

UNIVERSITÄT LEIPZIG

Medizinische Fakultät

Tag 2: Vorlesung 4

Wahrscheinlichkeitsrechnung / Diagnostik

Dr. rer. med. habil. Marita Ziepert /
Dr. rer. nat. Meinhard Mende,
IMISE, Universität Leipzig

imise.

Lernziele Wahrscheinlichkeitsrechnung I

Am Ende dieses Abschnitts können Sie

- Ereignisse logisch verknüpfen,
- disjunkte und nicht disjunkte Ereignisse sowie abhängige von nicht abhängigen Ereignissen unterscheiden,
- die klassische Definition der Wahrscheinlichkeit anwenden,
- Wahrscheinlichkeiten mit den Rechenregeln (Additions-, Subtraktions- und Multiplikationsregel) berechnen.

Lernziele Diagnostik

Am Ende dieses Abschnitts

- können Sie Wahrscheinlichkeiten für Test- und Diagnoseergebnisse angeben,
- können Sie Prävalenz, Sensitivität, Spezifität und prädiktive Werte anwenden und berechnen
- wissen Sie, dass PPW und NPW von der Prävalenz abhängen
- kennen Sie ROC – Kurven und
- können Sie diagnostische Wahrscheinlichkeiten grafisch darstellen

Entscheidungen unter Unsicherheit treffen

- Jeder Mensch ist ein hoch komplexes Wesen: D. h., das Ergebnis vieler Prozesse ist nicht vorhersagbar. Das nennen wir Zufall.
- Aus statistischer Sicht hat es ähnliche Folgen, ob wir Details nicht wissen (ignorance) oder ob der Ausgang unsicher ist.
- Das Wissensgebiet Stochastik modelliert den Zufall und bietet Lösungen mit Hilfe von Wahrscheinlichkeiten an.

- Ziel: Gute statistische Ausbildung für Ärzte, Gesundheitsmanager und Patienten für informierte, gemeinsame Therapieentscheidung

(vgl. Gigerenzer G et al. Helping Doctors and Patients Make Sense of Health Statistics. Psychological Science in the Public Interest 2008;8(2):53-96)

Beispiel 4.1: NT-proBNP – Marker für VHF

ESC HEART FAILURE ORIGINAL
ESC Heart Failure 2022; **9**: 100–109
Published online 30 November 2021 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/ehf2.13703

NT-proBNP as a marker for atrial fibrillation and heart failure in four observational outpatient trials

Stefanie M. Werhahn[1,2], Christian Becker[1], Meinhard Mende[3], Helge Haarmann[1], Kathleen Nolte[1], Ulrich Laufs[4], Samira Zeynalova[3,5], Markus Löffler[3,5], Nikolaos Dagres[6], Daniela Husser[6], Marcus Dörr[7,8], Stefan Gross[7,8], Stephan B. Felix[7,8], Astrid Petersmann[9], Christoph Herrmann-Lingen[2,10], Lutz Binder[2,11], Martin Scherer[12], Gerd Hasenfuß[1,2], Burkert Pieske[13,14,15], Frank Edelmann[1,2,13,14] and Rolf Wachter[1,2,4*]

- NT-proBNP als Marker für Herzinsuffizienz etabliert:
  - Assoziation hoher NT-proBNP-Werte bei Herzinsuffizienten mit hörerem Letalitätsrisiko
  - NT-proBNP signalisiert hohe Wandspannung bei diastolischer Herz-insuffizienz (HFpEF)
- Relativ neue Erkenntnis:
  Hohe NT-proBNP-Werte sind mit Vorhofflimmern (AF) assoziiert.

B 4.1: Studiendetails

Hintergrund:

Der Zusammenhang NT-proBNP – AF (atrial fibrillation) wurde anhand mehrerer Kohorten gezeigt, darunter der DIAST-CHF-Studie

Studiendesign: Diast-CHF-Studie

- Beobachtungsstudie zur diastolischen Herzinsuffzienz, heute HFpEF
- Einschluss von n = 1929 Probanden mit HFpEF bzw. Risikofaktoren für Herzinsuffizienz, darunter 202 Kontrollen
- Daten zu AF und NT-proBNP von n = 1547 Probanden
- Optimaler Cutoff: 338 pg/ml

4.1 Wahrscheinlichkeitstheoretischer Hintergrund

Geschichte der Wahrscheinlichkeitsrechnung

Chevalier de Méré

(französischer Glücksspieler der Renaissance)



Gonick & Smith:
Statistik in Cartoons

Sein Freund und Mathematiker Blaise Pascal (1623-1666) beantwortete ihm die Frage und entwickelte zusammen mit Pierre de Fermat die Wahrscheinlichkeitsrechnung.

4.1.1 Ereignisraum

Erinnerung: Wissenschaft braucht Modelle

- Leben ist sehr komplex.
- Um es zu erforschen muss man abstrahieren: Unwichtiges weg-lassen, vereinfachen, sich auf das Wesentliche konzentrieren.
- Man bildet ein Modell mit vorge-gebenen Bedingungen.

George E. P. Box:
Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.
Box GEP, Draper NR. Empirical model-building and response surfaces. John Wiley & Sons 1987



Erinnerung: Wissenschaft braucht Modelle

- Trotz großer natürlicher Variabilität gelingt es, Gesetzmäßigkeiten abzuleiten.
- Z. B. Preis für Bordeaux-Wein:

$$Preis \sim Temperatur_{Wachstumsperiode} + Niederschlag_{Erntezeit} + Gesamtniederschlag_{verg.Winter}$$

- Daniel Kahnemann: Schnelles Denken, langsames Denken.
  Dt. Ausgabe: Siedler Verlag München 2011, S. 275ff



Zufallsexperiment

Ein Stichprobenexperiment heißt Zufallsexperiment, wenn es folgende Eigenschaften aufweist:

1. Es gibt einen Versuchsplan für eine Stichprobe von Beobachtungseinheiten.
2. Dieser ist grundsätzlich beliebig oft in gleicher Weise wiederholbar.
3. Es sind genau definierte Ergebnisse möglich.
4. Das Ergebnis ist für die individuelle Beobachtungseinheit (BE) nicht vorhersagbar.

Grundraum

Der Grundraum ist die Menge aller Ergebnisse.

zum Beispiel: Münzwurf

Ergebnisse: Zahl (Z) und Wappen (W)
Grundraum: {Z, W}

Wurf mit einem Würfel

Ergebnisse: 1, 2, 3, 4, 5, 6
Grundraum: {1, 2, 3, 4, 5, 6}

Ereignis

Ein Ereignis ist eine Menge von Ergebnissen.

zum Beispiel: Ereignis =
{Augensumme gleich 4 bei einem Wurf mit zwei Würfeln}
wird erzielt in folgenden Fällen:

Ein Ereignis, welches nur ein mögliches Ergebnis enthält, nennt man Elementarereignis.

Die Menge aller möglichen Ergebnisse wird als sicheres Ereignis S bezeichnet.

Die Menge, die kein mögliches Ergebnis enthält, wird als unmögliches Ereignis bezeichnet.

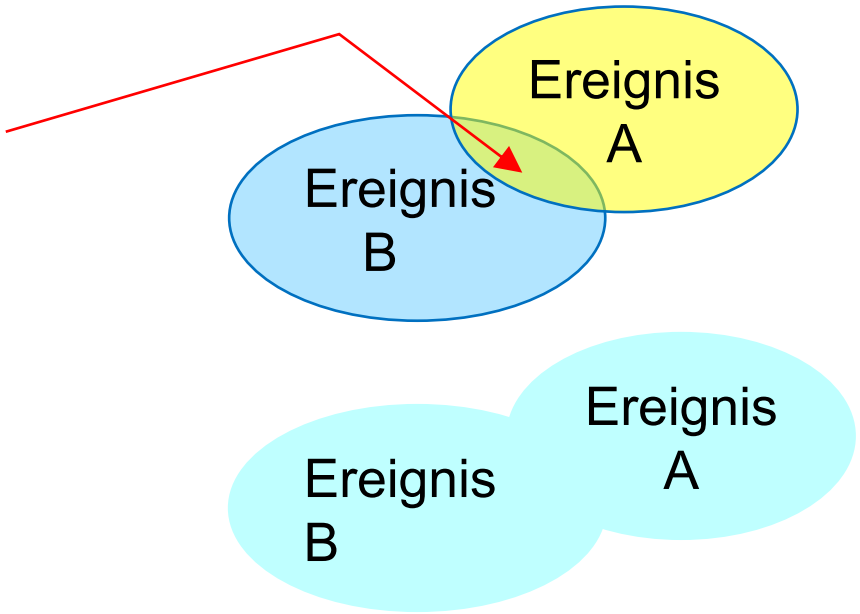
Logische Verknüpfungen von Ereignissen

Aus zwei (oder mehr) Ereignissen A und B kann man neue Ereignisse definieren:

$A \cap B$ *(A* und *B):*
Ereignis A und Ereignis B treten beide ein.

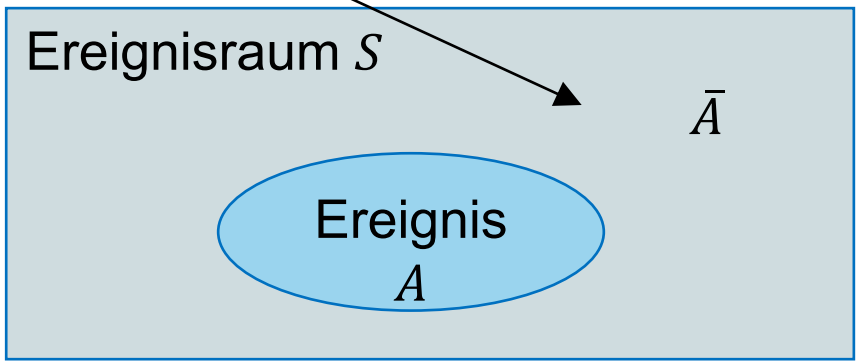
$A \cup B$ *(A* oder *B):*
Ereignis A oder Ereignis B (oder beide) treten ein.

Venn-Diagramm:



Komplementäre und disjunkte Ereignisse

- $\bar{A}$ (Nicht *A*): *Das Ereignis A tritt nicht ein.*



- A und B sind disjunkt genau dann, wenn $A \cap B = \emptyset$, d. h. A und B schließen einander aus.
- A und $\bar{A}$ sind per definitionem disjunkt

Unabhängigkeit

Ereignisse A und B zweier Zufallsprozesse sind voneinander unabhängig, wenn sie wechselseitig nicht informativ sind, d. h., aus dem (Nicht-)Eintreten des Ereignisses A kann nicht auf das (Nicht-)Eintreten des Ereignisses B geschlossen werden.

Zum Beispiel: Würfeln mit 2 verschiedenen Würfeln

4.1.2 Konzepte für Wahrscheinlichkeit

Klassische Definition der Wahrscheinlichkeit

Für ein beliebiges Ereignis A ist die Wahrscheinlichkeit (Laplace)

$$P(A) = \frac{\textit{Anzahl der für A günstigen Ergebnisse}}{\textit{Anzahl der möglichen Ergebnisse}}$$

Dabei wird vorausgesetzt, dass die möglichen Ergebnisse austauschbar, also *gleichberechtigt* sind (Elementarereignisse).

Zum Beispiel: Die Wahrscheinlichkeit bei 1 x Würfeln das Ereignis eine gerade Augenzahl zu werfen:

$$P(gerade\ Augenzahl) =$$

Dieses Modell eignet sich nur für endlich viele, vorher bekannte Elementarereignisse.

Gesetz der großen Zahl I

Johann Bernoulli 1667 – 1748:

Das Gesetz der großen Zahl sagt sinngemäß aus, dass die Streuung von relativen Häufigkeiten beliebig klein wird, wenn nur der Stichprobenumfang hinreichend groß wird.

Zum Beispiel: Werfen mit einem Würfel,

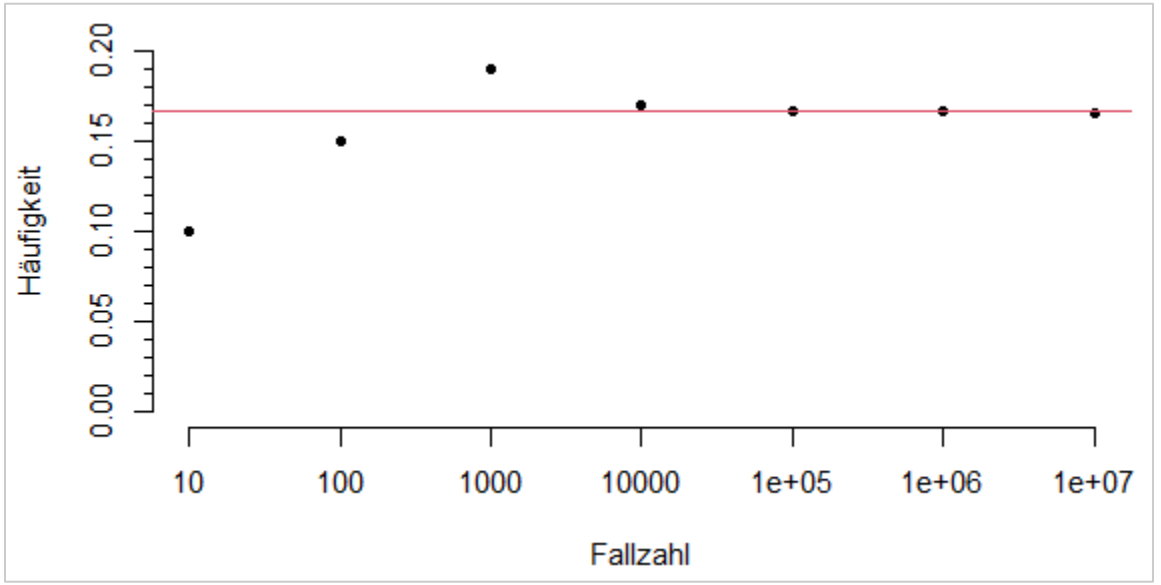
Eintreten des Ereignisses {Augenzahl 1}



Gesetz der großen Zahl II

Gedankenexperiment:

- Wir würfeln 10-, 100-, 1000-mal usw. und tragen die relativen Häufigkeiten ab.
- Dann strebt die relative Häufigkeit für wachsende Fallzahl n gegen die Wahrscheinlichkeit.
- Das wurde hier mit der Software R simuliert.



Frequentistischer Wahrscheinlichkeitsbegriff

- Das Eintreten oder Ausbleiben eines Ereignisses A eines einzelnen Zufallsexperimentes ist nicht vorhersagbar.
- n-fache Wiederholung des gleichen Zufallsexperimentes
- Beobachtung: Die relative Häufigkeit des Eintretens von A ist relativ stabil.
- Schlussfolgerung: Es existiert ein fester wahrer Wert, um den die relative Häufigkeit des zufälligen Ereignisses schwankt und dem sie sich mit wachsendem n nähert.
- Diesen Wert nennt man die Wahrscheinlichkeit des zufälligen Ereignisses A, die man mit $P(A)$ bezeichnet.

Schätzung einer Wahrscheinlichkeit

- Die Wahrscheinlichkeit wird durch ein Zufallsexperiment geschätzt:

$\hat{p}$ = relative Häufigkeit von *A* in einer großen Versuchsserie

- Maß für die Genauigkeit: die Standardabweichung *s*:

$$\hat{s} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Anzahl der Beobachtungseinheiten
Stichprobengröße

Axiome der Wahrscheinlichkeitsrechnung I

Andrej Nikolajewitsch Kolmogorow entwickelte 1930 eine auf Axiomen basierende Definition von Wahrscheinlichkeit:
Wir betrachten eine Menge von möglichen Ereignissen.

1. Jedem Ereignis *A* wird eine Zahl $P(A)$ zwischen *0* und *1* zugeordnet:

$$0 \leq P(A) \leq 1 \quad \text{Nicht-Negativität}$$

P (A) heißt Wahrscheinlichkeit von *A*

2. Das sichere Ereignis hat die Wahrscheinlichkeit *1:*

$$P(S) = 1 \quad \text{Normierung}$$

Axiome der Wahrscheinlichkeitsrechnung II

3. Für zwei disjunkte Ereignisse (→ Folie 10) gilt:

   Die Wahrscheinlichkeit für das Eintreten eines der Ereignisse *A* oder *B* ergibt sich als Addition der Wahrscheinlichkeiten der beiden Ereignisse *A* bzw. *B*

$$P(A \cup B) = P(A) + P(B)$$ Additivität

oder

Auf Axiomen - unbeweisbaren Grundsätzen – basieren ganze Theorien. Häufig entsprechen sie unserem gewöhnlichen Denken.

4.1.3 Rechnen mit Wahrscheinlichkeiten

Rechnen mit Wahrscheinlichkeiten

Zweiwertige Logik:

- Ein Ereignis D kann entweder eintreten oder auch nicht, z. B.
- Es ist sicher, entweder die Krankheit zu haben (D^+) oder nicht (D^-).

Aus Kolmogorovs Axiom 3 folgt die

Allgemeine Additionsregel:

$$P(D^+) + P(D^-) = P(D^+ \cup D^-) = 1$$

Bedingte Wahrscheinlichkeiten

Beispiel 4.1 b: Diast-CHF-Studie

In einer Kohorte von 1727 Patienten mit kardiovaskulären Risikofaktoren soll die Frage untersucht werden, ob die Diagnose KHK (K) von der Diagnose arterielle Hypertonie (H) abhängt.

Was vermuten Sie?

Gegeben sind folgende Häufigkeiten:

KNHI
Kompetenznetz
Herzinsuffizienz

	KHK- (K^-)	**KHK+**(K^+)	$\sum$
ohne Hypertonie (H^-)	161	26 (14%)	187
mit Hypertonie (H^+)	1183	357 (23%)	1540
$\sum$	1344	383	1727

Bedingte Wahrscheinlichkeiten

- Für Szenarien wie dieses wurde das Konzept bedingter Wahrscheinlichkeiten eingeführt:
- Die Bedingung H schränkt den Wahrscheinlichkeitsraum auf H ein.
- Die bedingte Wahrscheinlichkeit des Ereignisses *K*, gegeben, dass das Ereignis *H* eingetreten ist, wird definiert durch:

$$P(K|H) = \frac{P(K \cap H)}{P(H)}$$

und



Unabhängigkeit

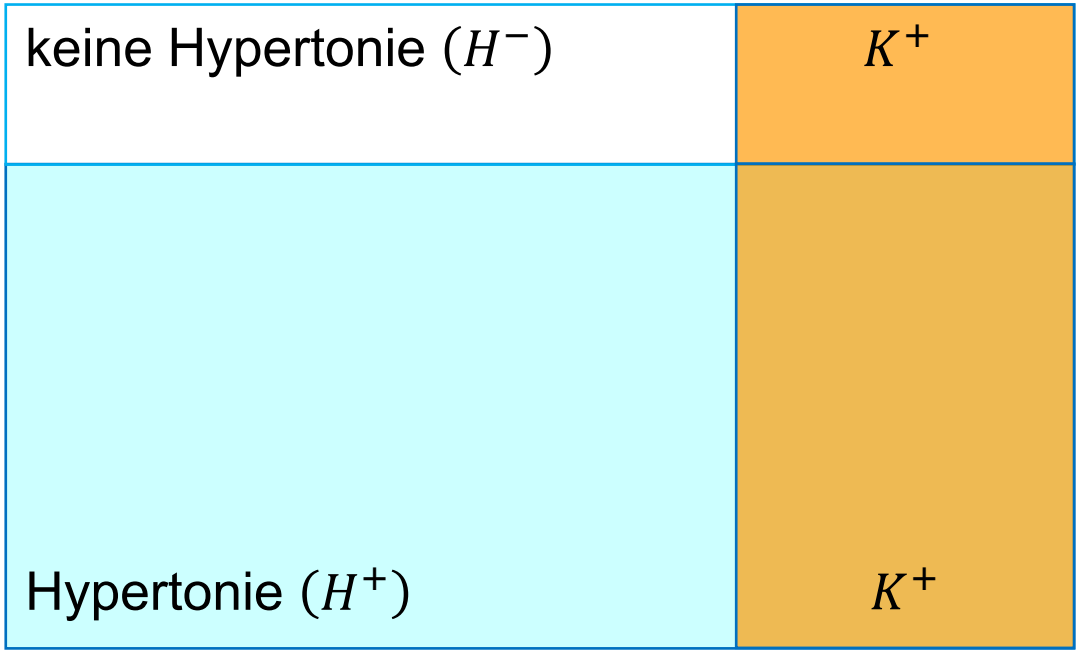
Wäre dagegen der Anteil mit KHK in beiden Subgruppen gleich, also

$P(K^+|H^-) = P(K^+|H^+)$,

- läge Unabhängigkeit der Merkmale H und K vor.
- Dann wären die bedingten Wahrscheinlichkeiten für K gleich der unbedingten:

$P(K^+|H^-) = P(K^+|H^+) = P(K^+)$

- die vereinfachte Multiplikationsregel

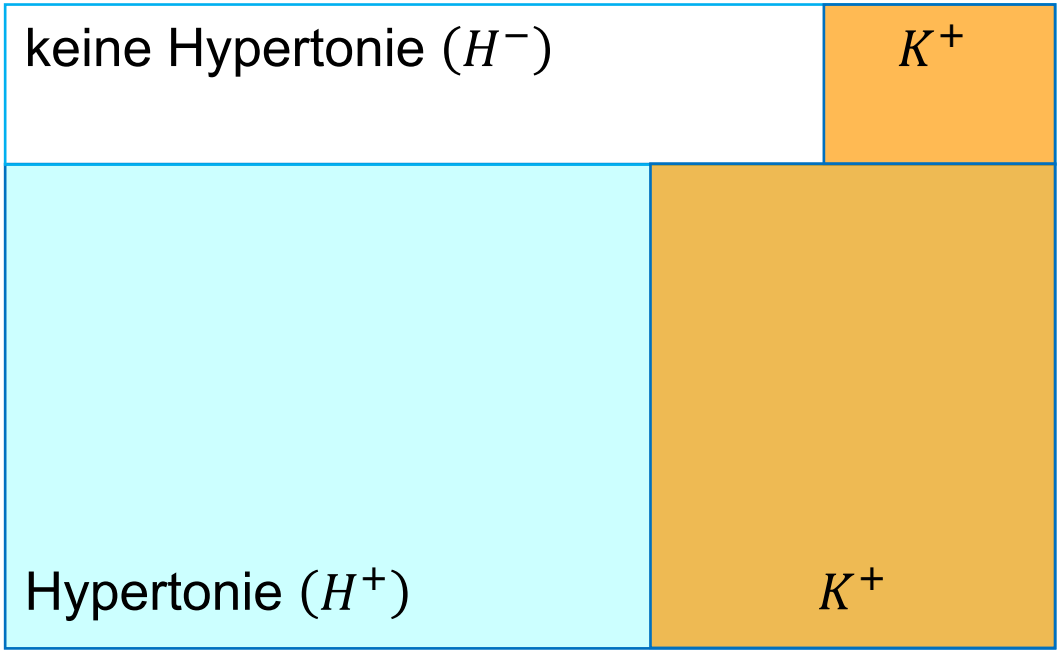


und es gälte

$P(H^+ \cap K^+) = P(H^+) \cdot P(K^+)$

Statistische Abhängigkeit (Assoziation)

- Die Mehrzahl dieser Patienten leidet unter hohem Blutdruck (H^+, Grafik nicht maß-stabsgetreu)
- Der Anteil mit KHK (K^+) unter Nicht-Hyper-tonikern beträgt nur etwa 2/3 des Anteils bei Hypertonikern:



$$P(K^+|H^-) < P(K^+|H^+)$$

- Dann sprechen wir von Abhängigkeit oder Assoziation.

Multiplikationsregel

Wahrscheinlichkeit des Durchschnitts von Ereignissen:

⇨ *2 (oder mehr) Ereignisse treten gleichzeitig auf*

Allgemeiner Multiplikationssatz:

Die Wahrscheinlichkeit für das gemeinsame Auftreten der Ereignisse A und B gilt:

$$P(A \cap B) = P(A \mid B) \cdot P(B)$$

Sind die Ereignisse unabhängig, folgt daraus:

$$P(A \cap B) = P(A) \cdot P(B) \quad (*)$$

Das heißt, für unabhängige Ereignisse gilt :

$$P(A|B) = P(A) \quad (**)$$

Die Gleichungen $(*)$ oder $(**)$ definieren Unabhängigkeit.

Additionsregel

Wahrscheinlichkeit der Vereinigung von Ereignissen:

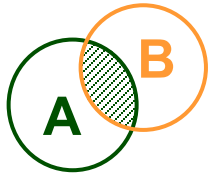
mindestens 1 von 2 (oder mehreren) Ereignissen tritt auf

Allgemeine Additionsregel

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

oder *und*

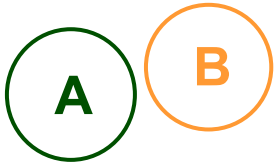
Venn-Diagramm:



Für disjunkte Ereignisse A und B vereinfacht sich die Formel zu:

$$P(A \cup B) = P(A) + P(B)$$

(vgl. Kolmogorovs 3. Axiom)



Subtraktionsregel

$$P(A) = 1 - P(\bar{A})$$

wird angewandt, wenn $P(\bar{A}) = P(nicht\ A)$ einfacher zu berechnen ist als $P(A)$.

Z. B.:

Wie hoch ist die Wahrscheinlichkeit des Ereignisses A, mit zwei Würfeln zwei unterschiedliche Augenzahlen zu würfeln?

$P(A) = 1 - P(nicht\ A)$

$=$

$=$

Wahrscheinlichkeitsbaum I

- Zusammengesetzte Ereignisse
- Baumdiagramm:
  sehr übersichtliches Schema zur Berechnung von Wahrscheinlichkeiten.

Zum Beispiel:

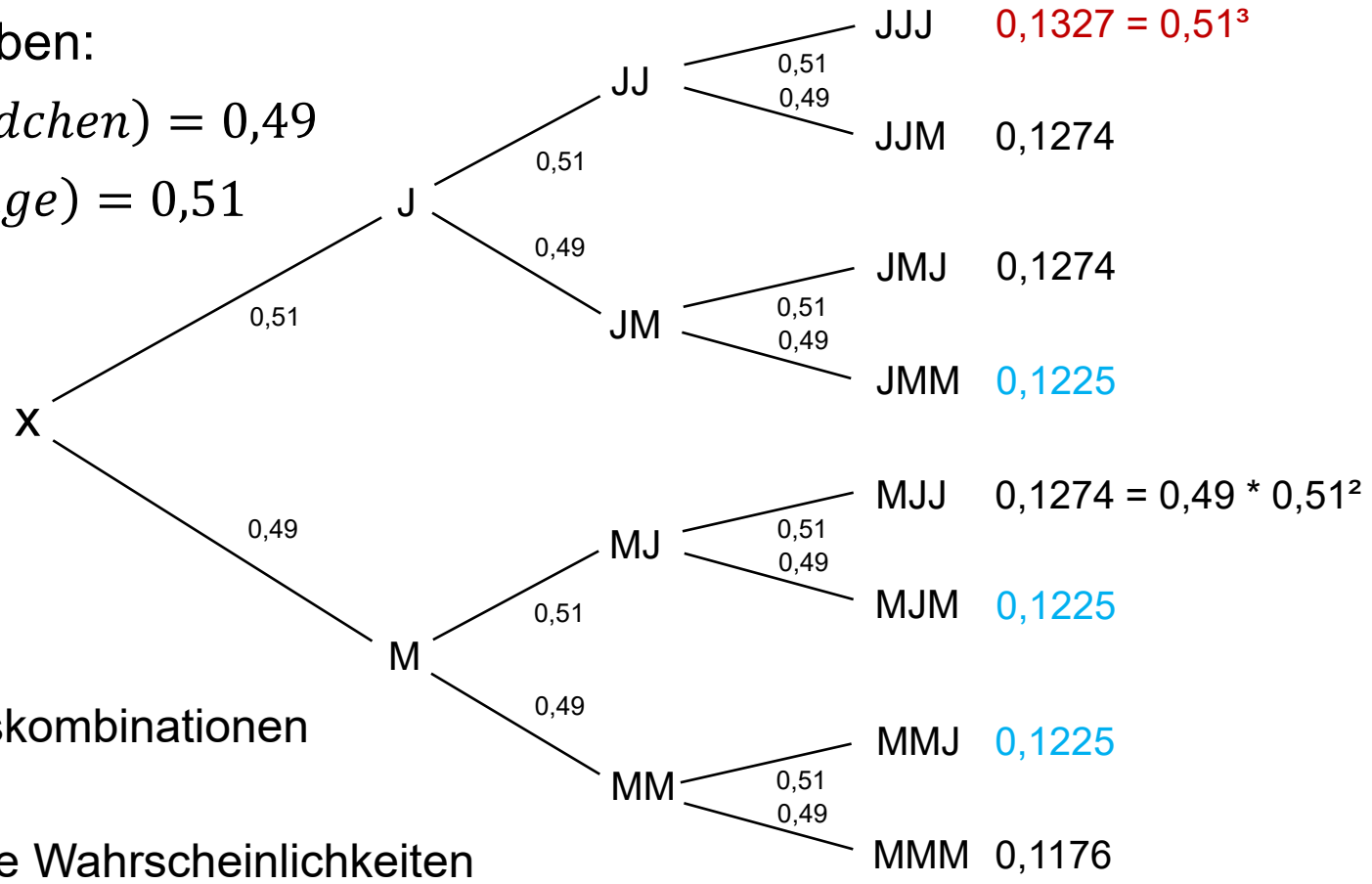
- In einer Familie wurden drei Kinder geboren;
- mit welchen Wahrscheinlichkeiten treten die möglichen Kindersequenzen auf?
  - Wie wahrscheinlich ist es, dass drei Jungen geboren wurden?
  - Wie hoch ist die (theoretische) Wahrscheinlichkeit, dass zwei Mädchen und ein Junge geboren wurden?

Wahrscheinlichkeitsbaum II

Gegeben:

$P(Mädchen) = 0{,}49$

$P(Junge) = 0{,}51$



Knoten:

Ereigniskombinationen

Kanten:

Bedingte Wahrscheinlichkeiten

Wahrscheinlichkeitsbaum III

- Nach dem Multiplikationssatz dürfen Wahrscheinlichkeiten entlang eines Astes multipliziert werden:

$$P(J\ J\ J) = P(J) \cdot P(J) \cdot P(J) = 0{,}51^3 = 0{,}1327 \approx 13{,}3\%$$

- Ereigniskombinationen auf verschiedenen Ästen sind disjunkt.
- Ihre Wahrscheinlichkeiten dürfen nach dem Additionssatz einfach addiert werden:

$$P(2M, 1J) = P(M\ M\ J) + P(M\ J\ M) + P(J\ M\ M)$$
$$= 3 \cdot P(MMJ) = 3 \cdot 0{,}1225 = 0{,}3675 \approx 36{,}8\%$$

CAVE: Diese Rechnungen setzen Unabhängigkeit der Geburten voraus.
Diese kann man nicht selbstverständlich voraussetzen.



de Mérés Frage

Können Sie de Mérés Frage (Folie 2) nun beantworten?

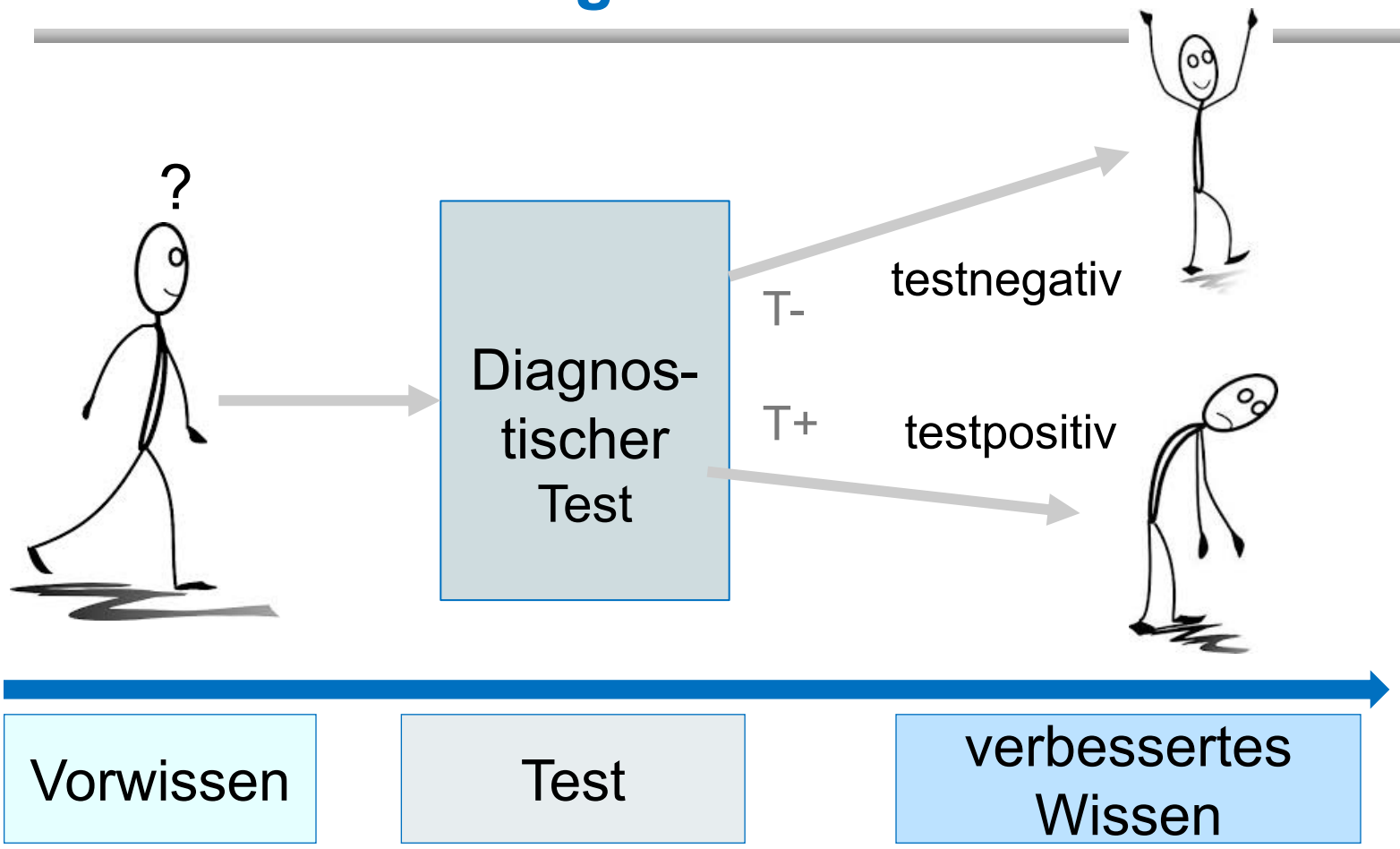
$$P(\text{mindestens eine 6 in vier Würfen mit einem Würfel}) =$$

$$P(\text{mindestens ein 6er}-\text{Pasch in 24 Würfen mit zwei Würfeln}) =$$

4.2 Mathematische Grundlagen der Diagnostik

4.2.1 Modellierung und Schätzung von Wahrscheinlichkeiten

Diagnostische Tests



AF und NT-proBNP: Fragestellungen

Charakteristika des Tests

- Welcher Anteil Probanden mit AF wird durch einen positiven Test richtig klassifiziert? (= Sensitivität)
- Wie hoch ist die Wahrscheinlichkeit, dass Probanden ohne AF ein negatives Testergebnis haben? (= Spezifität)

Prädiktion

- Mit welcher Wahrscheinlichkeit hat ein positiv Getesteter tatsächlich Vorhofflimmern? (= PPW)
- Mit welcher Wahrscheinlichkeit ist ein negativ Getesteter tatsächlich frei von AF? (= NPW)

Eignung des Tests

- Ist der Test zum Screening / zur endgültigen Diagnose geeignet?

Ergebnis NT-proBNP-Test

		NT-proBNP		Summe
		< 338 µmol/l	≥ 338 µmol/l	
Vorhof-Flimmern baseline	nein	1244 88,0%	169 12,0%	1413 (100%)
	ja	42 31,3%	92 68,7%	134 (100%)
Summe		1286	261	1547 (100%)

Erster Eindruck:

Typisches Diagnostik-Szenarium

- zwei binäre Ereignisvariablen: Erkrankung (n / j): D^-, D^+ verifiziert durch Goldstandard
- diagnostischer Test (negativ / positiv): T^-, T^+
- Das mündet unmittelbar in eine Kreuztabelle.
- Aus dem sachlogischen Zusammenhang ergeben sich die 4 Klassen RN (richtig negativ), FP (falsch positiv), FN und RP

		Diagnost. Test		
		T^-	T^+	
Erkrankung	D^-	RN	FP	
	D^+	FN	RP	

Formalisierung 1

- Es ergeben sich zunächst die (un)bedingten (Rand-) Wahrscheinlichkeiten:

		Diagnost. Test		
		T^-	T^+	
Erkran-kung	D^-			$P(D^-)$
	D^+			$P(D^+)$
		$P(T^-)$	$P(T^+)$	

Prävalenz in dieser Kohorte:

$$Prev = \frac{134}{1547} = 8{,}7\%$$

Anteile von negativem / positivem Test:

Formalisierung 2

- Fokussiert man auf je eine Zeile, ergeben sich bedingte Wahrscheinlichkeiten bezogen auf Erkrankung oder nicht:

		Diagnost. Test	
		T^-	T^+
Erkrankung	D^-		
	D^+	$P(T^-\|D^+)$	$P(T^+\|D^+)$
		… unter der Bedingung, dass der Proband erkrankt ist	

Formalisierung 3

- Fokussiert man auf je eine Spalte, ergeben sich bedingte Wahrscheinlichkeiten bedingt auf das Testergebnis:

		Diagnost. Test		
		T^-	T^+	
Erkran-kung	D^-		$P(D^-\|T^+)$	
	D^+		$P(D^+\|T^+)$	

CAVE:

I. Allg. gilt $P(D^-|T^+) \neq P(T^+|D^-)$, d. h.

man darf Ereignis und Bedingung nicht vertauschen!



Prävalenz

		NT-proBNP		
		T^-	T^+	
AF	D^-	1244	169	1413
	D^+	42	92	134
		1286	261	1547

Die Prävalenz,

die Wahrscheinlichkeit für das Vorliegen von Vorhofflimmern (D+) in der Grundgesamtheit, kann hier empirisch aus der Studienkohorte geschätzt werden:

$$\hat{P}(D^+) = \frac{134}{1547} = 0{,}0866 \triangleq 8{,}7\%$$

Wahrscheinlichkeit für Krankheitsausschluss

Analog ergibt sich die Wahrscheinlichkeit für den Krankheitsausschluss:

$$\hat{P}(D^-) = \frac{1413}{1547} = 0{,}913 \triangleq 91{,}3\%$$

		NT-proBNP		
		T^-	T^+	
AF	D^-	1244	169	1413
	D^+	42	92	134
		1286	261	1547

Anderer Rechenweg:

Nach Subtraktionsregel gilt: $\hat{P}(D^-) = 1 - \hat{P}(D^+)$

$$= 1 - 0{,}087 = 91{,}3\%$$

Wahrscheinlichkeiten für richtige Diagnosen

Weitere Wahrscheinlichkeiten können unmittelbar aus der Kreuztabelle geschätzt werden:

		Diagnost. Test		
		T^-	T^+	
AF	D^-	1244	169	1413
	D^+	42	92	134
		1286	261	1547

- Wahrscheinlichkeit für die richtig positive Diagnose (RP) aus den Daten geschätzt:

$$\hat{P}\,[\text{positive Diagnose UND positiver Test}] = \hat{P}(RP) = \hat{P}(D^+ \cap T^+)$$
$$= \frac{92}{1\,547} = 0{,}059 = 5{,}9\%$$

- Analog: Wahrscheinlichkeit für die richtig negative Diagnose (RN):

$$\hat{P}\,[negative\ \text{Diagnose UND negativer Test}] = \hat{P}(RN) = \hat{P}(D^- \cap T^-)$$
$$= \frac{1\,244}{1\,547} = 0{,}804 = 80{,}4\%$$

Güte (Kenngrößen) diagnostischer Tests

Frage:

Wie groß ist die Wahrscheinlichkeit zutreffender Testausgänge?

Sensitivität: Positiver Test bei Kranken (entdeckt Kranke) $P(T^+ \mid D^+)$

Spezifität: Negativer Test bei Gesunden (schließt Krankheit aus) $P(T^- \mid D^-)$

Beide Kenngrößen müssen validiert werden, ehe der Test in die Praxis überführt werden kann.

Testcharakteristikum 1: Sensitivität

Sensitivität = Anteil Test-Positiver unter Erkrankten

$$Sens = \frac{P(T^+ \cap D^+)}{P(D^+)} = P(T^+|D^+)$$

		Diagnost. Test		
		T^-	T^+	
AF	D^-	1244	169	1413
	D^+	42	92	134
		1286	261	1547

- Die Bedingung D+ schränkt auf die untere Zeile ein und die Sensitivität kann aus den Daten geschätzt werden durch:

$$\widehat{Sens} = \frac{\#\ Testpostitiv\ und\ erkrankt}{\#\ Erkrankt},$$

hier: $\widehat{Sens} = \frac{92}{134} = 0{,}687 = 68{,}7\%$

Testcharakteristikum 2: Spezifität

Spezifität = Anteil Test-Negativer unter Nicht-Erkrankten:

$$Spez = \frac{P(T^- \cap D^-)}{P(D^-)} = P(T^-|D^-)$$

–

		Diagnost. Test		
		T^-	T^+	
AF	D^-	1244	169	1413
	D^+	42	92	134
		1286	261	1547

– Die Bedingung D^- schränkt auf die obere Zeile ein und die Spezifität kann aus den Daten geschätzt werden durch:

$$\widehat{Spez} = \frac{\#\ Testnegativ\ und\ nicht\ erkrankt}{\#\ Nicht\ erkrankt}$$

hier: $\widehat{Spez} = \frac{1\,244}{1\,413} = 0{,}880 = 88{,}0\%$

Kein Testverfahren ist fehlerfrei

Leider treten gelegentlich falsche (inkonsistente) Testergebnisse auf:

		Diagnost. Test		
		T^-	T^+	
AF	D^-	1244	169	1413
	D^+	42	92	134
		1286	261	1547

- Wahrscheinlichkeit für die falsch positive Diagnose (FP) aus den Daten geschätzt:

$$\hat{P}\,[\text{negative Diagnose UND positiver Test}] = \hat{P}(FP) = \hat{P}(D^- \cap T^+)$$
$$= \frac{169}{1\,547} = 0{,}102 = 10{,}2\%$$

- Ebenso gibt es falsch negative Diagnosen (FN):

$$\hat{P}\,[positive\ \text{Diagnose UND negativer Test}] = \hat{P}(FN) = \hat{P}(D^+ \cap T^-)$$
$$= \frac{42}{1\,547} = 0{,}027 = 2{,}7\%$$

Kosten falscher Testergebnisse

Falsch positiv

- Weitere belastende diagnostische Verfahren folgen (hier: Langzeit-EKG)
- Ggf. erleidet der Patient dadurch Schaden.
- Psychische Belastung durch falsche Diagnose

Falsch negativ

- Erkrankung bleibt unerkannt.
- Patient wiegt sich in falscher Sicherheit.
- hier: Unerkanntes Vorhofflimmern kann zur Ursache eines Schlaganfalls werden.

Wahrscheinlichkeiten - Zusammenfassung

Testergebnis (NT-proBNP)				
		T- (negativ)	**T+** (positiv)	
7-Tages-Langzeit-EKG	**D-** (kein Vorhofflimmern)	$\hat{P}(T-\cap D-)$ 0,804 RN	$\hat{P}(T+\cap D-)$ 0,102 FP	$\hat{P}(D-)$ 0,906
	D+ (AF)	$\hat{P}(T-\cap D+)$ 0,027 FN	$\hat{P}(T+\cap D+)$ 0,059 RP	$\hat{P}(D+)$ 0,086
		$\hat{P}(T-)$ 0,831	$\hat{P}(T+)$ 0,161	0,992

Rundungsabweichung

Bewertung des diagnostischen Tests 1

Die Kenngrößen des Tests entscheiden, ob der Test im entsprechenden Szenario sinnvoll ist.

- Die Spezifität ist ziemlich hoch (88%), nicht Erkrankte werden also relativ gut klassifiziert.
- Nur etwa 2 von 3 an AF Erkrankten werden identifiziert (Sensitivität). NT-proBNP könnte als Vortest Risikoprobanden finden, von denen dann ein Langzeit-EKG genommen wird.

4.2.2 Vorhersagen auf Grund diagnostischer Tests

Ziel – Prädiktive Werte bestimmen

Hauptfrage:

Ein Patient hatte einen positiven Test – wie hoch ist nun seine Wahrscheinlichkeit, wirklich erkrankt zu sein?

Das heißt Positiv-Prädiktiver Wert (PPW, engl. PPV)

Formalisiert:

- gegeben sind $P(T^+|D^+)$ und $P(T^-|D^-)$,
- gesucht ist aber: $\mathrm{PPW} = P(D^+|T^+)$

- Dafür muss die Basiswahrscheinlichkeit (meist Prävalenz $\hat{P}(D^+)$) bekannt sein.

Lösung 1: PPW aus der Kreuztabelle

PPW := Wahrscheinlichkeit, dass die Erkrankung tatsächlich vorliegt, wenn der Test positiv ist

⇨ Vorhersage

	T-	**T+**
D-	RN $P(T-\cap D-)$	FP $P(T+\cap D-)$
D+	FN $P(T-\cap D+)$	RP $P(T+\cap D+)$
Spaltensumme	RN+FN $P(T-)$	FP+RP $P(T+)$

$$PPW = P(D^+|T^+) = \frac{P(T^+ \cap D^+)}{P(T^+)} = \frac{\#RP}{(\#FP + \#RP)}$$

PPW am Beispiel 4.1

- 92 von 261 der positiv getestetem Patienten hatten tatsächlich Vorhofflimmern.

- Daraus kann man den PPW schätzen:

		Diagnost. Test		
		T^-	T^+	
AF	D^-	1244	169	1413
	D^+	42	92	134
		1286	261	1547

$$\widehat{PPW} = \frac{\#\ postitiv\ getestet\ und\ erkrankt}{\#\ positiv\ getestet} = P(D^+|T^+)$$

$$= \frac{92}{261} = 0{,}352 = 35{,}2\%$$

Erkenntnisgewinn durch den Test:

- Die Wahrscheinlichkeit ist von 8,7% (Prävalenz) auf das Vierfache gestiegen.

Lösung 2: Satz von Bayes

(Thomas Bayes, 1701 - 1761)

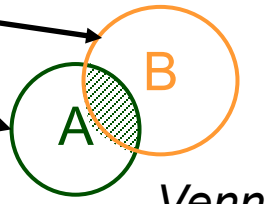
$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})}$$

Der Satz von Bayes verbindet $P(A|B)$ mit $P(B|A)$, wenn $P(A)$ gegeben ist.

Herleitung: Aus dem Multiplikationssatz folgt:

$$P(A \cap B) = P(A|B) \cdot P(B) \quad \text{und} \quad P(A \cap B) = P(B|A) \cdot P(A) \quad (*)$$

und n. Def. $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(A \cap B) + P(\bar{A} \cap B)}$



Venn-Diagramm

und wegen $(*)$ $= \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})}$

Bayessche Formel

- Gegeben: $P(T^+|D^+)$, $P(T^-|D^-)$, $P(D^+)$,
- Gesucht: $P(D^+|T^+)$

Erkrankung D		Testergebnis T		
		T^-	T^+	
	D^-	$P(T^-\|D^-)$		$P(D^-)$
	D^+	$P(T^- \cap D^+)$	$P(T^+ \cap D^+)$	$P(D^+)$
		$P(\bar{T}^-)$	$P(T^+)$	

$$P(D^+|T^+) = \frac{P(T^+|D^+) \cdot P(D^+)}{P(T^+|D^+) \cdot P(D^+) + P(T^+|D^-) \cdot P(D^-)}$$

- Durch die Bayessche Formel gelingt der Paradigmenwechsel, dass Ereignis und Bedingung die Rollen tauschen.

Anwendung der Bayes-Formel

$$P(D^+|T^+) = \frac{P(T^+|D^+)\cdot P(D^+)}{P(T^+|D^+)\cdot P(D^+)+P(T^+|D^-)\cdot P(D^-)} \quad (*)$$

Wir kennen:

- $Sens = P(T^+|D^+) = 0{,}687$
- $Spez = P(T^-|D^-) = 0{,}804$
- $\text{Präv} = P(D^+) = 0{,}0866$

Es fehlen noch:

- $P(D^-) \quad = 1 - P(D^+)$
- $P(T^+|D^-) \quad = 1 - P(T^-|D^-)$
- eingesetzt in Formel (*)

		Diagnost. Test		
		T^-	T^+	
AF	D^-	1244	169	1413
	D^+	42	92	134
		1286	261	1547

$$P(D^+|T^+) = \frac{P(T^+|D^+)\cdot P(D^+)}{P(T^+|D^+)\cdot P(D^+) + \{1 - P(T^-|D^-)\}\cdot\{1 - P(D^+)\}}$$

Anwendung der Bayes-Formel 2

Ersetzen wir nun die bedingten Wahrscheinlichkeiten durch die diagnostischen Kenngrößen, lautet die Bayessche Formel für Diagnostik:

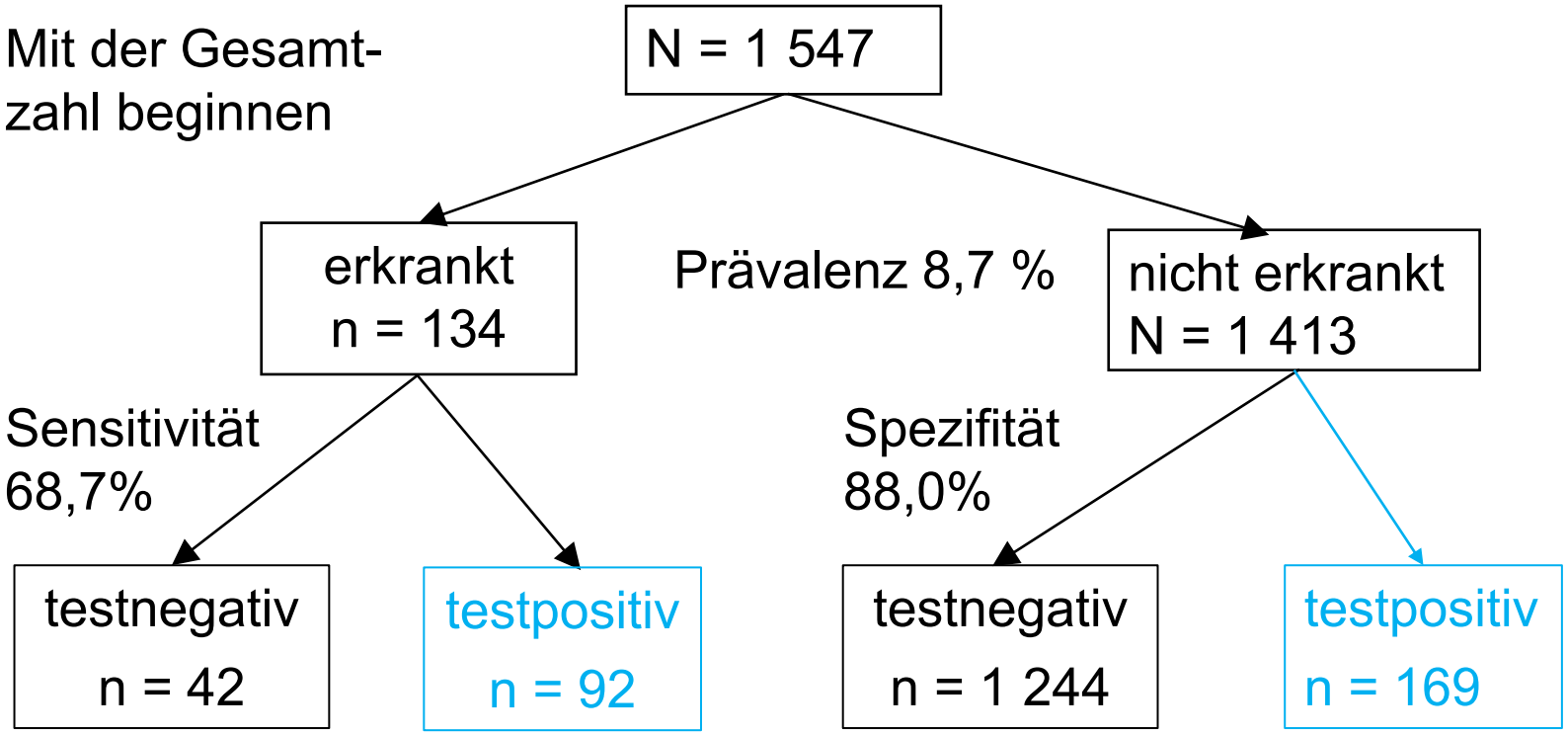
		Diagnost. Test		
		T^-	T^+	
AF	D^-	1244	169	1413
	D^+	42	92	134
		1286	261	1547

$$PPW = \frac{Sens \cdot Präv}{Sens \cdot Präv + \{1 - Spez\} \cdot \{1 - Präv\}}$$

Mit unseren Zahlen folgt: $PPW = \frac{0{,}687 \cdot 0{,}0866}{0{,}687 \cdot 0{,}0866 + \{1 - 0{,}88\} \cdot \{1 - 0{,}0866\}}$

bzw. $PPW = \frac{0{,}687 \cdot 0{,}0866}{0{,}687 \cdot 0{,}0866 + 0{,}12 \cdot 0{,}9134} = 0{,}352 = 35{,}2\%$

Lösung 3: PPW mit Ereignisbaum



Positiv-prädiktiver Wert $:= \frac{\#\ richtig\ Positive}{\#\ Testpositive} =$

Negativ prädiktiver Wert (PPW)

Mit ähnlichen Formeln kann man den Negativ-Prädiktiven Wert (engl. NPV) ermitteln:

		Diagnost. Test		
		T^-	T^+	
AF	D^-	1244	169	1413
	D^+	42	92	134
		1286	261	1547

$$\widehat{NPW} = \frac{P(T^- \cap D^-)}{P(T^-)} = P(D^-|T^-)$$

$$= \frac{Spez \cdot (1-Präv)}{Spez \cdot (1-Präv) + (1-Sens) \cdot Präv}$$

$$= \frac{0{,}88 \cdot (1-0{,}087)}{0.88 \cdot (1-0{,}087) + (1-0{,}687) \cdot 0{,}087} = 0{,}967 = 96{,}7\%$$

Personen mit negativem Test können sich also ziemlich sicher sein, kein Vorhofflimmern zu haben.

NPW aus der Tabelle

Aus der Tabelle können wir ablesen:

		Diagnost. Test		
		T^-	T^+	
AF	D^-	1244	169	1413
	D^+	42	92	134
		1286	261	1547

$$NPW = \frac{\#\ Richtig\ negativ}{\#\ Negativ\ getestet}$$

$$NPW = \frac{1\,244}{1\,286} = 0{,}967 = 96{,}7\%$$

Vor dem Test hatte ein Proband die Wsk. $(100 - 6{,}7\%) = 93{,}3\%$ nicht an AF erkrankt zu sein, bei negativem Test ist es die Wsk. $96{,}7\%$,

also auch hier hat es einen Wissenszuwachs gegeben.

Prävalenz – prädiktive Werte

Sensitivität und Spezifität hängen als Eigenschaften des diagnostischen Tests nicht vom Szenario (also z.B. der Prävalenz) ab.

Frage: Wie verhalten sich die Vorhersagewerte (PPW, NPW), wenn sich die Prävalenz ändert?

Antwort: Die Basiswahrscheinlichkeit (i.d.R. Prävalenz) bestimmt ganz wesentlich die prädiktiven Werte PPW und NPW.

- Ist die Kreuztabelle (vgl. Folien 43 ff.) nicht repräsentativ für die Grundgesamtheit (z. B. Fall-Kontroll-Studie), muss man die Prävalenz nach dem Satz von Bayes bestimmen.
- Das geht am einfachsten mit Hilfe eines Baumdiagramms.

Vorhersagewerte abhängig von der Prävalenz

Entspricht der Anteil der Erkrankten in der Untersuchungs-population nicht der Prävalenz in der Population, in der der Test eingesetzt werden soll, so müssen die Vorhersagewerte über die Bayessche Formel bestimmt werden:

$$PPW = P(D^+|T^+) = \frac{Prävalenz \cdot Sensitivität}{Prävalenz \cdot Sensitivität + (1 - Prävalenz) \cdot (1 - Spezifität)}$$

$$NPW = P(D^-|T^-) = \frac{(1 - Prävalenz) \cdot Spezifität}{(1 - Prävalenz) \cdot Spezifität + Prävalenz \cdot (1 - Sensitivität)}$$

Das wird am einfachsten mit Hilfe eines passenden Ereignis-diagramms realisiert.

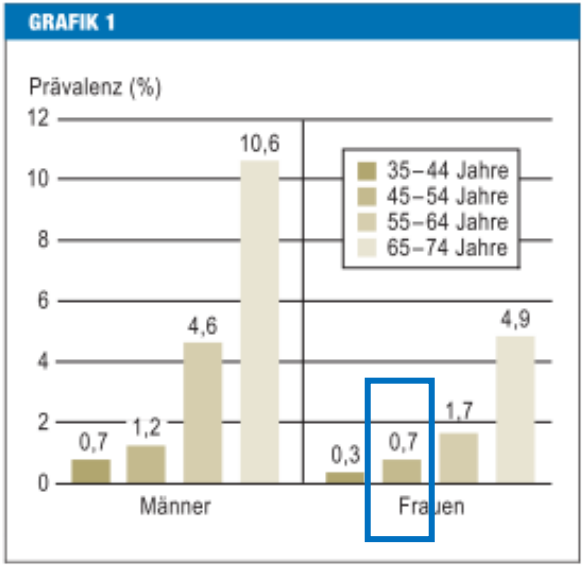
Andere Prävalenz

In der Studienkohorte (Prävalenz 8,7%) war der PPW = 35,2%.

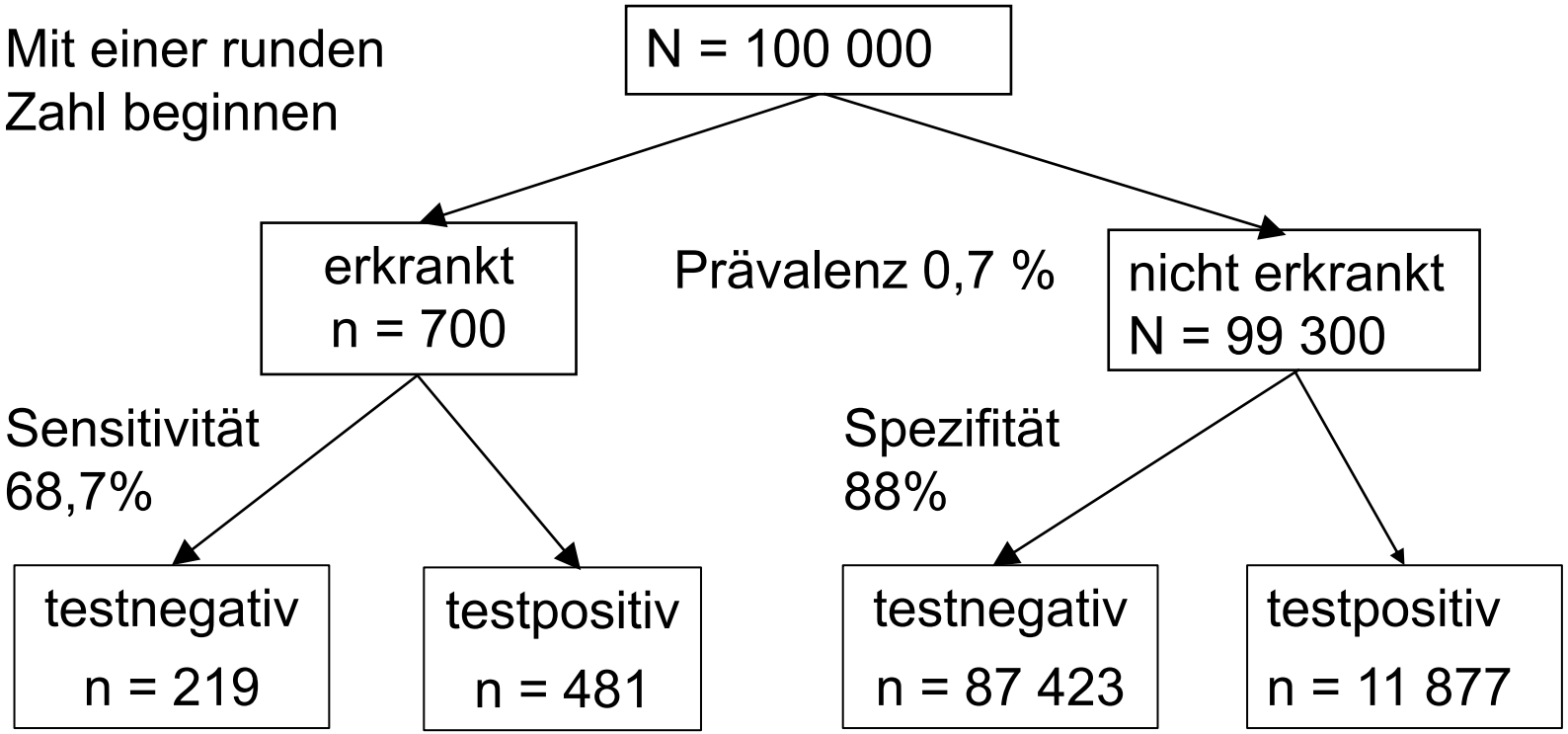
- Angenommen, bei Frau Vera Hof (51 Jahre) wird ein NT-proBNP-Wert von 520 µmol/l gemessen. Mit welcher Wahrscheinlichkeit hat sie Vorhofflimmern?
- Durch die GHS-Studie wurde die Prävalenz von AF wie folgt bestimmt (n = 210 867,

  Schnabel et al. DÄbl 2012; 109(16):293)
- Die Studie ermittelte für Frauen von 45 – 54 Jahren eine Prävalenz von 0,7%.
- Schätzen Sie: Wie hoch ist der PPW für Frau Vera Hof?

30%, 10%, 5%, 3%, 1%,

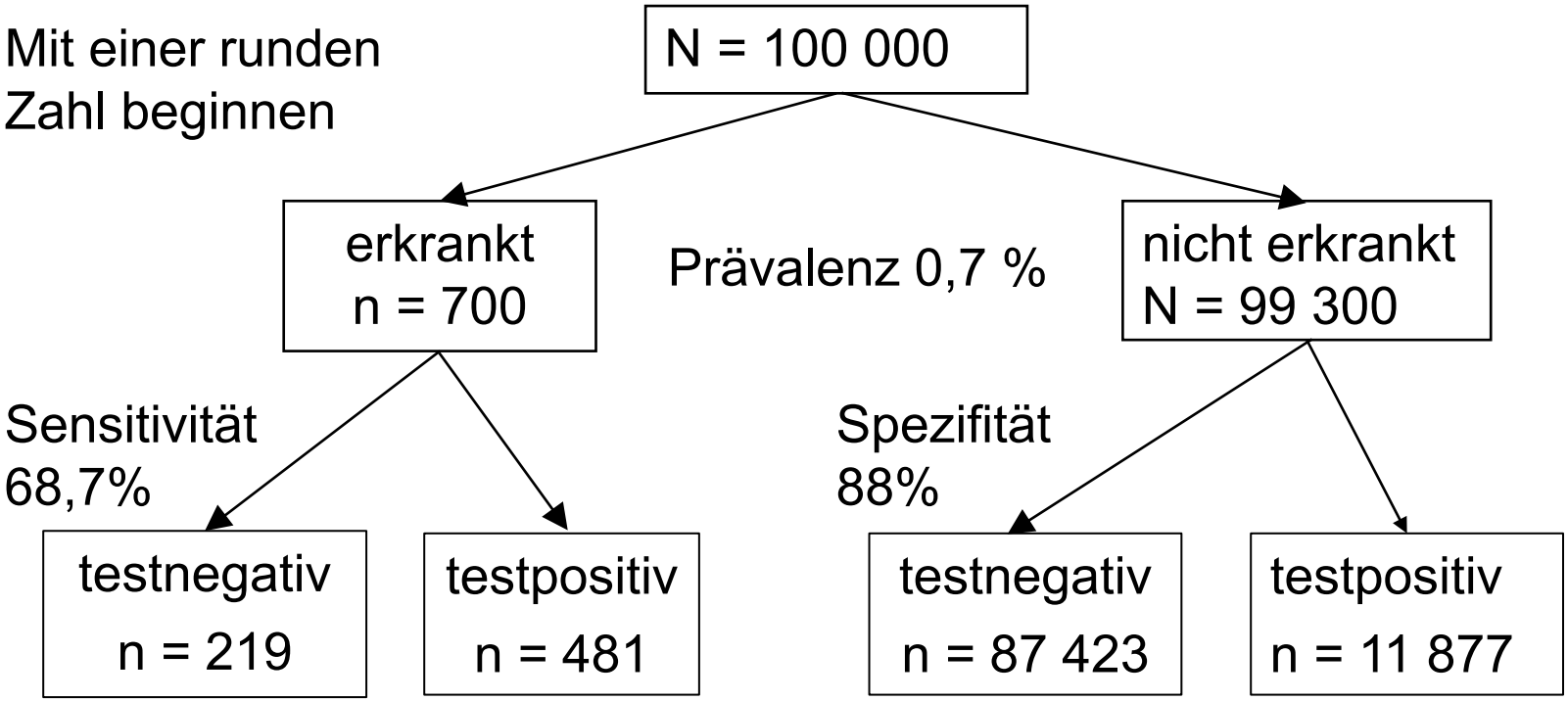


Lösung 3a: PPW mit Ereignisbaum



Positiv-prädiktiver Wert $:= \frac{\#\ richtig\ Positive}{\#\ Testpositive} =$

Lösung 3b: NPW mit Ereignisbaum



Negativ-prädiktiver Wert $:= \frac{\#\, richtig\ Negative}{\#\, Testnegative} =$

Bei negativem Testausgang kann sich Frau Hof ziemlich sicher sein, kein AF zu haben.

4.2.3 Ausblicke

1. Die Find-AF-2-Studie

Uhe T et al. Am Heart J 2023;265:66

Intensive heart rhythm monitoring to decrease ischemic stroke and systemic embolism—the Find-AF 2 study—rationale and design

Tobias Uhe, MD[a,#], Katrin Wasser, MD[b,#], Mark Weber-Krüger, MD[c,#], Wolf-Rüdiger Schäbitz, MD[d], Martin Köhrmann, MD[e], Johannes Brachmann, MD[f], Ulrich Laufs, MD[a], Martin Dichgans, MD[g,h,i], Götz Gelbrich, PhD[j,k], David Petroff, PhD[l], Christiane Prettin, PhD[l], Dominik Michalski, MD[m], Andrea Kraft, MD[n], Thorleif Etgen, MD[o], Peter D. Schellinger, MD[p], Hassan Soda, MD[q], Florian Bethke, MD[r], Michael Ertl, MD[s], Bernd Kallmünzer, MD[t], Martin Grond, MD[u], Katharina Althaus, MD[v], Gerhard F. Hamann, MD[w], Meinhard Mende, PhD[l], Marcus Wagner, PhD[l], Sonja Gröschel, MD[x], Timo Uphaus, MD[x], Klaus Gröschel, MD[x,†], and Rolf Wachter, MD[a,y,z,†], on behalf of the Find-AF 2 study group

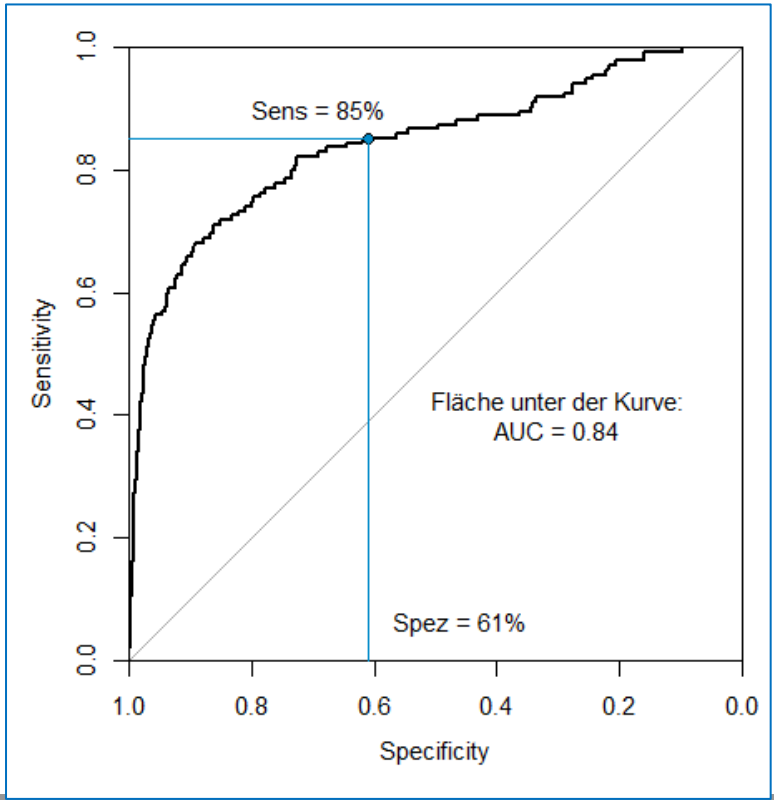
Göttingen, Germany; Essen, Germany; Coburg, Germany; Munich, Germany; Würzburg, Germany; Leipzig, Germany; Halle, Germany; Traunstein, Germany; Minden, Germany; Bad Neustadt, Germany; Ibbenbüren, Germany; Augsburg, Germany; Bielefeld, Germany

- Probanden werden nach dem 1. Schlaganfall randomisiert
  - Kontrollgruppe: 24h-EKG lt. Leitlinie
  - erweiterte Diagnostik: 7-Tages-EKG bzw. implantierter Event-Recorder
- wird AF diagnostiziert → Antikoagulation
- Endpunkt: Schlaganfall oder systemische Embolie
- Nachbeobachtung 2 – 6 Jahre

2. ROC – Kurve I

Soll die Diagnose anhand eines metrischen Merkmals getroffen werden, braucht man einen Schwellwert. Den zu finden hilft die ROC (Receiver-Operating-Characteristic)-Kurve.

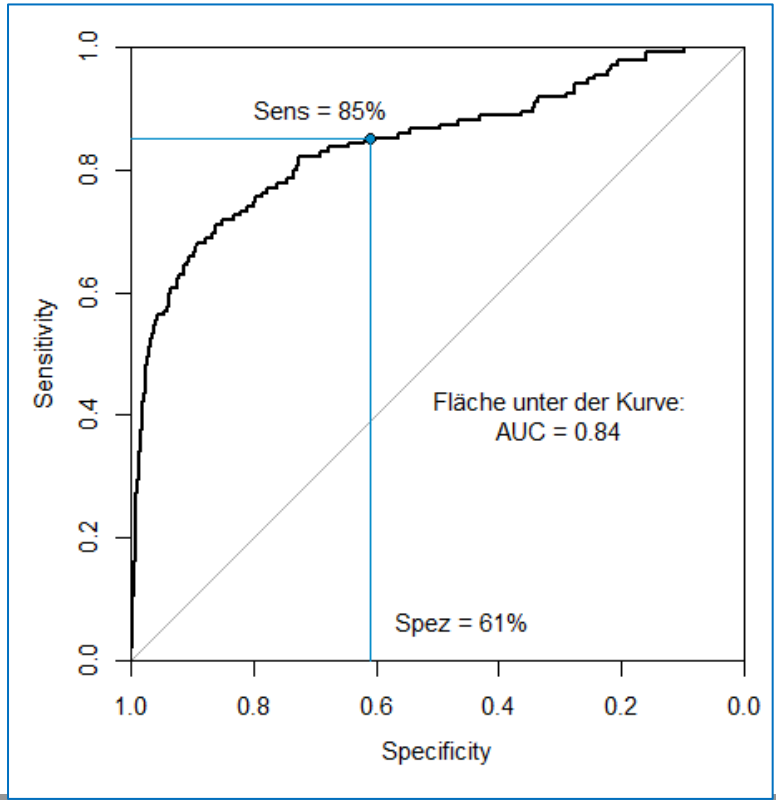
- Für alle x im Wertebereich des Merkmals
  - Anteil der richtig Positiven (Y-Achse, Sensitivität)
  - Anteil falsch Positiver (X-Achse, 1 - Spezifität).



2. ROC – Kurve II

Ein Merkmal trennt umso besser Erkrankte von Gesunden, je mehr die ROC-Kurve die Ecke oben links ausfüllt. Das misst die

- AUC (area under the curve)
- Die Hauptdiagonale $AUC = 0{,}5$ entspricht einer zufälligen Diagnose (z. B. per Münzwurf).

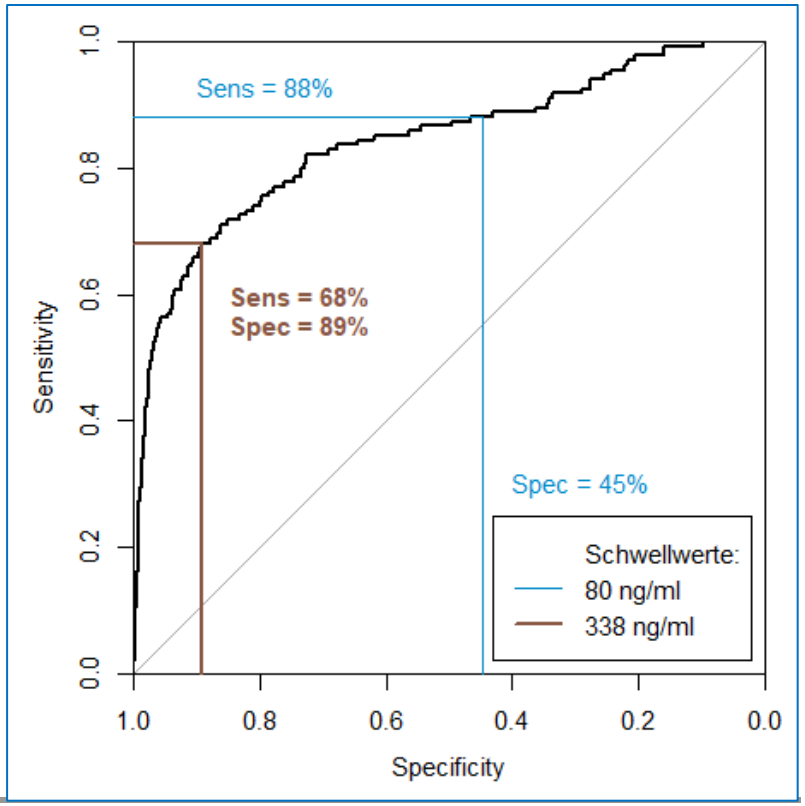


2. ROC – Kurve III

- Senkt man den Schwellwert,
  - steigt die Sensitivität, aber
  - fällt die Spezifität
- Erhöht man den Schwellwert,
  - sinkt die Sensitivität, aber
  - steigt die Spezifität

- Youden-Index: maximale Summe von Sensitivität und Spezifität
  - hier: bei 338 ng / ml
  - Sens = 68 %, Spec = 89%

Der optimale Schwellwert ist abhängig vom Testzusammenhang!

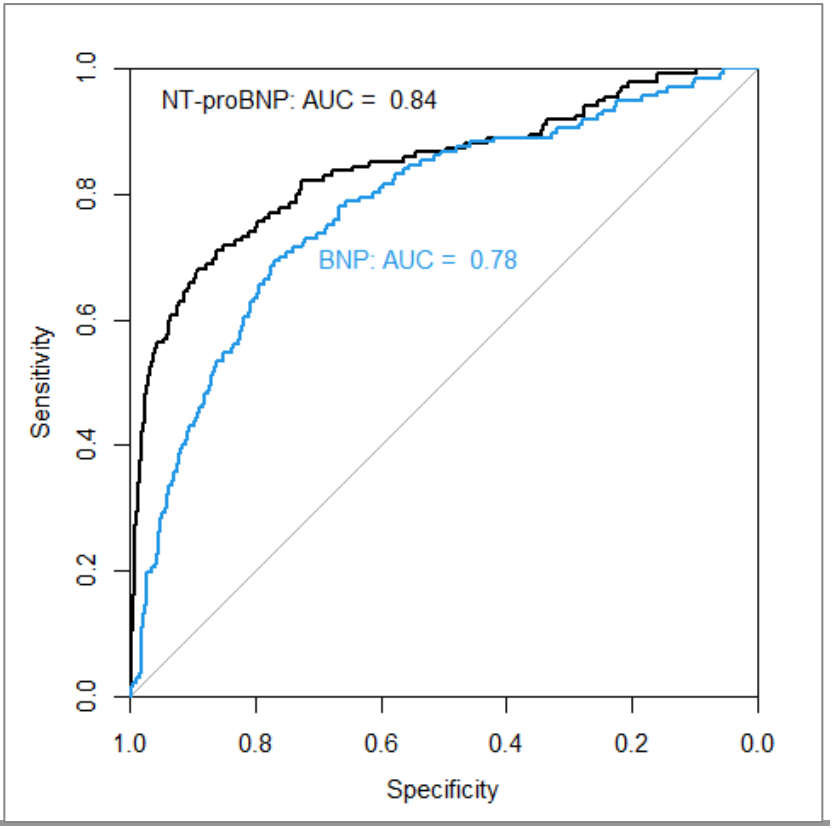


Vergleich von ROC - Kurven

Zeichnet man zwei ROC-Kurven in ein Koordinatensystem, kann man diese vergleichen:

- welches Merkmal besser zwischen krank / gesund unterscheidet
- ggf. welches Merkmal in bestimmten Bereichen besser ist.

Programme, wie z. B. R oder SPSS, bieten Tests für den Vergleich zweier AUCs an.



3. Kommunikation statistischer Sachverhalte

Historisches Drama: 1987-Konferenz über AIDS

- 22 Blutspender nach positivem ELISA-Test,
- 7 davon begingen Suizid
- selbst wenn ELISA- und WB-Test positiv gewesen wären, wäre die Chance für Infektion 50% (Prävalenz damals ca. 1 von 10 000.)

Wie vermittelt man verständlich statistische Ergebnisse?

- Herxheimer A. Communicating with patients about harms and risks. PLoS Med 2005; 2(2): e42
- Gigerenzer et al. Psychological Science in the Public Interest 2008;8(2):56-96.

Vision: gemeinsame Entscheidungsfindung von gut informierten Patienten und Ärzten.

Kommunikation statistischer Sachverhalte II

1. Es gibt in der Welt keine Sicherheiten, nur Wahrscheinlichkeiten.

~~*Der Test war positiv, Sie haben Krankheit XY!*~~

   – *Sie hatten einen positiven Test. In Ihrer Risikogruppe haben ca. 10 von 100 die Krankheit XY.*

2. Kommunizieren Sie Wahrscheinlichkeiten mit natürlichen Zahlen!

~~*Die Wahrscheinlichkeit für einen positiven Test unter der Bedingung, dass die Personen erkrankt ist, beträgt 93,2%*~~

   – *Bei 932 von 1 000 Erkrankten ist der Test positiv.*
   – *Aber auch 15 Nicht Erkrankte haben einen positiven Test.*
   – *Sie sind testpositiv. Von 1 000 positiv Getesteten sind im Mittel 473 tatsächlich erkrankt, also nur knapp die Hälfte.*

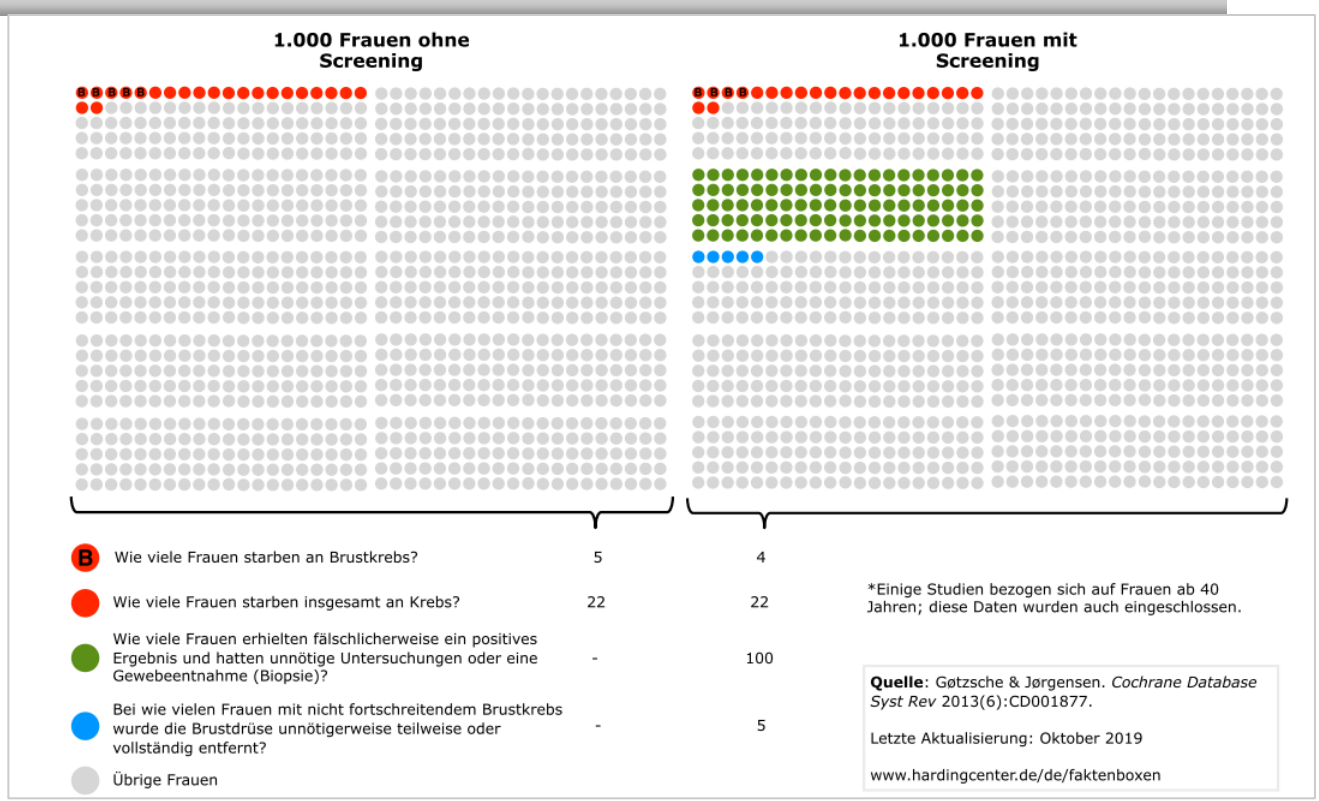
Kommunikation statistischer Sachverhalte III

3. Klären Sie die Patienten insbesondere bei Screening-Tests über die Folgen falsch negativer und falsch positiver Tests auf, z. B.
   - Durch einen positiven Test wird im Mittel bei 90 von 100 erkrankten Probanden die Erkrankung gefunden.
   - Einen positiven Test gibt es aber auch im Mittel bei 5 von 100 nicht Erkrankten.
   - Nach einem positiven Test führen wir i.d.R. eine Gewebeentnahme durch. Das ist mit diesen Risiken … verbunden.

Faktenboxen

Quelle:

https://hardingcenter.de/sites/default/files/2020-05/Icon%20Array_Mammographie_DE_20191012.pdf



- Faktenboxen stellen die Risiken ohne und mit Screening gegenüber
- Hier: Mammographie-Screening
- weitere Faktenboxen beim Harding-Zentrum für Risikokompetenz

Take home

Sensititivität und Spezifität charakterisieren die Güte eines Tests:

- Sensitivität: Wsk. für ein positives Ergebnis des Testes T unter der Voraussetzung, dass die Krankheit D vorliegt.
- Spezifität: Wsk. für ein negatives Ergebnis des Testes T unter der Voraussetzung, dass die Krankheit D nicht vorliegt.

Falsche Testergebnisse haben i.d.R. für den Probanden ungünstige Folgen.

Prädiktive Werte bestimmt man mit Hilfe der Bayesschen Formel, am einfachsten mit einem Ereignisbaum.

Kommunizieren Sie Wahrscheinlichkeitsaussagen mit Verantwortung und am besten mit Hilfe natürlicher Zahlen!

Grundzug Bayesscher Statistik

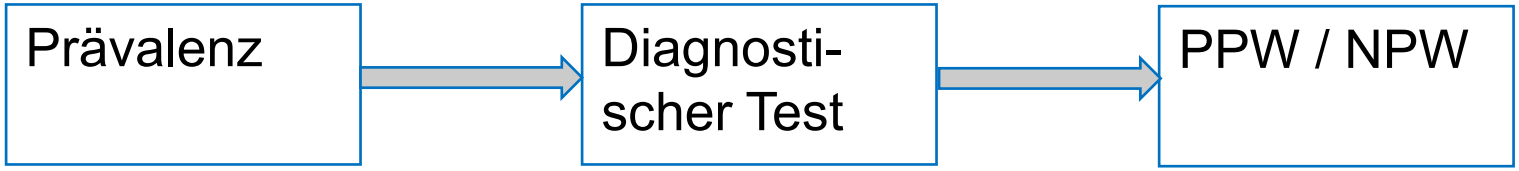
P

- Alles Vorwissen wird in einer A-priori-Verteilung quantifiziert.



- Kombination der Ergebnisse des Zufallsexperiments mit der Basisverteilung führt zur A-posteriori-Verteilung,
- die verbessertes Wissen widerspiegelt.

Am Beispiel des diagnostischen Tests:





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