



Allgemeine Empfehlungen der Spurenkommission zur statistischen Bewertung von DNA-Datenbank-Treffern

**P. M. Schneider, H. Schneider, R. Fimmers, W. Keil,
G. Molsberger, W. Pflug, T. Rothämel, M. Eckert,
H. Pfeiffer, B. Brinkmann**

Publikation in der Zeitschrift „Rechtsmedizin“ im April 2010

Die deutsche DNA-Analyse-Datei

- DAD-Statistik Ende 2009:
 - 668.700 Personen
 - 166.500 nicht zugeordnete Spuren
- Treffer bisher:
 - 73.100 Spur-Personen-Treffer
 - 20.700 Spur-Spur-Treffer
- Jede dritte neu eingestellte Spur erzielt einen Treffer mit einer Person oder einer anderen Spur

Das Problem der Teilprofile

- 55.000 Datensätze von Personen mit nur fünf STR-Systemen
- 2.000 Datensätze von Spuren mit sechs STR-Systemen
- 7.500 Datensätze von Spuren mit sieben STR-Systemen
- Im internationalen Datenaustausch (gemäß Vertrag von Prüm) fällt SE33 weg:
 - Fast alle Datensätze der DAD haben nur sieben Systeme!

Behandlung von Datenbanktreffern

- Folge eines Datenbanktreffers:
 - Ergebnisüberprüfung zur Richtigkeit der Daten
 - Häufig erfolgt Anklage durch die Staatsanwaltschaft ohne spezielle Biostatistik!
- Auf Anforderung:
 - „Treffergutachten“ meist mit Angabe der Häufigkeit des DNA-Profiles in der Bevölkerung
 - Danach Anklage durch die StA
- **Was ist hierbei relevant?**
 - Die **Häufigkeit** dieses DNA-Profiles in der allgemeinen **Bevölkerung**, oder die **Wahrscheinlichkeit, in der Datenbank** ein übereinstimmendes Profil zu finden?

Datenbank-Treffer – eine Lotterie?



- Ein Tipp (6 aus 49) - Gewinnchance 1 : 13.000.000
- Eine Million Tipps - Gewinnchance 1 : 13

Trefferwahrscheinlichkeit

- Ein-Personen-Spur ohne Datenbank-Recherche
 - **Frequenz f** des DNA-Profiles entspricht der Wahrscheinlichkeit der zufälligen Übereinstimmung einer beliebigen Person aus der Bevölkerung mit der Spur
(entspricht Einzeltipp im Lotto mit Gewinnchance 1 : 13.000.000)
 - d.h. es müssen 13 Mio. beliebige Personen untersucht werden, um einen Zufallstreffer zu erhalten.

Trefferwahrscheinlichkeit
= Identitätswahrscheinlichkeit
(engl. *match probability*)



Trefferwahrscheinlichkeit in Datenbank

- Ein-Personen-Spur mit Treffer nach Recherche
 - 13 Mio. beliebige Personen = ein Zufallstreffer
 - Werden diese Personen zufällig auf 13 Töpfe mit je einer Million aufgeteilt ...



...dann ist die Chance auf einen Zufallstreffer für jeden einzelnen Topf 1 : 13

Trefferwahrscheinlichkeit in Datenbank

- Ein-Personen-Spur nach Datenbank-Recherche

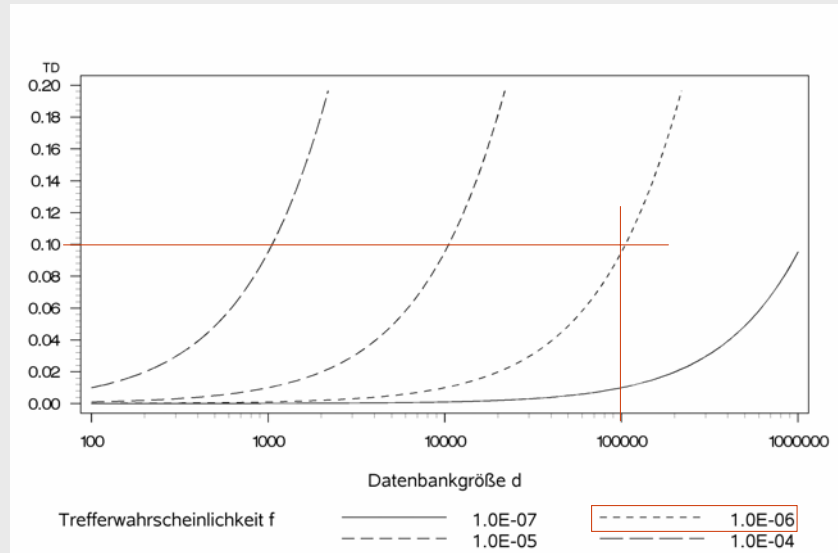
f - Frequenz des gesuchten DNA-Profiles

d - Größe der Datenbank

- Die Wahrscheinlichkeit für (mindestens) einen Datenbanktreffer ist

$$T_D \approx f \cdot d$$

Einfluss der Datenbankgröße auf die Wahrscheinlichkeit eines Treffers bei Berücksichtigung unterschiedlicher Häufigkeiten des DNA-Profiles



Berechnungsbeispiel nach Datenbanktreffer

- Tatortspur S mit 10 STR-Systemen
 - Häufigkeit (Mittleuropäer) $f = 1,0 \times 10^{-11}$
- Recherche in einer Datenbank
 - Größe $d = 1,0 \times 10^6$

$$T_D \approx f \cdot d = 1,0 \cdot 10^{-11} \cdot 1,0 \cdot 10^6 = 1,0 \cdot 10^{-5}$$

- "Die Datenbankrecherche hat eine Übereinstimmung der Merkmalskombination der Spur mit dem Datensatz einer Person ergeben. Die Wahrscheinlichkeit für das zufällige Auftreten eines solchen Ereignisses zum Zeitpunkt der Recherche beträgt ca. eins in 100.000."

Berechnung des Likelihood-Quotienten (ohne Datenbank-Recherche)

H_0 : Die Spur S stammt vom Tatverdächtigen.

- Dies entspricht der Wahrscheinlichkeit $P=1$ für die Beobachtung der in der Spur vorliegenden Merkmalskombination unter Annahme von H_0 .

H_1 : Die Spur S stammt von einer unbekanntem und mit dem Tatverdächtigen unverwandten Person aus der Vergleichsbevölkerung.

- Die Wahrscheinlichkeit P entspricht dann der Frequenz f der beobachteten Merkmalskombination in der relevanten Vergleichsbevölkerung.

$$LQ = \frac{P(S | H_0)}{P(S | H_1)} = \frac{1}{f}$$

Berechnung des LQ nach Datenbank-Recherche

H_0 : Die Spur S stammt von einer Person aus der Datenbank.

- *Frage*: Wie groß ist die Wahrscheinlichkeit für die vorgefundene Merkmalskombination unter der Annahme, dass eine beliebige Person aus der Datenbank Spurenleger ist?
 - Die Merkmale der Personen in der Datenbank sind bekannt.
 - Wenn Spurenleger in Datenbank vorhanden:
Wahrscheinlichkeit für die Beobachtung der Spur = relative Häufigkeit der Merkmalskombinationen in der Datenbank.
Wenn genau ein Treffer in der Datenbank erzielt wird, entspricht diese Wahrscheinlichkeit in einer Datenbank, in der genau eine Person enthalten ist, dem Wert 1.
- Wenn mehr als eine Person in der Datenbank erfasst ist, dann:

$$P(S | H_0) = \frac{1}{d}$$

Berechnung des LQ nach Datenbank-Recherche

H_1 : Die Spur S stammt von einer Person, die nicht in der Datenbank erfasst ist.

- Frage: Wie groß ist die Wahrscheinlichkeit für die beobachtete Merkmalskombination unter der Annahme, eine beliebige Person aus der Vergleichsbevölkerung ist Spurenleger?
- Diese Wahrscheinlichkeit P entspricht wieder der Häufigkeit des DNA-Profiles in der relevanten Vergleichsbevölkerung:

$$P(S | H_1) = f$$

- Also lautet der gesamte Ausdruck:

$$LQ = \frac{P(S | H_0)}{P(S | H_1)} = \frac{1}{d \cdot f}$$

Beispiel für LQ-Berechnung nach Datenbanktreffer

- Tatortspur S mit 10 STR-Systemen
 - Häufigkeit (Mitteleuropäer) $f = 1,0 \times 10^{-11}$
- Recherche in einer Datenbank
 - Größe $d = 1,0 \times 10^6$

$$LQ = \frac{1}{d \cdot f} = \frac{1}{10^6 \cdot 10^{-11}} = 1 \cdot 10^5$$

- „Die Datenbankrecherche hat eine Übereinstimmung der Merkmalskombination der Spur mit dem Datensatz einer Person ergeben. Die Wahrscheinlichkeit für die Beobachtung der Spur ist 100.000 Mal größer, wenn der Spurenleger aus der Datenbank stammt, als wenn er aus der Vergleichsbevölkerung stammt.“

Beispiel einer DAD-Recherche

(durchgeführt am 16.04.2009, d = 630.162)

Anzahl STR-Systeme*	Genotyp-Frequenz	Treffer erwartet	Treffer beobachtet
6 Systeme (D21S11, VWA, TH01, FGA, D3S1358, D8S1179)	1 in 2,2 Mio.	0,28	0
5 Systeme (D21S11, VWA, TH01, FGA, D3S1358)	1 in 225.000	3	5
4 Systeme (D21S11, VWA, TH01, FGA)	1 in 25.000	25	48

*ohne SE33, da international nicht vergleichbar

Ergänzung von STR-Daten nach Datenbank-Treffer

- DAD-Treffer auf der Basis von acht Systemen
- Eine erneut entnommene Probe wird für insgesamt 11 STR-Systeme untersucht
- Wie wird gerechnet?
 1. Korrigierte DAD-Statistik nach Treffer für acht Systeme
 2. Die drei zusätzlichen Systeme können als unabhängige Information ohne Datenbank-Korrektur einbezogen werden

Beispiel einer ergänzenden Berechnung

Biostatistik nach Datenbank-Recherche		DAD-Größe Personen	
		12/2003	278.000
Fall	I-999/09	Dez 04	320.000
Spur	D-1234/09	12/2005	367.000
Person	N.N. 1972	12/2006	439.000
		12/2007	525.000
		12/2008	612.000
		09/2009	658.000
	Datum	Wert	Frequenz
		(1 in .. Pers.)	
Trefferwahrscheinlichkeit (DNA odds)	2,12E+10		4,72E-11
Datenbankgröße	06.12.2007 525.000		
Trefferwahrscheinlichkeit nach Recherche	40.381		2,48E-05
zusätzliche Systeme (DNA odds)	10.000		
komb. Trefferwahrscheinlichkeit	4,04E+08		

Empfehlung der Spurenkommission

- Relevant für die biostatistische Beurteilung einer Übereinstimmung zwischen einer Spur und einer Person nach einer DAD-Recherche ist neben der Häufigkeit der Merkmalskombination grundsätzlich auch die Anzahl der Personen, die zum Zeitpunkt des Treffers in der Datenbank erfasst waren.
- Diese Information muss daher bei jedem Recherche-Ergebnis mit angegeben sein und als Grundlage der Bewertung herangezogen werden.