Die Vierfeldertafel

R. Bender[1], St. Lange[2]

[1] Fakultät für Gesundheitswissenschaften, AG Epidemiologie und medizinische Statistik, Universität Bielefeld
[2] Abteilung für Medizinische Informatik, Biometrie und Epidemiologie der Ruhr-Universität Bochum

Eine Vierfeldertafel ist eine (2×2)-Anordnung absoluter Häufigkeiten, die sich bei einer gleichzeitigen Betrachtung von zwei binären Merkmalen ergibt. Ein binäres Merkmal ist eine Variable mit nur zwei möglichen Ausprägungen (zum Beispiel *Krankheit* ja/nein, *Behandlung* ja/nein, *Erfolg* ja/nein, *Diagnose* positiv/negativ). Haben die betrachteten Merkmale mehr als zwei Ausprägungen ergibt sich der allgemeine Fall einer (r×c)-Kontingenztafel. Die wichtigsten Maße zur Beschreibung der Zusammenhänge zwischen qualitativen Daten lassen sich jedoch am besten anhand einer Vierfeldertafel darstellen. Häufige Anwendungen sind der Risikovergleich zweier Gruppen und die Evaluierung diagnostischer Tests. In **Tab.1** finden sich beispielhaft die Häufigkeiten von Diabetikern mit und ohne Entwicklung einer Neuropathie innerhalb von 5 Jahren der Interventions- und der Kontrollgruppe des Diabetes Control and Complications Trial (DCCT) (7). Die Intervention bestand hierbei in der Anwendung einer intensivierten Insulintherapie im Vergleich zur gewöhnlichen Insulintherapie (Kontrolle).

Tab.1 Vierfeldertafel zur Untersuchung des Effekts einer intensivierten Insulintherapie auf die Entwicklung einer Neuropathie in 5 Jahren bei 622 Diabetikern (7)

		Neuropathie		
		ja	**nein**	Summe
Gruppe	**Kontrolle**	**52**	**255**	307
	Intervention	**21**	**294**	315
Summe		73	549	622

Die Grundlage aller Maße zur Beschreibung qualitativer Daten bildet die **Wahrscheinlichkeit.** Eine Wahrscheinlichkeit quantifiziert die Eintrittshäufigkeit eines Ereignisses mit Hilfe von Werten aus dem Intervall [0,1]. Sehr wahrscheinliche Ereignisse besitzen eine Wahrscheinlichkeit nahe an 1, sehr unwahrscheinliche Ereignisse eine Wahrscheinlichkeit nahe an 0. Wahrscheinlichkeiten können durch relative Häufigkeiten geschätzt werden. So beträgt die (geschätzte) Wahrscheinlichkeit, in 5 Jahren eine Neuropathie zu entwickeln in der Kontrollgruppe 52/307 = 16,9% und in der Interventionsgruppe 21/315 = 6,7%. Ein **Risiko** ist die Wahrscheinlichkeit für ein unerwünschtes Ereignis (zum Beispiel Krankheit, Tod). In einem verallgemeinerten Sinn wird der Begriff »Risiko« aber auch als Synonym für andere Maße zur Quantifizierung der Eintrittshäufigkeit unerwünschter Ereignisse verwendet. Dies ist häufig verwirrend, da unterschiedliche Maße auch unterschiedliche Zahlen ergeben können.

Eine **Chance** ist das Verhältnis der Wahrscheinlichkeit, dass ein Ereignis eintritt, zur Wahrscheinlichkeit, dass das Ereignis nicht eintritt. Sei p eine Wahrscheinlichkeit, so ist die zugehörige Chance gegeben durch $odds = p/(1\text{-p})$. In Worten wird eine Chance meist nicht als Dezimalzahl sondern durch ein (gerundetes) Verhältnis angegeben. Sei zum Beispiel $p = 2/3$, so entspricht das einer Chance von 2:1. Zu einer gegebenen Chance kann man die zugehörige Wahrscheinlichkeit berechnen durch $p = odds/(1+odds)$. Man kann damit die Eintrittshäufigkeit eines Ereignisses wahlweise als Wahrscheinlichkeit oder auch als Chance darstellen. Beide Darstellungsformen sind mathematisch äquivalent und ineinander umrechenbar. Nach **Tab.1** beträgt die (geschätzte) Chance, in 5 Jahren eine Neuropathie zu entwickeln, in der Kontrollgruppe 52/255 = 0,204 = 1:5 und in der Interventionsgruppe 21/294 = 0,071 = 1:14.

Um die Risiken zweier Gruppen zu vergleichen gibt es eine Reihe von Maßen, die im Folgenden kurz erläutert werden. Ein **relatives Risiko** (RR) ist das Verhältnis zweier Risiken. Ist zum Beispiel q das Risiko der Kontrollgruppe und p das Risiko der Interventionsgruppe, so ist das relative Risiko der Kontrolle im Vergleich zur Intervention gegeben durch $RR = q/p$.

Ein **Chancenverhältnis** (OR) ist das Verhältnis zweier Chancen. Mit den Risiko-Definitionen von oben ist das Chancenverhältnis gegeben durch $OR = q\,(1\text{-}p)/p\,(1\text{-}q)$. Das RR bei gewöhnlicher Insulintherapie in 5 Jahren eine Neuropathie zu entwickeln im Vergleich zur intensivierten Insulintherapie kann nach **Tab.1** durch RR = 0,169/0,067 = 2,5 geschätzt werden. Das entsprechende OR erhält man durch OR = (52×294)/(255×21) = 2,9.

Ein RR ist leichter interpretierbar als ein OR, kann aber bei retrospektiven Fall-Kontroll-Studien nicht sinnvoll berechnet werden, da die relative Häufigkeit der Krankheit durch das Verhältnis der Fälle und Kontrollen vom Untersucher bestimmt wird. Das OR kann jedoch auch in retrospektiven Studien sinnvoll geschätzt werden. Sind die betrachteten Risiken sehr klein (< 10%), so liefern RR und OR nahezu identische Werte und das OR kann auch als Schätzung für das RR verwendet werden. In allen anderen Fällen liefert das OR jedoch extremere Werte als das entsprechende RR (6). Sei zum Beispiel $q = 0{,}9$ und $p = 0{,}5$, so erhält man RR = 1,8 und OR = 9.

Neben dem Quotienten zum Vergleich zweier Risiken kann auch die Differenz betrachtet werden. Häufig wird diese Differenz auf das Risiko der Kontrollgruppe bezogen. Damit erhält man die relative Risikodifferenz, die durch $(q\text{-}p)/q$ gegeben ist. Sie wird häufig in % ausgedrückt und stellt eine Prozentzahl von einer Prozentzahl dar. Im Falle eines präventiven Effekts spricht man von einer **relativen Risiko-Reduktion** (RRR), im Fall eines schädlichen Effekts von einem **relativen**

Dtsch. Med. Wschr. 2001; 126: T 36–T 38

Exzess-Risiko. Die RRR einer intensivierten im Vergleich zur gewöhnlichen Insulintherapie kann durch (0,169 – 0,067/ 0,169 = 0,604 = 60,4% geschätzt werden. Häufig werden in Studien, die einen Behandlungseffekt zeigen wollen, nur RR oder RRR angegeben. Mit diesen relativen Maßen, insbesondere der relativen Risiko-Reduktion bzw. dem relativen Exzess-Risiko, lassen sich oft auch dann eindrucksvolle Zahlen erzeugen, wenn der absolute Effekt der Behandlung gering ist. Relative Maße können definitionsgemäß nicht zwischen absolut hohen und geringen Effekten unterscheiden.

Um den absoluten Effekt einer Behandlung zu beschreiben, benötigt man also auch absolute Maße. Das einfachste Maß ist die absolute Risiko-Differenz *q-p*, die im Falle eines präventiven Effekts als **absolute Risiko-Reduktion** (ARR) bezeichnet wird. Bei der Untersuchung von schädlichen Einflussfaktoren wird diese Differenz auch als (absolutes) Exzess-Risiko oder **attributales Risiko** bezeichnet. Die geschätzte ARR einer intensivierten im Vergleich zu einer gewöhnlichen Insulintherapie errechnet sich durch 0,169 – 0,067 = 0,102).

Ein weiteres absolutes Maß ist die Zahl **»Number Needed to Treat«** (NNT) (5), die definiert ist als Kehrwert von ARR, das heißt NNT = 1/ARR. Das Maß »Number Needed to Treat« beinhaltet im Prinzip die gleiche Information wie ARR, ist aber besser interpretierbar. NNT ist gerade die Zahl von Patienten, die der Intervention unterzogen werden müssen, um *ein* (unerwünschtes) Ereignis zu verhindern. Hierbei wird die Höhe des Basis-Risikos mitberücksichtigt. Je geringer die Wahrscheinlichkeit, dass ein Ereignis eintritt, desto höher ist die Zahl der Patienten, die behandelt werden müssen, um *ein* Ereignis zu verhindern. Es müssen etwa NNT = 1/0,102 = 9,8 ≈ 10 Patienten 5 Jahre mit intensivierter anstelle von gewöhnlicher Insulintherapie behandelt werden, um einen Fall von Neuropathie zu verhindern.

Ausführlichere Beschreibungen der Maße zur Darstellung eines Behandlungseffekts, die aus Vierfeldertafeln abgeleitet werden, findet man in der Literatur (4, 9, 10).

Ein weiteres wichtiges Anwendungsgebiet der Vierfeldertafel stellen diagnostische und Screening-Tests dar. Es werden die positiven und negativen Resultate des zu untersuchenden Tests den entsprechenden tatsächlichen Resultaten gegenübergestellt. Ist der tatsächliche Gesundheitszustand unbekannt, so werden als Ersatz die Resultate eines entsprechenden Goldstandards verwendet. In **Tab.2** findet man beispielhaft die Daten einer Studie aus den USA, in der die Effizienz des Hämoccult-Tests zum Screening auf ein kolorektales Karzinom untersucht wurde (1).

Die grundlegenden Effizienzmaße eines diagnostischen Tests sind **Sensitivität** und **Spezifität** (2). Die Sensitivität ist definiert als Anteil der positiven Tests unter den Kranken und die Spezifität als Anteil der negativen Tests unter den Gesunden. Der Hämoccult-Test (HemeSelect) hat nach **Tab.2** eine Sensitivität von 22/32 = 69% und eine Spezifität von 7043/ 7461 = 94%.

Sensitivität und Spezifität beschreiben die allgemeine Güte eines diagnostischen Tests. In der klinischen Anwendung beantworten diese Maße aber nicht die Frage nach der Wahrscheinlichkeit für das Vorliegen der Krankheit nach Durchführung des Tests. Für die diagnostische Situation in der klinischen Praxis sind daher die prädiktiven Werte wichtiger (3). Der **positive prädiktive Wert** (PPV) ist definiert als Anteil der Kranken unter allen Test-Positiven und der **negative prädiktive Wert** (NPV) als Anteil der Gesunden unter den Test-Negativen. Zu beachten ist, dass in Studien zur Evaluierung diagnostischer Tests in der Regel die untersuchten Individuen keine Zufallsstichprobe aus der interessierenden Population darstellen, sondern dass häufig zwei Stichproben vorliegen, eine aus der Population der Gesunden und eine aus der Population der Kranken. Die Stichprobenumfänge werden hierbei vom Untersucher vorgegeben, so dass die Krankheitsprävalenz, das heißt die Wahrscheinlichkeit für das Vorliegen der Erkrankung vor Testdurchführung (A-priori-Wahrscheinlichkeit), nicht schätzbar ist und aus anderen Quellen abgeleitet werden muss. In Abhängigkeit von Sensitivität, Spezifität und Prävalenz lassen sich dann die prädiktiven Werte berechnen durch (3):

$$PPV = \text{Sensitivität} \times \text{Prävalenz} / [\text{Sensititvität} \times \text{Prävalenz} + (1\text{-Prävalenz}) \times (1\text{-Prävalenz})]$$

$$NPV = \text{Spezifität} \times (1\text{- Prävalenz}) / [\text{Spezifität} \times (1\text{- Prävalenz}) + (1\text{-Sensitivität}) \times \text{Prävalenz}]$$

Im Beispiel des Hämoccult-Tests ist die Prävalenz des kolorektalen Karzinoms (Diagnose innerhalb von 2 Jahren) aus den vorhandenen Häufigkeiten sinnvoll schätzbar durch 32/

Tab.2 Vierfeldertafel zur Untersuchung des Hämoccult-Test zum Screening auf ein kolorektales Karzinom bei 7493 Personen (1)

		kolorektales Karzinom ja	nein	Summe
Hämoccult-Test	+	22	418	440
	-	10	7043	7053
Summe		32	7461	7493

Tab.3 Übersetzungen (deutsch – englisch)

Vierfeldertafel	2 by 2 table
Kontingenztafel	contingency table
Kreuzklassifikation	cross tabulation
binär	binary
Wahrscheinlichkeit	probability
Risiko	risk
Chancen	odds
relatives Risiko	relative risk (oder: risk ratio)
Chancenverhältnis	odds ratio
Fall-Kontroll-Studie	case control study
Exzess-Risiko	excess risk
attributables Risiko	attributable risk
Sensitivität	sensitivity
Spezifität	specificity
positiver (negativer) prädiktiver Wert	positive (negative) predictive value
Prävalenz	prevalence

7493 = 0,4%. Daher lassen sich hier die prädiktiven Werte direkt einfacher schätzen durch PPV = 22/440 = 5% und NPV = 7043/7053 = 99,9%. Das bedeutet, dass nur 5% der Screening-Positiven tatsächlich ein kolorektales Karzinom haben. Trotz hoher Werte für Sensitivität und Spezifität ist der positive prädiktive Wert gering, wenn die Prävalenz der betrachteten Krankheit niedrig ist.

Ausführliche Übersichten über Methoden der Diagnose-Evaluierung findet man in der Literatur (4, 8, 11). Die Übersetzungen der diskutierten Begriffe zeigt **Tab. 3**.

kurzgefasst: **In einer Vierfeldertafel lassen sich absolute Häufigkeiten, die sich bei der Betrachtung zweier binärer Merkmale ergeben, darstellen. Mit diesen Zahlen lassen sich ableiten:**

- **Wahrscheinlichkeit (Eintrittshäufigkeit eines Ereignisses),**
- **Chance (Verhältinis der Wahscheinlichkeit, dass ein Ereignis eintritt zur Wahrscheinlichkeit, dass das Ereignis nicht eintritt),**
- **relatives Risiko (Verhältnis zweier Risiken)**
- **Chancenverhältnis**
- **Relative Risiko-Reduktion bzw. relatives Exzessrisiko**
- **Absolute Risikoreduktion**
- **Attribuales Risiko**
- **Number needed to treat**

Speziell im Rahmen einer diagnostischen Studie:
- **Sensitivität**
- **Spezifität**
- **prädiktive Werte (falls »echte« Prävalenzen vorliegen).**

Literatur

1 Allison JE, Tekawa IS, Ransom LJ, Adrain AL. A comparison of fecal occult-blood tests for colorectal-cancer screening. New Engl J Med 1996; 334: 155–159
2 Altman DG, Bland JM. Diagnostic tests 1: Sensitivity and specificity. Brit med J 1994; 308: 1552
3 Altman DG, Bland JM. Diagnostic tests 2: Predicitive values. Brit med J 1994; 309: 102
4 Bender R. Interpretation von Effizienzmaßnahmen der Vierfeldertafel für Diagnostik und Behandlung. Z Gastroenterologie 1999; (im Druck)
5 Cook RJ, Sackett DL. The number needed to treat: A clinically useful measure of treatment effect. Brit med J 1995; 310: 452–454 Correction: Brit med J 1995; 310: 1056
6 Davies HTO, Crombie IK, Tavakoli M. When can odds ratios mislead? Brit med J 1998; 316: 989–991
7 DCCT Research Group. The effect of intensive diabetes therapy on the development and progression of neuropathy. Ann Intern Med 1995; 122: 561–568
8 Jaeschke R, Guyatt G, Sackett DL for the Evidence-Based Medicine Working Group. Users´ guides to the medical literature III. How to use an article about a diagnostic test. A. Are the results of the study valid? J Amer med Ass 1994; 271: 389–391
9 Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V for the Evidence-Based Medicine Working Group. Users´ guides to the medical literature. IV. How to use an article about harm. J Amer med Ass 1994; 271: 1615–1619
10 Oxman AD, Cook DJ, Guyatt GH for the Evidence-Based Medicine Working Group. Users´ guides to the medical literature. VI. How to use an overview. J Amer med Ass 1994; 272: 1367–1371
11 Richter K, Lange S. Methoden der Diagnoseevaluierung. Internist 1997; 38: 325–336

Korrespondenz
Dr. Ralf Bender
Fakultät für Gesundheitswissenschaften
AG3: Epidemiologie und medizinische Statistik
Universität Bielefeld
Postfach 100131
33501 Bielefeld
E-Mail: ralf.bender@uni-bielefeld.de