

Das Rechnen mit Wahrscheinlichkeiten in der Humangenetik

Welches Allel an einem Genort oder Marker an ein Kind vererbt wird, folgt i. Allg. einer klaren Gesetzmäßigkeit, aber das Ereignis ist probabilistisch, es ist nicht sicher vorhersagbar. Den Zufall beschreibt man durch Wahrscheinlichkeiten, mit denen nach definierten Regeln gerechnet wird. Dazu formuliert man ein Wahrscheinlichkeitsmodell, das die Vorgänge in der Realität für die Fragestellung mit ihren wichtigsten Aspekten beschreibt. Diese Model-

lierung erlaubt es uns, Wahrscheinlichkeiten zu berechnen.

Grundregeln der Wahrscheinlichkeitsrechnung

In **Infobox 1** sind die Axiome und die daraus folgenden wichtigsten Rechenregeln für Wahrscheinlichkeiten zusammengestellt.

Ereignisse oder Ereignismengen lassen sich gut graphisch veranschaulichen, wie in **Abb. 1** gezeigt. In einem einfachen Modell stellt man sich die Wahrscheinlichkeit als Fläche relativ zur Grundgesamtheit Ω vor, die immer die Wahrscheinlichkeit 1 hat.

Bei der Komplementregel addieren sich die Wahrscheinlichkeiten zu 1 (**Abb. 1a**). Zum Beispiel betrachten wir bei den Risikoberechnungen oft die Wahrscheinlichkeit, dass eine Person Überträgerin einer Krankheitsmutation ist oder nicht. Haben die Mengen keine Ereignisse gemeinsam (sie sind disjunkt), wie in **Abb. 1b**, dann kann man für die Wahrscheinlichkeit von A oder B die Einzelwahrscheinlichkeiten addieren. Komplementäre Ereignisse sind durch ihre Definition disjunkt. Besonders wichtig sind bedingte Ereignisse; z. B. wollen wir wissen, wie groß unsere Wahrscheinlichkeit ist, krank zu sein, wenn wir wissen, ein diagnostischer Test ist positiv. Auch im Alltagsleben entsprechen bedingte Wahrscheinlichkeiten unserer differenzierteren Einschätzung von unsicheren Situationen, weil es Erfahrungen oder Vorwissen gibt. So ist die Wahrscheinlichkeit, bei Gewit-

ter vom Blitz erschlagen zu werden, insgesamt sehr gering, aber hoch, wenn wir uns bei Gewitter auf der Spitze des höchsten Berges in der Umgebung befinden.

Graphisch ist die Bedingung auf B ($B|B$) (**Abb. 1d**) die Einschränkung der Grundgesamtheit Ω auf B, die bedingte Wahrscheinlichkeiten $P(A|B)$ entspricht dem Flächenverhältnis von $A \cap B$ zu B (**Abb. 1c** und Regel 4). Die Regel von der totalen Wahrscheinlichkeit (Box 1, Regel 6) ist sehr nützlich, wenn man über die gesuchten Wahrscheinlichkeiten in disjunkten Untergruppen Aussagen machen kann. Beispielweise bei einem Paar, die beide heterozygote Überträger für dieselbe autosomal-rezessive Erbkrankheit sind und wegen einer Zwillingschwangerschaft wissen wollen, wie groß die Wahrscheinlichkeit ist, dass beide Feten nicht betroffen sind. Hier betrachtet man die beiden Fälle „eineiige Zwillinge“ (EZ) und „zweieiige Zwillinge“ (ZZ) getrennt und berechnet die allgemeine Wahrscheinlichkeit mit der Regel von der totalen Wahrscheinlichkeit. Als Wahrscheinlichkeiten für EZ und ZZ verwendet man die Populationshäufigkeiten.

Sind Ereignisse A und B unabhängig, so ist die Wahrscheinlichkeit für A und B das Produkt der Einzelwahrscheinlichkeiten. Wir gehen davon aus, dass ein Elternteil mit dem Genotyp Aa jedes Allel mit Wahrscheinlichkeit 0,5 an ein Kind weitergibt und dass dieser Prozess bei mehreren Kindern statistisch unabhängig ist.

Infobox 1

Definitionen

Ereignisse oder Ereignismengen werden durch Großbuchstaben abgekürzt, z. B. A
 $P(A)$ = die Wahrscheinlichkeit, dass das Ereignis A eintritt
 Ω die Menge aller Ereignisse im jeweiligen Modell (Grundgesamtheit)
 $A \cup B$ = A oder B oder A und B
 $A \cap B$ = Ereignisse, die in A und in B der Schnittmenge liegen
 $A - B$ = Ereignisse, die in A, aber nicht in B liegen
 $\bar{A} = \Omega - A$, die komplementäre Ereignismenge oder das Gegenereignis
 $A|B$ = A unter der Bedingung B
 A und B heißen unabhängig, wenn
 $P(A|B) = P(A)$

Regeln

- $0 \leq P(A) \leq 1$
- $P(\Omega) = 1$
- $P(A \cup B) = P(A) + P(B)$, falls A und B sich gegenseitig ausschließen
- bedingte Wahrscheinlichkeit
 $P(A|B) = P(A \cap B) / P(B)$
- die Komplementregel: $P(A) = 1 - P(\bar{A})$ und
- die Regel von der totalen Wahrscheinlichkeit: $P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A})$

C. Fischer · T. Grimm

Das Rechnen mit Wahrscheinlichkeiten in der Humangenetik

Zusammenfassung

Für die Risikoberechnungen in Familien werden die wichtigsten Begriffe und Regeln aus der Wahrscheinlichkeitsrechnung erläutert und die Beschreibung Mendelscher Erbgänge formalisiert.

Schlüsselwörter

Wahrscheinlichkeit · Bayes-Theorem · Mendelsche Segregation · Genetisches Modell · Likelihood-Funktionen

Probability calculations in human genetics

Abstract

The notation and rules of probability calculus for risk calculations in families and the parameters of Mendelian models are described.

Keywords

Probability · Bayes theorem · Mendelian segregation · Genetic model · Likelihood functions

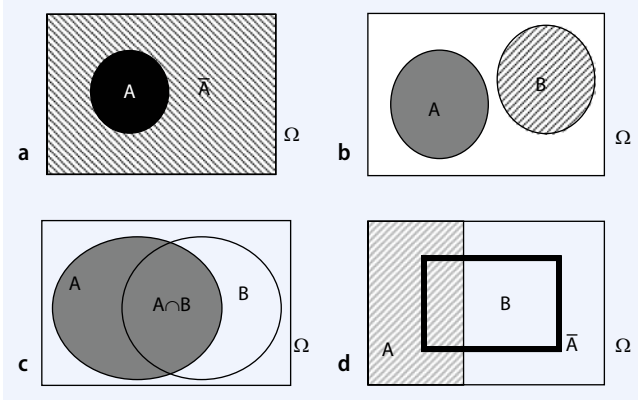


Abb. 1 ▲ Ereignismengen als Diagramme mit der Grundgesamtheit Ω . **a** Ereignismenge A und \bar{A} , ihr Komplement oder Gegenereignis, **b** sich gegenseitig ausschließende, disjunkte Ereignismengen, **c** Ereignisse mit nichtleerer Schnittmenge, **d** Zerlegung der Grundgesamtheit

Mendelsche Segregation und genetisches Modell

Die Weitergabe von Allelen an einen Genort von einem Elternteil an mehrere Nachkommen wird durch eine Binomial- oder Multinomialverteilung beschrieben. Zum Beispiel beträgt die Wahrscheinlichkeit $(0,5)^3$, dass das erste und das zweite und das dritte Kind eines heterozygoten Paares (Vater Aa und Mutter Aa) den Genotyp Aa haben.

Dies ist ein Beispiel für eine diskrete Zufallsgröße. Können bei einer Messgröße beliebig viele Werte vorkommen, wie etwa beim Blutdruck oder bei der Zeit bis zum Eintreten einer Krankheit, dann handelt es sich um stetige (auch quantitative) Zufallsvariablen. Hier ist die Wahrscheinlichkeit für das Auftreten eines bestimmten einzelnen Wertes 0. Daher betrachtet man die Wahrscheinlichkeiten von Bereichen wie etwa „vor dem Alter von 50 Jahren an Brustkrebs (BC) zu erkranken“. Die Verteilung ist durch eine Verteilungsfunktion

$$F(x) = P(BC - \text{Erkrankungsalter} \leq x)$$

beschrieben. In **Abb. 2** ist die Verteilungsfunktion für die Zeit bis zum Auftreten von Brustkrebs für Frauen mit einer Krankheitsmutation unter Annahme des Claus-Modells dargestellt. Danach liegt die Wahrscheinlichkeit, bis zum Alter von 50 Jahren an Brustkrebs zu erkranken, bei 29%, bis zum Alter von 60 Jahren bei 50% und die Lebenszeitwahrscheinlichkeit bei 85%.

Risikoberechnungen für Mendelsche Krankheiten, aber auch für genetisch komplexe Krankheiten mit monogenen Varianten basieren auf einem genetischen Modell mit den dazugehörigen Parametern. Im einfachen Fall sind dies Allelfrequenz des krankheitsverursachenden Allels, Penetranz und Phänotypkopierate. Wenn Fälle in der Familie aufgetreten sind, spielt die Populationsfrequenz nur eine geringe Rolle, für Handrechnungen lässt man sie außer Acht. Werte für die genetischen Parameter erhält man durch Schätzungen anhand von Stichproben. Man wählt das genetische Modell mit den Parametern, die das Vererbungsmuster in der Stichprobe am besten erklären (Maximum-Likelihood-Prinzip). Wenn mehrere Gene jeweils mit einem kleinen Einfluss auf das Krankheitsrisiko beteiligt sind, ist die Modellierung i. Allg. schwierig, da sich die verschiedenen genetischen Modelle statistisch nicht voneinander unterscheiden lassen.

Grundlage der Risikoberechnung in Familien ist ein genetisches Modell für die Krankheit. Die Mendelsche Segregation ist das einfachste und am meisten verwendete Modell für den Vererbungsmechanismus beim Menschen.

Die Beziehung zwischen Genotyp und Phänotyp, dem beobachtbaren Merkmal, wird durch die Penetranz beschrieben. Sie ist die bedingte Wahrscheinlichkeit, dass bei einer Person mit einem bestimmten Genotyp ein Phänotyp vorliegt:

$$\text{Penetranz} = P(\text{Phänotyp} | \text{Genotyp})$$

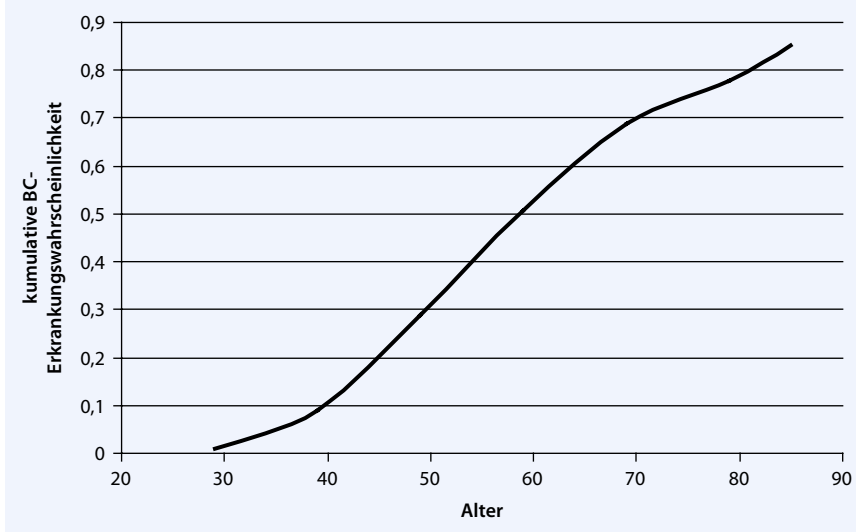


Abb. 2 ▲ Wahrscheinlichkeit für eine Frau, die BRCA1- oder BRCA2-Mutationsträgerin ist, bis zu einem bestimmten Alter an Brustkrebs (BC) zu erkranken [1]

Häufig wird angenommen, dass der Genort A biallelisch ist. Das disponierende Allel sei mit A und das normale Allel mit a bezeichnet. Damit hat ein Individuum vier mögliche geordnete Genotypen (AA, Aa, aA, aa) und drei mögliche ungeordnete Genotypen (AA, Aa, aa).

Der Phänotyp sei erkrankt, K, oder nicht erkrankt, G. Zur Beschreibung des Genortes gehört zunächst die Verteilung der Allelhäufigkeiten in der Population

$$P(A) = p \text{ und } P(a) = 1 - p = q$$

sowie die Penetranzen (Pen)

$$Pen_{AA} = P(K | AA)$$

$$Pen_{Aa} = P(K | Aa)$$

$$Pen_{aA} = P(K | aA)$$

$$Pen_{aa} = P(K | aa)$$

Diese Notation lässt sich auf mehrere Allele am Genort analog erweitern.

Die Allelhäufigkeiten und die Penetranzen sind zusammen die Parameter, die ein Gen beschreiben, das der Mendelschen Segregation folgt. Diese müssen bei der betrachteten Krankheit mittels Stichproben geschätzt werden. Bei klassischen monogenen Erkrankungen wird das Modell durch ein einzelnes Gen bestimmt und die Penetranzen nehmen nur die Werte 0 und 1 an. Bei einer klassischen autosomal-dominanten Erkrankung er-

kranken alle Träger des disponierenden Allels A, also

$$Pen_{AA} = Pen_{Aa} = Pen_{aA} = 1$$

und

$$Pen_{aa} = 0$$

Bei einer klassischen autosomal-rezessiven Erkrankung sind nur die Homozygoten AA erkrankt, also

$$Pen_{AA} = 1$$

und

$$Pen_{Aa} = Pen_{aA} = Pen_{aa} = 0$$

Als Verallgemeinerung lässt man zu, dass die Penetranzen zwischen null und eins liegen können. Setzte man diese Parameter als bekannt voraus, liefern sie Informationen über den i. Allg. unbekanntem Genotyp am Krankheitsgenort. Wenn z. B. gilt: $Pen_{aa}=0$ (man sagt, es liegen keine Phänokopien vor), dann können Kranke nicht den Genotyp aa haben. Liegen sowohl reduzierte Penetranz als auch Phänokopien vor, können Kranke und Gesunde alle Genotypen tragen, das genetische Modell erlaubt Wahrscheinlichkeitsaussagen darüber.

Oft sind die Penetranzen zusätzlich von anderen Faktoren abhängig. Bei Brustkrebs sind sie vom Geschlecht und vom Alter abhängig.

Infobox 2

Bayessche Formel

Beispiel 1: Eine Person trägt eine Krankheitsmutation am Krankheitsgenort, ist also Aa oder AA, das Gegenereignis \bar{A} beschreibt den Genotyp zweier Normalallele aa; G= gesund:

$$P(Aa, AA | G) = \frac{P(G | Aa, AA)P(Aa, AA)}{P(G)} = \frac{P(G | Aa, aa)P(Aa, AA)}{P(G | Aa, AA)P(Aa, AA) + P(G | aa)P(aa)}$$

Beispiel 2: Person II.1 in **Abb. 3** fragt nach ihrer Wahrscheinlichkeit, heterozygot (Aa) zu sein, Penetranz=0,6 und keine Phänokopien, vereinfachend wird nur der Genotyp Aa betrachtet:

$$P(Aa | G) = \frac{P(G | Aa)P(Aa)}{P(G)} = \frac{0,6 * 0,5}{0,6 * 0,5 + 1 * 0,5} = 0,286$$

Die Bayessche Formel

In der Genetik wollen wir Aussagen über Wahrscheinlichkeiten in Familien machen und oft den Genotyp am Krankheitsgenort berücksichtigen. Daher spielen abhängige Ereignisse eine Rolle. Wie wahrscheinlich ist die Erkrankung, wenn einen bestimmten Genotyp vorliegt? Die Wahrscheinlichkeit für das Auftreten einer monogenen Krankheit bei einer Ratsuchenden ist sehr unterschiedlich, je nach ihrem Genotyp am verursachenden Genort, oder je nach ihrer Familienvorgeschichte. Das wichtigste Hilfsmittel für diese Berechnungen ist die Bayessche Formel (**Infobox 2**). Mit ihrer Hilfe kann man bedingte Wahrscheinlichkeiten durch Herumdrehen der Bedingung berechnen, daher heißt sie auch Umkehrformel. Wissen wir z. B. aufgrund eines validierten genetischen Modells einer autosomal-dominanten Erkrankung, dass eine Person mit einer Krankheitsmutation, Ereignis heterozygot Aa, mit Wahrscheinlichkeit 0,6 erkrankt, $P(Aa)=0,6$, und eine Person mit dem Genotyp aa mit Wahrscheinlichkeit 0, $P(aa)=0$, dann können wir mithilfe der Bayesschen Formel umgekehrt berechnen, mit welcher Wahrscheinlichkeit bei einer gesunden Tochter einer Erkrankten (heterozygote Überträ-

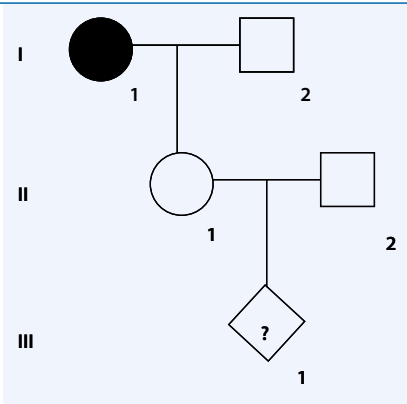


Abb. 3 ▲ Stammbaum mit einer autosomal-dominanten Erkrankung mit reduzierter Penetranz

gerin mit aA) eine Krankheitsmutation vorliegt: $P(Aa|G)$ ($G = \text{gesund}, K = \text{krank}$). Die bedingten Wahrscheinlichkeiten auf der rechten Seite der Bayesschen Formel sind bekannt: $P(G|A) = P(G|Aa)$ ist die Penetranz und $P(G|\bar{A}) = P(G|aa) = 1 - P(K|aa)$ entspricht 1-Phänokopierate. Nun brauchen wir nur noch Informationen über die Genotypwahrscheinlichkeiten $P(Aa)$ und $P(aa)$, die sog. A-priori-Wahrscheinlichkeiten. Sie müssen entsprechend der Situation gewählt werden, in der man die Berechnung beginnt. Das sind die Populationshäufigkeiten, wenn keine Familienvorgeschichte oder andere Informationen zur Verfügung stehen oder wenn ein Fall vorliegt wie z. B. im Stammbaum in **Abb. 3**, dort ist II.1 aufgrund der Transmissionswahrscheinlichkeiten mit 50%iger Wahrscheinlichkeit Aa oder aa am Krankheitsgenort.

Benutzung des Bayesschen Rechentableaus

Man kann die Wahrscheinlichkeiten in einem Rechentableau darstellen (**Tab. 1**) und die Berechnung so von Hand oder mit einfachen Hilfsprogrammen (z. B. Excel) durchführen, was wir für verschiedene Beispiele in den Artikeln dieses Heftes angewandt haben.

Die Tabelle enthält folgende Informationen:

1. Die A-priori-Wahrscheinlichkeiten für die möglichen Genotypen Aa und aa ergeben sich hier aus der Mendelschen Regel.
2. Die bedingten Wahrscheinlichkeiten für ein beobachtetes Phänomen

Hier steht eine Anzeige.

 Springer

Tab. 1 Berechnung für den Stammbaum in **Abb. 1**

Penetranz=0,6			
II.1	Aa	aa	Summe
1: A-priori-Wahrscheinlichkeit	0,5	0,5	
2: Bedingte Wahrscheinlichkeit: II.1 gesund	1-0,6	1	
3: Verbundene Wahrscheinlichkeit	0,2	0,5	0,7
4: A-posteriori-Wahrscheinlichkeit	0,2/0,7 ≈ 0,286	0,5/0,7 ≈ 0,714	
Erkrankungswahrscheinlichkeit für ein Kind von II.1 ≈ 0,086.			

- „II.1 ist gesund“, wenn angenommen wird, die Spaltenüberschrift ist wahr.
- Die verbundenen Wahrscheinlichkeiten, das Produkt aus den vorangegangenen Einträgen in der jeweiligen Spalte. Die Summe der verbundenen Wahrscheinlichkeiten müssen nicht 1 ergeben, daher ist eine Normierung zur Bestimmung der A-posteriori-Wahrscheinlichkeit erforderlich (Zeile 4)
 - Die verbundenen Wahrscheinlichkeiten müssen so normiert werden, dass sie als Summe 1 ergeben, man dividiert sie durch ihre Summe. Da zusätzlich Informationen berücksichtigt wurden, nennt man sie A-posteriori-Wahrscheinlichkeit.
 - Erkrankungswahrscheinlichkeit für ein Kind von II.1 ist dann: A-posteriori-Heterozygotenwahrscheinlichkeit * 0,5 * Penetranz;
hier: $0,286 * 0,5 * 0,6 = 0,086$

Zur Berechnung der Heterozygotenwahrscheinlichkeit von II.1 und der Erkrankungswahrscheinlichkeit eines Kindes von II.1 haben wir also die folgenden Schritte durchgeführt:

- Stammbaum (**Abb. 3**)
- Genetisches Modell:
autosomal-dominanter Erbgang mit zwei Allelen im Hardy-Weinberg-Gleichgewicht:
a=Normal-Allel mit einer Frequenz p und
A=Defekt-Allel mit einer Frequenz q
die Penetranz beträgt: Pen=0,6
keine Phänokopien
keine Neumutation
- A-priori-Wahrscheinlichkeit:
Die Tochter (Ratsuchende) ist entweder Überträgerin (Aa) einer autosomal-dominanten Krankheit mit 60%iger Penetranz oder nicht (aa):
 $P(Aa)=0,5$ und $P(aa)=0,5$

- Bedingte Wahrscheinlichkeit:
Die Tochter ist gesund (G) unter der Bedingung, dass sie Überträgerin ist: $P(G|Aa)=1-Pen=1-0,6=0,4$ oder sie ist gesund, weil sie keine Überträgerin ist: $P(G|aa)=1$
- Berechnung in **Tab. 1**
- Erkrankungswahrscheinlichkeit für ein Kind von II.1:
 $P(III.1 = K) = 0,286 * 0,5 * 0,6 = 0,086$

Fazit für die Praxis

Diese kurze Einführung in die Grundlagen der Wahrscheinlichkeitsrechnung und der Bayesschen Formel zeigt, wie wichtig diese Grundlagen für die Risikoberechnung in der genetischen Beratung sind.

Korrespondenzadresse

Dr. C. Fischer
Institut für Humangenetik
Universität Heidelberg
Im Neuenheimer Feld 366
69120 Heidelberg
cfischer@uni-hd.de

Interessenkonflikt. Die korrespondierende Autorin gibt an, dass kein Interessenkonflikt besteht.

Literatur

- Bickeböller H, Fischer C (2007) Einführung in die Genetische Epidemiologie. Springer, Heidelberg
- Claus EB, Risch N, Thompson WD (1991) Genetic analysis of breast cancer in the cancer and steroid hormone study. Am J Hum Genet 48(2):232–421
- Weiß C, Rzany B (2010) Basiswissen Medizinische Statistik. Springer Medizin, Heidelberg

Verlust eines Gen-Regulators entscheidend für seltene Form von Hautkrebs

Wissenschaftlern ist es gelungen, einen spezifischen Genverlust beim Sézary-Syndrom nachzuweisen, einer besonders aggressiven Form des T-Zell-Lymphoms aus der Gruppe der primären Hautlymphome.

Die Forscher untersuchten hoch aufgereinigte Tumorzellen von Patienten mit Sézary-Syndrom mit der so genannten „array comparative genomic hybridization“-Technik auf bisher unbekannte genetische Veränderungen. Dabei identifizierten sie deletierte Bereiche im Erbgut dieser Tumorzellen, die für den Transkriptionsfaktor E2A kodieren. E2A ist wichtig für die natürliche Lymphozyten-Entwicklung, und ein Verlust dieses Gens führt bei Mäusen zur Entstehung aggressiver T-Zell-Lymphome. Ein Genverlust in einer der verschiedenen humanen Lymphomklassen war bisher jedoch noch nicht gefunden worden.

Die Forscher identifizierten zudem mehrere E2A-regulierte Gene und Signalwege in den Tumorzellen, deren Deregulation zur Entwicklung und zum schnellen Wachstum der Tumoren beitragen können. Damit kommt E2A im Menschen erstmals die Stellung eines Tumorsuppressors zu. Die Forscher hoffen, dass diese Erkenntnisse Grundlage für die Entwicklung neuer Behandlungskonzepte sein könnten.

Literatur:

Steinger A, Möbs M, Ullmann R et al. (2011) Genomic loss of the putative tumor suppressor gene E2A in human lymphoma. J Exp Med 208:1585-1593

Quelle: Charité – Universitätsmedizin Berlin, www.charite.de