

## Die Beziehung zwischen genotypischen und phänotypischen Vaterschaftsplausibilitäten

Max P. Baur

Institut für Medizinische Statistik, Dokumentation und Datenverarbeitung der Universität Bonn und Institut für Gerichtliche Medizin der Universität Bonn (BRD)

Eingegangen am 25. Oktober 1974

### The Relation between Genotypical and Phenotypical Plausibilities of Paternity

*Summary.* Applying the formula of Okajima [2, 3] to the genotypical possibilities of a given phenotypical child-mother-putativ father constellation one can compute for each of these possibilities a genotypical probability of paternity. It is possible to determine the discrete distribution function, the mean value, and the variance of these genotypical probabilities of paternity and it is shown that the phenotypical probability of paternity is equal to the mean value of the genotypical probabilities of paternity.

*Zusammenfassung.* Wendet man bei gegebener phänotypischer Kind-Mutter-Putativvater-Konstellation die Formel von Okajima [2, 3] auf die genotypisch möglichen Fälle an, so erhält man für jede dieser Möglichkeiten einen genotypischen  $W$ -Wert. Die diskrete Verteilungsfunktion, der Erwartungswert und die Varianz dieser genotypischen  $W$ -Werte lassen sich bestimmen, und es zeigt sich, daß der phänotypische  $W$ -Wert gleich dem Erwartungswert der genotypischen  $W$ -Werte ist.

*Key words:* Blutgruppen, Biostatistik — Vaterschaftswahrscheinlichkeit, Genotyp, genotypische  $W$ -Werte, phänotypischer  $W$ -Wert.

Ausgangspunkt für die meistverwendete Methode zur biostatistischen Abstammungsbegutachtung [1] ist die phänotypische Konstellation des Doublets Mutter—Kind und des Eventualvaters in den untersuchten Blutgruppensystemen. Mit Hilfe der Formel von Okajima [2, 3]

$$\frac{Y_p}{X_p} = \frac{V_p \cdot S(M_g \times K_g)}{S(V_g \times M_g \times K_g)} \quad (1)$$

werden die Werte  $Y_p$  und  $X_p$  durch Symbole für die Genhäufigkeiten dargestellt, wobei hier wie auch im folgenden der Index  $p$  angibt, daß die indizierte Größe sich auf die phänotypische Konstellation bezieht, während der Index  $g$  anzeigen, daß es sich hierbei um eine genotypische Größe handelt.

Hierbei bedeuten:

$V_p$	Wahrscheinlichkeit des Phänotyps des Eventualvaters.
$S(M_g \times K_g)$	Summe der Wahrscheinlichkeiten aller bei den vorliegenden Phänotypen möglichen genotypischen Mutter-Kind-Verbindungen.

$S(V_g \times M_g \times K_g)$  Summe der Wahrscheinlichkeiten aller bei den vorliegenden Phänotypen möglichen genotypischen Eventualvater-Mutter-Kind-Verbindungen.

Die so bestimmten Werte für  $Y_p$  und  $X_p$  werden in den Ansatz von Essen-Möller [4] eingesetzt und damit die Vaterschaftsplausibilität für die phänotypische Konstellation

$$W_p = \frac{q \cdot X_p}{q \cdot X_p + (1-q) \cdot Y_p} = \frac{1}{1 + \frac{(1-q) \cdot Y_p}{q \cdot X_p}} \quad (2)$$

berechnet. Hierbei bedeutet  $q$  die a priori-Wahrscheinlichkeit für Vaterschaft, die in den meisten Fällen gleich 0,5 gesetzt wird, wodurch sich die Essen-Möller-Formel zu

$$W_p = \frac{X_p}{X_p + Y_p} = \frac{1}{1 + \frac{Y_p}{X_p}} \quad (3)$$

vereinfacht.

Auf Grund der Verwendung unvollständiger Blutgruppensysteme ist es jedoch sehr häufig der Fall, daß bei vorgegebener phänotypischer Konstellation mehrere genotypische Interpretationen vorhanden sind.

Hat man allgemein bei einer phänotypischen Eventualvater-Mutter-Kind-Konstellation in mehreren Blutgruppensystemen für die phänotypische Blutgruppenkombination des Eventualvaters  $n$  verschiedene genotypische Interpretationen und für die phänotypische Blutgruppenkombination des Dubletts Mutter—Kind  $m$  verschiedene genotypische Interpretationen, so ergeben sich für diesen einen phänotypischen Fall  $n \cdot m$  genotypische Möglichkeiten.

Es ist nicht schwer, die Formel von Okajima [2, 3] für diese genotypischen Terzette umzuwandeln, und man erhält für  $1 \leq i \leq n$  und  $1 \leq j \leq m$

$$\frac{Y_{i,j}}{X_{i,j}} = \frac{(V_g)_i \cdot (M_g \times K_g)_j}{(V_g \times M_g \times K_g)_{i,j}}. \quad (4)$$

Hierbei bezeichnet nun wie auch im folgenden der Index  $i$  die  $i$ -te Genotypenkonstellation des Eventualvaters und der Index  $j$  die  $j$ -te Genotypenkonstellation des Dubletts Mutter—Kind. Weiter ist analog zum phänotypischen Fall:

- |                                     |   |
|-------------------------------------|---|
| $(V_g)_i$                           | die Wahrscheinlichkeit der $i$ -ten Genotypenkonstellation des Eventualvaters,  |
| $(M_g \times K_g)_j$                | die Wahrscheinlichkeit der Genotypenkombination des $j$ -ten Dubletts,  |
| $(V_g \times M_g \times K_g)_{i,j}$ | die Wahrscheinlichkeit der Genotypenkombination der $j$ -ten Dublettenkonstellation mit der $i$ -ten Genotypmöglichkeit des Eventualvaters unter der Voraussetzung der Vaterschaft. |

Setzt man die Größen  $X_{i,j}$  und  $Y_{i,j}$  in die Essen-Möller-Formel ein, so ergibt sich für jede der  $n \cdot m$  möglichen genotypischen Konstellationen die Vaterschaftsplausibilität

$$W_{i,j} = \frac{q \cdot X_{i,j}}{q \cdot X_{i,j} + (1-q) \cdot Y_{i,j}} = \frac{1}{1 + \frac{(1-q) \cdot Y_{i,j}}{q \cdot X_{i,j}}}. \quad (5)$$

Man erhält also an Stelle der allgemeinen phänotypischen Vaterschaftsplausibilität  $W_p$  die genotypische Vaterschaftsplausibilität  $P_{i,j}$  für die genotypisch mögliche Terzettkonstellation  $(i,j)$ , die aus der  $i$ -ten Genotypenkonstellation des Eventualvaters und der  $j$ -ten Genotypenkonstellation des Dubletts Mutter—Kind besteht. Da es für den Gutachter in vielen Fällen unmöglich ist, die Genotypen genau zu bestimmen, ist es als nächstes nötig, die Wahrscheinlichkeit  $P_{i,j}$  des genotypischen Terzetts  $(i,j)$  innerhalb des phänotypischen Falls zu bestimmen.

Betrachtet man die auch der Okajima-Formel zugrundeliegenden Genfrequenzen als Wahrscheinlichkeiten (und nicht als Schätzwerte für diese Wahrscheinlichkeiten — eine Voraussetzung, die auch in der phänotypischen Betrachtung ständig benutzt wird), so erhält man diese Wahrscheinlichkeit durch

$$P_{i,j} = \frac{q \cdot X_{i,j} + (1-q) \cdot Y_{i,j}}{\sum_{k=1}^n \sum_{l=1}^m (q \cdot X_{k,l} + (1-q) \cdot Y_{k,l})}. \quad (6)$$

Man hat also nun jeder Genotypenkonstellation  $(i,j)$  und ihrer zugehörigen Vaterschaftsplausibilität  $W_{i,j}$  ihre Wahrscheinlichkeit  $P_{i,j}$  innerhalb des gegebenen Phänotyps zugeordnet und kann folglich die diskrete Verteilungsfunktion  $F$  der  $W_{i,j}$ -Werte angeben:

$$F(z) = P(W_g \leqq z) = \sum_{(i,j): W_{i,j} \leqq z} P_{i,j}. \quad (7)$$

Für jede Verteilungsfunktion ist es von Interesse, den Erwartungswert und die Varianz zu berechnen, und man erhält als wohl interessantestes Ergebnis dieser Betrachtungen:

$$\begin{aligned} E(W_g) &= \sum_{i=1}^n \sum_{j=1}^m W_{i,j} \cdot P_{i,j} \\ &= \sum_{i=1}^n \sum_{j=1}^m \left[ \frac{q \cdot X_{i,j}}{q \cdot X_{i,j} + (1-q) \cdot Y_{i,j}} \cdot \frac{q \cdot X_{i,j} + (1-q) \cdot Y_{i,j}}{\sum_{k=1}^n \sum_{l=1}^m (q \cdot X_{k,l} + (1-q) \cdot Y_{k,l})} \right] \\ &= \frac{q \cdot \sum_{i=1}^n \sum_{j=1}^m X_{i,j}}{q \cdot \sum_{i=1}^n \sum_{j=1}^m X_{i,j} + (1-q) \cdot \sum_{i=1}^n \sum_{j=1}^m Y_{i,j}} \\ &= \frac{q \cdot \sum_{i=1}^n \sum_{j=1}^m X_{i,j}}{q \cdot \sum_{i=1}^n \sum_{j=1}^m X_{i,j} + (1-q) \cdot \sum_{i=1}^n \sum_{j=1}^m (V_g)_i \cdot (M_g \times K_g)_j} \\ &= \frac{q \cdot \sum_{i=1}^n \sum_{j=1}^m X_{i,j}}{q \cdot \sum_{i=1}^n \sum_{j=1}^m X_{i,j} + (1-q) \cdot \left( \sum_{i=1}^n (V_g)_i \cdot \sum_{j=1}^m (M_g \times K_g)_j \right)}. \end{aligned} \quad (8)$$

Vergleicht man (8) mit den Bezeichnungen der Okajima-Formel, so ergibt sich unmittelbar, daß

$$\sum_{i=1}^n \sum_{j=1}^m X_{i,j} = S(V_g \times M_g \times K_g) \quad (9)$$

der Summe der Wahrscheinlichkeiten aller genotypisch möglichen Eventualvater-Mutter-Kind-Terzette, also gleich dem phänotypischen  $X_p$  ist.

Weiter ist

$$\sum_{i=1}^n (V_g)_i \quad (10)$$

offensichtlich gleich der Wahrscheinlichkeit  $V_p$  des Phänotyps des Eventualvaters und

$$\sum_{j=1}^m (M_g \times K_g)_j = S(M_g \times K_g), \quad (11)$$

und folglich ist

$$\left( \sum_{i=1}^n (V_g)_i \cdot \sum_{j=1}^m (M_g \times K_g)_j \right) = Y_p. \quad (12)$$

Setzt man dieses oben ein, so erhält man

$$E(W_g) = \sum_{i=1}^n \sum_{j=1}^m W_{i,j} \cdot P_{i,j} = \frac{q \cdot X_p}{q \cdot X_p + (1-q) \cdot Y_p} = W_p. \quad (13)$$

Der Erwartungswert der genotypischen  $W_g$ -Werte ist also gleich dem phänotypischen Wert  $W_p$ .

Damit ergibt sich für die Varianz  $s^2$  der  $W_g$  unmittelbar

$$s^2 = \sum_{i=1}^n \sum_{j=1}^m (W_{i,j} - W_p)^2 \cdot P_{i,j}. \quad (14)$$

Betrachtet man abschließend nun die Verteilungsfunktion  $F$  der  $W_g$ , so gilt auf Grund eines allgemeinen Satzes aus der Wahrscheinlichkeitstheorie (s. z. B. Vogel [5], S. 53)

$$E(W_g) = \int_{-\infty}^{+\infty} x dF(x) = \int_0^{+\infty} (1 - F(x)) dx - \int_{-\infty}^0 F(x) dx. \quad (15)$$

Da in diesem speziellen Fall die Verteilungsfunktion  $F$  auf dem Intervall  $(-\infty, 0)$  konstant gleich Null ist und auf dem Intervall  $(+1, +\infty)$  konstant gleich Eins ist, vereinfacht sich (15) zu

$$E(W_g) = \int_0^1 (1 - F(x)) dx = 1 - \int_0^1 F(x) dx. \quad (16)$$

Wie schon gezeigt wurde, ist aber  $E(W_g) = W_p$  und folglich

$$1 - E(W_g) = 1 - W_p = \int_0^1 F(x) dx. \quad (17)$$

Da  $(1 - W_p)$  gleich der Irrtumswahrscheinlichkeit bei Entscheidung für Vaterschaft ist, ist also die Fläche unter der Verteilungsfunktion  $F$  der  $W_g$  ein Maß für diese Irrtumswahrscheinlichkeit. Da es sich bei den  $W_g$  um diskret verteilte

zufällige Größen handelt, ist die zugehörige Verteilungsfunktion  $F$  eine Treppenfunktion, und das Integral läßt sich elementar auswerten

$$1 - W_p = \int_0^1 F(x) dx = \sum_{i=1}^n \sum_{j=1}^m (1 - W_{i,j}) \cdot P_{i,j}. \quad (18)$$

Die phänotypische Irrtumswahrscheinlichkeit ist also wie nicht anders zu erwarten gleich der Summe der genotypischen Irrtumswahrscheinlichkeiten, multipliziert mit der Wahrscheinlichkeit  $P_{i,j}(W_{i,j})$ .

### Diskussion

Wie die theoretischen Betrachtungen gezeigt haben, ist das Ergebnis des phänotypischen Falls  $W_p$  gleich dem Erwartungswert der  $W_g$ , die man bei Berechnung der genotypisch möglichen Konstellationen erhält. Weiterhin gewinnt man durch diese Betrachtungen mit Hilfe der Varianz oder Standardabweichung der  $W_g$  ein Maß für die Streuung der  $W_g$  um ihren Erwartungswert  $W_p$ . Außerdem veranschaulicht die graphische Darstellung der Verteilungskurve  $F$  der  $W_g$  durch die Fläche unter dieser Kurve die Größe der Irrtumswahrscheinlichkeit bei positiver Entscheidung des Falles.

Durch diese Art der Betrachtung, deren wichtigstes Ergebnis weiterhin der phänotypische  $W$ -Wert, wie er in der bisherigen Begutachtung angegeben wurde, ist, wird der Charakter dieses  $W$ -Wertes als Erwartungswert (Mittelwert) der genotypischen  $W$ -Werte hervorgehoben. Es handelt sich also nicht um ein von der Methode nach Essen-Möller abweichendes Verfahren, sondern ist eine Erweiterung dieses Verfahrens, die als zusätzliche Information die Varianz oder Streuung der genotypischen  $W$ -Werte liefert und damit den bisher allein angegebenen phänotypischen  $W$ -Wert relativiert.

Ist durch erweiterte Familienuntersuchung die genotypische Konstellation eindeutig bestimmt, so muß selbstverständlich, wie bei Hummel [6] beschrieben, der alleingültige genotypische  $W$ -Wert als Vaterschaftsplausibilität angegeben werden, dessen Streuung — abgesehen von der Streuung auf Grund der empirisch bestimmten Genfrequenzen — dann gleich Null ist.

Es ist offensichtlich, daß es für den einzelnen Gutachter unmöglich ist, die Berechnungen, die für die genotypische Interpretation notwendig sind, von Hand durchzuführen, da bei Verwendung mehrerer Blutgruppensysteme in manchen theoretischen Fällen in einer phänotypischen Kombination mehr als  $5 \times 10^{11}$  genotypische Konstellationen enthalten sind. Aus diesem Grund wurde ein Programm erstellt, das die genotypische Betrachtung durchführt, den genotypischen Erwartungswert der  $W_g$  (also auch den phänotypischen Wert  $W_p$ ) und die Standardabweichung der  $W_g$  berechnet und zur Veranschaulichung eine Graphik der Verteilungsfunktion  $F$  der  $W_g$  liefert.

Eine erste Version dieses Programms wurde schon in einem Vortrag vorgeführt [7], und eine ausführliche Beschreibung der letzten Version und ihrer Möglichkeiten ist Gegenstand weiterer Arbeiten [8, 9].

Dieses Programm wurde in FORTRAN IV geschrieben und benötigte für bei uns bisher erstellte Gutachten in keinem Fall mehr als 2 min Rechenzeit auf der Rechenanlage der Universität Bonn vom Typ IBM 370—168.

Als Dateneingabe wird für jedes bearbeitete Blutgruppensystem eine Lochkarte benötigt, so daß die Kosten dieser Auswertungsmethode nicht höher liegen dürften als die der bisherigen biostatistischen Auswertung.

### Literatur

1. Rittner, Ch., Wehner, H. D.: Forensische Serologie. In: B. Mueller, Gerichtliche Medizin, 2. Aufl. Berlin-Heidelberg-New York: Springer 1974
2. Hummel, K.: Die medizinische Vaterschaftsbegutachtung mit biostatistischem Beweis. Stuttgart: G. Fischer 1961
3. Okajima, M.: Probability of paternity in Rh blood groups. Acta Genet. med. (Roma) **7**, 321 (1958)
4. Essen-Möller, E.: Die Beweiskraft der Ähnlichkeit im Vaterschaftsnachweis; theoretische Grundlagen. Mitt. Anthropol. Ges. (Wien) **68**, 9 (1938)
5. Vogel, W.: Wahrscheinlichkeitstheorie. Göttingen: Vandenhoeck & Ruprecht 1970
6. Hummel, K.: Biostatistische Abstammungsbegutachtung, Tabellenband I. Stuttgart: G. Fischer 1971
7. Baur, M. P.: Computer-Programm zur Berechnung von Vaterschaftswahrscheinlichkeiten. Vortrag auf der 5. Internationalen Tagung der Gesellschaft für forensische Blutgruppenkunde, Amsterdam 1973
8. Baur, M. P., Rittner, Ch.: Programmsystem zur Abstammungsbegutachtung und Berechnung der Plausibilität für Vaterschaft mit Hilfe der Information serologischer Befunde. I. Programmbeschreibung. In Vorbereitung
9. Rittner, Ch., Baur, M. P.: Programmsystem zur Abstammungsbegutachtung und Berechnung der Plausibilität für Vaterschaft mit Hilfe der Information serologischer Befunde. II. Anwendungsbeispiele aus dem Bonner Gutachtenmaterial. In Vorbereitung

Dipl.-Math. Max P. Baur  
Institut für Medizinische Statistik,  
Dokumentation und Datenverarbeitung  
der Universität  
D-5300 Bonn, Venusberg  
Bundesrepublik Deutschland