a + b)	Anteil der Merkmalsträger im Zielkriterium bei unauffälliger Ausprägung des Risikofaktors	0,10	0,012	0,40	0,07
Risiko($RF_{auffällig}$)	d/(c + d)	Anteil der Merkmalträger im Zielkriterium bei auffälliger Ausprägung des Risikofaktors	0,90	0,50	0,80	0,31
Relatives Risiko	$\frac{d(a+b)}{b(c+d)}$	Faktor, um den sich das Risiko bei auffälliger Ausprägung des Risikofaktors in Referenz zu unauffälliger Ausprägung des Risikofaktors verändert	9	41	2	4,46
Absolute Risikoreduktion	$d(a+b)-d(c+d)$	Differenz des Risikos bei auffälliger und bei unauffälliger Ausprägung des Risikofaktors	0,80	0,49	0,40	0,24
Odds($RF_{unauffällig}$)	b/a	Odds (Risikoquotient) für Merkmalsträger im Zielkriterium, bei unauffälliger Ausprägung des Risikofaktors	0,1	0,01	0,6	0,07
Odds ($RF_{auffällig}$)	d/c	Odds (Risikoquotient) für Merkmalsträger im Zielkriterium, bei auffälliger Ausprägung des Risikofaktors	9	1	4	0,4
Odds Ratio (Quotenverhältnis)	(d/c)/(b/a) bzw. (a·d)/(b·c)	Faktor, um den sich die Odds bei auffälliger Ausprägung des Risikofaktors in Referenz zu unauffälliger Ausprägung des Risikofaktors verändern	81	81	6	6
ln(Odds Ratio)	ln(OR)	Logarithmiertes Quotenverhältnis	4,39	4,39	1,79	1,79
Loggit($ODDS_{auffällig}$)	ln(b/a)	Logarithmierte Odds bei unauffälliger Ausprägung des Risikofaktors	−2,20	−2,30	−0,41	−2,60
Loggit($ODDS_{unauffällig}$)	ln(d/c)	Logarithmierte Odds bei unauffälliger Ausprägung des Risikofaktors	2,20	0	1,39	−0,81
Logitreduktion	ln(d/c) – ln(b/a)	Differenz der Logits bei auffälliger und bei unauffälliger Ausprägung des Risikofaktors	4,40	2,30	1,80	1,79

Sensitivität in den Beispielen C und D ist der Merkmalszusammenhang geringer als in den Beispielen A und B. Weiterhin treten in den Beispielen A und C auffällige Ausprägungen im Zielkriterium ‚Frühberentung' mit 50 % deutlich häufiger auf als in den Beispielen B und D mit 10 %.

In ▶ **Tab. 3** sind für jede Ausprägung des Risikofaktors die numerischen Ausprägungen des Risikos, der Odds und der Logits angegeben. Aus diesen Werten können die Zusammenhangsmaße Relatives Risiko und Odds Ratio direkt ermittelt werden.

Das **Relative Risiko (RR)** gibt den Quotienten der Risiken der beiden Vergleichsgruppen an. Das Risiko für die Ausprägung des Risikofaktors, deren Effekt abgebildet werden soll, steht dabei im Zähler. Das Risiko der Referenzbedingung steht im Nenner.

$$RR = \frac{Risiko(RF_{auffällig})}{Risiko(RF_{unauffällig})} = \frac{\frac{d}{c+d}}{\frac{b}{a+b}} = \frac{d(a+b)}{b(c+d)}$$

Das Relative Risiko entspricht dem Faktor, um den sich das Risiko bei auffälliger Ausprägung des Risikofaktors im Vergleich zur Referenzbedingung verändert. Das höchste Relative Risiko ergibt sich für Beispiel B: da das Risiko für ‚Frühberentung' bei auffälliger Ausprägung ‚nicht ergonomisch angepasste Sitze' mit 0,50 41-mal größer ist als in der Referenzbedingung ‚ergonomisch angepasste Sitze' mit 0,012.

Der Wertebereich des Relativen Risikos liegt zwischen 0 und (1/Risiko($RF_{unauffällig}$)). Dies bedeutet, dass das Relative Risiko nach Kenntnis des Risikos in der Referenzbedingung, ergonomisch angepasste Sitze' in Beispiel A nicht größer als 1/0,1 = 10, in Beispiel B nicht größer als 1/(10/820) = 82, in Beispiel C nicht größer als 1/0,4 = 2,5 und in Beispiel D nicht größer als 1/(60/870) = 14,5 werden kann. Diese Werte können nicht überschritten werden, da für die jeweiligen Werte das Risiko bei auffälligem Risikofaktor 100 % beträgt. Hieraus resultieren 2 zentrale Probleme des Relativen Risikos als Maß des Zusammenhangs eines Risikofaktors und eines Zielkriteriums.

Zum einen bildet sich derselbe Effekt eines Risikofaktors in unterschiedlichen Werten ab, je nachdem wie hoch das Risiko in der Kontrollbedingung ist. Angenommen, es wird der Vorhersagewert des kritischen Alkoholkonsums für auffällige Leberwerte untersucht. Würden die Daten in ▶ **Tab. 2** für dieses Beispiel gelten, so hätte dies folgende Bedeutung: In Studie A werden bereits leicht erhöhte Leberwerten als ‚auffällig' gewertet, in Studie B werden nur deutlich erhöhte Leberwerte als ‚auffällig' klassifiziert. In Studie A ergibt sich somit eine systematisch höhere Grundrate auffälliger Leberwerte als in Studie B. Entsprechend werden in Studie B – unter ansonsten gleichen Bedingungen – deutlich höhere Werte des Relativen Risikos gemessen, da die Grundrate im Nenner hier kleiner ist.

Zum anderen ergeben sich insbesondere für seltene Merkmalshäufigkeiten oft Interpretationsprobleme, da auf den ersten Blick beeindruckende Werte des Relativen Risikos mit empirisch wenig bedeutsamen Effekten einhergehen. Angenommen, von 100 000 Personen, die ein bestimmtes Lebensmittel regelmäßig verzehren, kann bei 6 Personen ein bestimmtes Gesundheitsproblem nachgewiesen werden (Risiko = 0,00006). Bei 100 000 Personen, die dieses Lebensmittel nicht regelmäßig verzehren, tritt dieses Problem nur bei 2 Personen auf (Risiko = 0,00002). Dann zeigt das Relative

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

Risiko mit dem Wert 3 eine Verdreifachung des Risikos an. Bei der Rezeption dieses Studienbefundes kann beim Bericht der Verdreifachung des Risikos der Eindruck eines gravierenden Gesundheitsrisikos entstehen. Deswegen ist das Relative Risiko bei seltenen Ereignissen um die Angabe der Information der absoluten Häufigkeit oder der Prozentzahl (bei 6 von 100 000 Personen bzw. 0,006 % aller Personen, die das Lebensmittel konsumieren, tritt das Problem auf) oder die Absolute Risiko-Reduktion zu ergänzen. Die Absolute Risikoreduktion (ARR) bzw. Risikoerhöhung entspricht einfach dem Betrag der Differenz der beiden Risiken. Im Beispiel beträgt der Wert 0,00006 – 0,00002 = 0,00004. Dies entspricht einer Erhöhung des prozentualen Anteils um 0,004 %, was als Information deutlich weniger beeindruckend wirkt als eine Verdreifachung des Risikos. Die Absolute Risikoreduktion ist deswegen als wichtige Information bei der Dokumentation von Interventionseffekten (Differenz des Risikos der Interventionsgruppe und des Risikos in der Kontrollbedingung) stets zu berichten.

In den Beispieldaten in ▶ **Tab. 2** wird die kritische Abhängigkeit der Maßzahlen von der Grundrate ebenfalls deutlich. In Beispiel B scheint der Risikofaktor ‚ergonomische Anpassung der Sitze' gemäß des Relativen Risikos deutlich stärker mit dem Zielkriterium ‚Frühberentung' assoziiert zu sein als in Beispiel A (RR = 41 vs. 9). Der hohe Wert in B ist jedoch insbesondere durch die geringe Ereignisrate (0,012) bei unauffälliger Ausprägung des Risikofaktors bedingt. Die Absolute Risikoreduktion (z. B. Abnahme des Risikos, wenn Sitze ergonomisch angepasst werden) weist hingegen für Beispiel A mit 0,8 (80 %) einen deutlich stärkeren Effekt aus als für Beispiel B mit 0,49 (49 %), sodass je nach Maßzahl ein unterschiedlicher Eindruck der Stärke des Zusammenhangs entsteht. Dieselbe Problematik zeigt sich ebenfalls etwas abgeschwächt beim Vergleich der Daten in Beispiel C und D: Hier verringert sich das Risiko für Frühberentung in C um 0,40 (40 %) und in D um 0,24 (24 %).

Die Berechnung der **Absoluten Risikoreduktion** ist für die Bewertung des Effekts hilfreich, da diese die Veränderung der Anteile oder der daraus abgeleiteten Prozentwerte durchaus anschaulich abbildet. Das intuitive Verständnis der Unterschiede von Anteilswerten kann jedoch ebenfalls mit Fehlbeurteilungen einhergehen, wenn Effekte aus unterschiedlichen Datenverteilungen verglichen werden. Wie oben erläutert wurde, sind Differenzen von Risiken aufgrund des nicht metrischen Datenniveaus statistisch nicht vergleichbar, da sich derselbe Effekt umso schwächer in Risikodifferenzen abbildet, je weiter die Risiken vom Wert 0,5 abweichen. Deswegen sollten beim statistischen Vergleich von Effekten die Differenzen der metrisch skalierten Logit-Werte zugrunde gelegt werden. Die **Logitreduktion** oder **Logiterhöhung** ermöglicht einen fairen Vergleich von Effekten für verschiedene Grundraten. Dieser korrigierende Effekt wird in den Datenbeispielen C und D deutlich: Während die Absolute Risikoreduktion für die Frühberentung mit 0,40 (40 %) und 0,24 (24 %) für das Beispiel C einen deutlich größeren Effekt anzeigt, wird durch die Logitreduktion von 1,80 bzw. 1,79 in beiden Beispielen eine quasi identische Zusammenhangsstärke angezeigt.

Für alle bisher behandelten Maße gilt, dass je nach Maß andere Schlussfolgerungen bei der vergleichenden Beurteilung der Zusammenhangsstärken in den 4 Beispielen resultieren. Berücksichtigt man das Konstruktionsrational der Beispiele, so unterscheiden sich jedoch sowohl A und B als auch C und D nur in der Grundrate mit der das Zielkriterium auffällige Werte annimmt bzw. Personen frühberentet werden. In A und B erfolgen Frühberentungen mit 50 % häufig und in C und D mit 10 % selten. Der tatsächliche Zusammenhang mit dem Zielkriterium ‚Frühberentung' bleibt in A und B bzw. in C und D jedoch aufgrund identischer Sensitivität und Spezifität gleich. Diese Invarianz der Zusammenhangsstärke – unabhängig von der Grundrate im Risikomerkmal – wird durch **Odds Ratio** (OR; Quotenverhältnis) abgebildet, das ähnlich wie das Relative Risiko definiert ist, jedoch werden hier die Odds anstatt der Risiken ins Verhältnis gesetzt:

$$OR = \frac{Odds(RF_{auffällig})}{Odds(RF_{unauffällig})} = \frac{d/c}{b/a} = \frac{a \cdot d}{b \cdot c}$$

Für Beispiel A und B ist Odds Ratio gleich 81: Das bedeutet, dass der Ereignisquotient Odds im Falle eines auffälligen Wertes im Risikofaktor um den Faktor 81 höher ausgeprägt ist, als wenn ein unauffälliger Wert im Risikofaktor vorliegt: Fährt ein Busfahrer auf einem ergonomisch nicht angepassten Sitz, so erhöht sich der Ereignisquotient für das Merkmal Frühberentung um den Faktor 81. Der durch die geringere Sensitivität (die Merkmalsausprägung ‚Frühberentung: ja' korrespondiert schwächer mit der Merkmalsausprägung ‚ergonomisch angepasste Sitze: nein') bedingte schwächere Zusammenhang in den Beispielen C und D spiegelt sich in einem geringeren Odds Ratio von 6 wider.

Wie für die Odds liegt der Wert von Odds Ratio zwischen 0 und $+\infty$. Besteht kein Zusammenhang zwischen den Merkmalen, so sind die Odds für beide Stufen des Risikomerkmals gleich und es ergibt sich eine Odds Ratio von 1. Für Werte kleiner 1 gilt: Auffällige Werte im Risikofaktor gehen eher mit unauffälligen Werten im Zielkriterium einher (negativer Zusammenhang). Für Werte größer 1 gilt: Auffällige Werte im Risikofaktor gehen eher mit auffälligen Werten im Zielkriterium einher (positiver Zusammenhang). Genau wie bei den Odds ist die Werteverteilung um den neutralen Punkt 1 extrem unsymmetrisch: Während negative Zusammenhänge im Bereich (0;1) abgebildet werden, erstreckt sich der Wertebereich für positive Zusammenhänge von 1 bis $+\infty$. Diese Asymmetrie kann – genau wie bei den Odds – durch Logarithmierung korrigiert werden. Für die ln(OR) liegt metrisches Skalenniveau vor, sodass die Differenzen verschiedener **ln(OR)** begründet verglichen werden dürfen [7–9] [3].

[3] Vertiefende Detailbetrachtung In den Datenbeispielen liegt der Wert von ln(OR) für A und B bei 4,39 und für C und D bei 1,79. Eine Vierfeldertafel, für die die Zusammenhangsstärke genau in der Mitte liegt, müsste also der Wert von ln(OR) 3,09 betragen. Dies wäre der Fall, wenn Odds Ratio = exp(3,09) = 22 entsprechen würde. Während Unsicherheitsbereiche oder Konfidenzintervalle für Odds Ratio unsymmetrisch um den Stichprobenkennwert verteilt sind, sind diese für ln(OR) symmetrisch. Für die Daten in Beispiel C beträgt bspw. das 95 %-Konfidenzintervall (4,23; 8,45) um den Stichprobenkennwert OR = 6. Für ln(OR) = 1,79 ist die Distanz zu den Konfidenzgrenzen (1,44; 2,14) mit 0,35 in positiver und negativer Richtung identisch.

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

Relatives Risiko: Faktor, um den sich das Risiko für einen auffälligen Wert im Zielkriterium verändert, wenn eine Person statt eines unauffälligen einen auffälligen Wert im Risikofaktor aufweist.
Odds Ratio: Faktor, um den sich der Ereignisquotient Odds für einen auffälligen Wert im Zielkriterium verändert, wenn eine Person statt eines unauffälligen einen auffälligen Wert im Risikofaktor aufweist.
ln(OR): Logarithmiertes Odds Ratio. Metrisches Assoziationsmaß, für das der Vergleich von Differenzen aussagekräftig ist.
Relatives Risiko < 1, Odds Ratio < 1 bzw. ln (Odds Ratio) < 0: Auffällige Werte im Risikofaktor (bzw. in diesem Falle eher ‚Schutzfaktor') gehen mit unauffälligen Werten im Zielkriterium einher.
Relatives Risiko = 1, Odds Ratio = 1 bzw. ln (Odds Ratio) = 0: Werte im Risikofaktor und Zielkriterium stehen nicht in Zusammenhang.
Relatives Risiko > 1, Odds Ratio > 1 bzw. ln (Odds Ratio) > 0: Auffällige Werte im Risikofaktor gehen mit auffälligen Werten im Zielkriterium einher.

Mittels der Datenbeispiele wurde exemplarisch demonstriert, dass Odds Ratio bei identischer Zusammenhangsstärke konstant bleibt, auch wenn die Grundhäufigkeit des Zielkriteriums variiert. Die Invarianz von Odds Ratio gilt auch, wenn die Grundhäufigkeit des Risikomerkmals oder die Grundhäufigkeit sowohl der Risiko- als auch der Zielvariablen simultan variieren [10]. Angenommen, es wird der Zusammenhang von Geschlecht und auffälligen Depressivitätswerten bestimmt. Studie 1 werde im November durchgeführt, sodass auffällige Depressivitätswerte häufiger auftreten als in Studie 2, die den Zusammenhang in Juli untersucht. Zudem sei dar Anteil von Frauen vs. Männern in beiden Studien verschieden. Für alle Maßzahlen, die das Risiko verrechnen, also insbesondere das Relative Risiko, variieren die Werte auch wenn der zugrunde liegende Zusammenhang zwischen Depressivität und Geschlecht identisch ist, da sich die Grundraten der Merkmale zwischen den Studien unterscheiden. Odds Ratio und ln(OR) bleiben jedoch in beiden Studien konstant. Angenommen, für Männer und Frauen wird der Zusammenhang des regelmäßigen Kontakts mit einem bestimmten Gefahrenstoff und dem Auftreten allergischer Reaktionen vergleichend untersucht. Männer haben zu 50 % regelmäßig Kontakt mit dem Gefahrenstoff, Frauen jedoch nur zu 10 %. Auch wenn für beide Geschlechter der regelmäßige Kontakt in gleichem Maße für die Entwicklung allergischer Reaktionen fördert, so wird das Relative Risiko für Frauen aufgrund der extremeren Grundrate geringer ausfallen als für Männer. Die geschlechtsinvariante Zusammenhangsstärke oder Risikoerhöhung wird lediglich valide durch identische Werte des Odds Ratio bzw. des ln(OR) für beide Geschlechtsgruppen abgebildet.

Insbesondere in Querschnittsstudien ist der Anteil von Stichprobenmitgliedern pro Vergleichsgruppe oft beliebig oder dieser Anteil variiert zwischen Studien [6]: Da risikobasierte Maße von dieser Grundrate systematisch beeinflusst werden, sind diese in der Regel nicht geeignet, den Zusammenhang valide abzubilden. Bleibt die zugrunde liegende Zusammenhangsstärke identisch, so bleibt die Ausprägung von Odds Ratio und ln(OR) auch bei variierenden Grundraten der Untersuchungsmerkmale stabil.

Obwohl Odds Ratio und ln(OR) in Bezug auf die Zusammenhangsstärke somit sehr vorteilhafte Eigenschaften aufweisen, sind trotzdem Probleme bei deren Interpretation zu berücksichtigen. So kann Odds Ratio als Faktor interpretiert werden, um den sich die Odds in Abhängigkeit von der Ausprägung des Risikofaktors verändern. Findet sich bspw. in einer epidemiologischen Studie ein Odds Ratio von 2,5 zwischen Rauchen und koronarer Herzerkrankung, so bedeutet dies, dass Rauchen mit einem um den Faktor 2,5 erhöhten Ereignisquotienten Odds einhergeht. Ein zentrales Problem liegt hier in der mangelnden Anschaulichkeit der Odds: Nicht selten wird unzulässiger Weise Odds im Sinne von ‚Chance' mit dem Begriff ‚Risiko' implizit oder explizit gleichgesetzt. Der oben ausführlich erläuterte Unterschied zwischen Risiko und Odds ist jedoch erheblich, und gemäß Odds Ratio wird in der Regel gemäß der numerischen Ausprägung eine systematisch und erheblich höhere Zusammenhangsstärke widergespiegelt als durch das Relative Risiko. Entsprechend ist große Vorsicht geboten, Odds Ratio nicht fälschlicherweise als Faktor, um den sich das ‚Risiko' verändert, zu interpretieren. Da Rezipienten von Statistiken in der Regel kein valides Konzept von Odds haben, ist die zusätzliche Angabe von Risiken und absoluten Häufigkeiten – trotz aller in diesem Beitrag problematisierten statistischen Probleme dieser Maße – unbedingt empfehlenswert, um möglichen Fehlinterpretationen vorzubeugen.

KERNBOTSCHAFT

Dichotome Datenformate werden in der rehabilitationswissenschaftlichen Forschung häufig zur Beschreibung und Analyse empirischer Verteilungen und Merkmalszusammenhänge verwendet. Für die angemessene Analyse und Ergebnisinterpretation muss zwischen Maßen unterschieden werden, die auf Risiken vs. Ereignisquotienten (Odds) basieren. Obwohl die Bedeutung von Risiken als relative Ereignishäufigkeiten am einfachsten inhaltlich nachvollzogen werden kann, besitzen Odds viele vorteilhafte Eigenschaften. Aus den Odds können Logits als metrische Kenngrößen der Ereignisraten ermittelt werden und das Verhältnis zweier Odds (Odds Ratio) bildet die Zusammenhangsstärke zwischen 2 dichotomen Merkmalen eindeutiger ab als der Quotient der Risiken (Relatives Risiko). In Interventionsstudien oder prospektiven Kohortenstudien dokumentiert das Relative Risiko oder die Absolute Risikoreduktion die Effekte von Einflussmerkmalen jedoch valider und verständlicher. Bei der Auswahl und Darstellung der Maßzahlen müssen generell neben statistische Eigenschaften auch Aspekte der verständlichen und fehlerfreien Interpretierbarkeit der Analysebefunde berücksichtigt werden.

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

Koordinatoren der Reihe „Methoden in der Rehabilitationsforschung“

Prof. Dr. Dr. Hermann Faller, Würzburg; Prof. Dr. Thomas Kohlmann, Greifswald; Prof. Dr. Markus Antonius Wirtz, Freiburg; Prof. Dr. Dr. Christian Zwingmann, Siegburg.
Interessenten, die einen Beitrag zur Reihe beisteuern möchten, werden gebeten, vorab Kontakt aufzunehmen: h.faller@uni-wuerzburg.de.

Interessenkonflikt

Der Autor gibt an, dass kein Interessenkonflikt besteht.

Literatur

[1] MacCallum RC, Zhang S, Preacher KJ et al. On the practice of dichotomization of quantitative variables. Psychological Methods 2002; 7: 19–40

[2] Ulrich R, Wirtz M. On the correlation of a naturally and an artificially dichotomized variable. British Journal of Mathematical and Statistical Psychology 2004; 57: 235–251

[3] Faller H. Sensitivität, Spezifität, positiver und negativer Vorhersagewert. Serie Methoden in der Rehabilitationsforschung. Rehabilitation 2005; 44: 44–49

[4] Wirtz M, Böcker M. Eigenschaften und Nutzen des Rasch-Modells in der klinischen Diagnostik. Rehabilitation 2007; 46: 238–245

[5] Muche R. Die logistische Regression - ein vielseitiges Analyseinstrument rehabilitationswissenschaftlicher Forschung. Rehabilitation 2008; 48: 56–62

[6] Benesch M, Raab-Steiner E. Klinische Studien lesen und verstehen. Wien: Facultas; 2013

[7] Urban D. Logit Analyse. Stuttgart: Lucius; 1998

[8] Agresti A. Categorical Data Analysis. New York: Wiley; 2002

[9] Fleiss JL. Effect sizes for dichotomous data. In: Cooper H, Hedges LV, Valentine JC.eds. The Handbook of Research Synthesis and Met-Analysis. New York: Russell Sage; 2009: 221–236

[10] Wirtz M, Ulrich R. Bivariate deskriptive Statistik. In: Holling H, Schmitz B, (Hrsg.). Handbuch der Psychologie – Methoden und Evaluation. Göttingen: Hogrefe; 2010: 288–301

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.