Virchows Arch. path. Anat. 337, 579—582 (1964)

Rückschlüsse von Sektionskollektiven

Bemerkungen zu der Arbeit von H. Grosse:
„Über ‚Berksons Fallacy' und die Selektion durch den Tod"
Virchows Archiv 337, 573—578 (1964)

Von

O. Mittmann

(Eingegangen am 16. März 1964)

Die Arbeit von Grosse über „Berksons Fallacy und die Selektion durch den Tod" regt dazu an, ganz allgemein die Frage zu untersuchen, ob und wie die statistischen Daten eines durch Auslese entstandenen Kollektivs verwendet werden können, um auf die relative Häufigkeit einer Merkmalskombination oder auf die Art der Korrelation zwischen zwei Merkmalen im Ausgangskollektiv zurückzuschließen. Das Ausgangskollektiv möge z.B. aus den Lebenden einer Bevölkerung oder aus den im Laufe eines Jahres Gestorbenen gebildet werden, während das durch Auslese entstandene Kollektiv z.B. die Gesamtheit derer sein möge, die von einem Pathologischen Institut im Laufe eines Jahres seziert werden. In jedem Kollektiv sollen vier Gruppen betrachtet werden, nämlich eine Gruppe A von Personen, die sowohl ein Merkmal 1 wie auch ein Merkmal 2 aufweisen, eine Gruppe B von Personen, die zwar das Merkmal 1, aber nicht das Merkmal 2 besitzen, eine Gruppe C von Personen, die das Merkmal 1 nicht, aber das Merkmal 2 aufweisen, und eine Gruppe D von Personen, die weder das Merkmal 1 noch das Merkmal 2 aufweisen. Die relativen Häufigkeiten der vier Gruppen seien im Ausgangskollektiv h bzw. k bzw. j bzw. l, und in dem durch Auslese entstandenen Kollektiv mögen dieselben vier Gruppen mit den relativen Häufigkeiten h' bzw. k' bzw. j' bzw. l' vorkommen. Die entsprechenden Auslesefaktoren seien w bzw. x bzw. y bzw. z, so daß

$$\begin{aligned} h' &= hw, \\ k' &= kx, \\ j' &= jy, \\ l' &= lz, \end{aligned}$$

wobei $h+k+j+l=1$ und $h'+k'+j'+l'=1$ ist. Die vier relativen Häufigkeiten in dem durch Auslese entstandenen Kollektiv mögen als bekannt vorausgesetzt werden, wogegen jede der vier relativen Häufigkeiten im Ausgangskollektiv für sich allein unbekannt sein möge, desgleichen jeder der Auslesefaktoren w, x, y und z für sich allein. Darf über die relativen Häufigkeiten im Ausgangskollektiv und über die Auslesefaktoren überhaupt nichts vorausgesetzt werden, dann liegen mehr Unbekannte als Gleichungen vor, und die Unbekannten sind nicht bestimmbar. Macht man spezielle Voraussetzungen über Beziehungen zwischen Unbekannten, so kann eine Bestimmbarkeit ermöglicht werden. Zur Erreichung einer solchen Bestimmbarkeit sollen folgende Voraussetzungen gemacht werden: Die Summe $h+k=p_1$ sei bekannt, desgleichen die Summe $h+j=p_2$. Der Besitz von Merkmal 1 oder Merkmal 2 soll eine verstärkte Auslese gegenüber denen bewirken, die weder Merkmal 1 noch Merkmal 2 besitzen. Der Besitz von Merkmal 1 und Merkmal 2 soll mindestens eine ebenso starke Auslese bewirken, wie das bei der stärkeren der durch Merkmal 1 oder Merkmal 2 bewirkten Auslesen der Fall ist. Über Beziehungen zwischen Auslesefaktoren werden im folgenden verschiedene Annahmen gemacht.

Rückschluß auf die Kombinationshäufigkeit im Ausgangskollektiv

Ist Merkmal 1 eine sehr ernste und Merkmal 2 eine sehr leichte Krankheit, so darf angenommen werden, daß die Personen, die nur Merkmal 1 aufweisen, nahezu ebenso stark der Auslese unterworfen sind wie die Personen, die sowohl Merkmal 1 wie auch Merkmal 2 besitzen, d.h. hier könnte man angenähert die Annahme machen, daß $w = x$ ist. Im Falle $w = x$ ist offenbar

$$\frac{h'}{k'} = \frac{h}{k}. \tag{1}$$

Hiernach ergäbe sich die relative Häufigkeit der Kombinationsfälle im Ausgangskollektiv zu

$$h = \frac{h' p_1}{h' + k'}.$$

Sind die beiden Merkmale zwei Krankheiten von ungefähr gleicher Schwere, so könnte angenommen werden, daß die Auslese bei beiden Krankheiten etwa gleichstark wirkt, d.h. daß $x = y$. In diesem Falle wäre offenbar

$$\frac{k'}{j'} = \frac{k}{j}, \tag{2}$$

und die relative Häufigkeit der Kombinationsfälle wird hier

$$h = p_1 - (p_1 - p_2)\frac{k'}{k' - j'} = p_2 - (p_2 - p_1)\frac{j'}{j' - k'}.$$

Im allgemeinen Falle werden weder die Voraussetzungen der Formel (1) noch der Formel (2) gegeben sein. Vielmehr wird im allgemeinen w größer als x wie auch y sein, x und y werden voneinander verschieden sein, und z wird kleiner als y wie auch x sein. Um die Auslesevorgänge im allgemeinen Fall in Formeln zu fassen, könnte man sich als geeignetes Modell eine Urne vorstellen, in der große und kleine Kugeln vorhanden sind, wobei sich unter den großen wie auch unter den kleinen Kugeln rauhe (gut griffige) und glatte (schlecht griffige) befinden. Dann stellt die Kugelgröße ein Merkmal 1 und die Rauheit der Oberfläche ein Merkmal 2 dar. Die Kugeln mit Merkmal 1 und Merkmal 2 haben die relative Häufigkeit h, die Kugeln mit Merkmal 1 und ohne Merkmal 2 haben die relative Häufigkeit k, die Kugeln mit Merkmal 2 und ohne Merkmal 1 haben die relative Häufigkeit j, und die Kugeln ohne Merkmal 1 und ohne Merkmal 2 haben die relative Häufigkeit l. Im Hinblick auf Ziehungen von Kugeln aus der Urne nehmen wir noch an, die Ziehbarkeit einer Kugel möge bei Vorhandensein von Merkmal 1 das t_1-fache der Ziehbarkeit einer Kugel ohne Merkmal 1 betragen, und die Ziehbarkeit einer Kugel mit dem Merkmal 2 möge das t_2-fache der Ziehbarkeit einer Kugel ohne Merkmal 2 betragen. Erfolgen nun Kugelziehungen aus der Urne, wobei die Kugeln „zurückgelegt“ oder bei großem

Urneninhalt auch „nicht zurückgelegt" werden können, so bilden die Kugelziehungen ein Kollektiv, für das

$$h' = \frac{h}{s} t_1 t_2 ,$$
$$k' = \frac{k}{s} t_1 ,$$
$$j' = \frac{j}{s} t_2 ,$$
$$l' = \frac{l}{s} 1 ,$$

wobei $s = h\ t_1 t_2 + k\ t_1 + j\ t_2 + l$ ist. Hieraus folgt $\frac{h'}{k'} = \frac{h}{k} t_2$, ferner $\frac{h'}{j'} = \frac{h}{j} t_1$ und $\frac{h'}{l'} = \frac{h}{l} t_1 t_2$, also $\frac{h'}{l'} \frac{l}{h} = \frac{h'}{j'} \frac{j}{h} \frac{h'}{k'} \frac{k}{h}$, woraus sich nach Kürzung mehrerer Faktoren die Formel

$$\frac{h' l'}{j' k'} = \frac{h l}{j k} \tag{3}$$

ergibt. Der nach Formel (3) berechenbare h-Wert bestimmt sich aus der quadratischen Gleichung

$$h^2 - \frac{1 + (p_1 + p_2)(c - 1)}{c - 1} h + \frac{p_1 p_2 c}{c - 1} = 0 \text{ mit } c = \frac{h' l'}{j' k'} .$$

Rückschluß auf die Art der Korrelation im Ausgangskollektiv

Wenn im Ausgangskollektiv Unabhängigkeit zwischen den Merkmalen 1 und 2 besteht, ist $hl - jk = 0$, d.h. es ist $hl = jk$ und $\frac{hl}{jk} = 1$. Bei positiver Korrelation zwischen den beiden Merkmalen ist $hl - jk > 1$, d.h. es ist $hl > jk$ und $\frac{hl}{jk} > 1$. Wenn in dem durch Auslese entstandenen Kollektiv Unabhängigkeit zwischen den Merkmalen 1 und 2 besteht, ist entsprechend $h' l' - j' k' = 0$, d.h. es ist $h' l' = j' k'$ und $\frac{h' l'}{j' k'} = 1$. Bei positiver Korrelation zwischen den Merkmalen ist $h' l' - j' k' > 1$, d.h. es ist $h' l' > j' k'$ und $\frac{h' l'}{j' k'} > 1$. Die Gleichung $\frac{h' l'}{j' k'} = \frac{h l}{l k}$ würde demnach besagen, daß zwischen den beiden Merkmalen im Ausgangskollektiv Unabhängigkeit besteht, falls in dem durch Auslese entstandenen Kollektiv Unabhängigkeit besteht (und umgekehrt), während zwischen den beiden Merkmalen im Ausgangskollektiv irgendeine positive Korrelation besteht, falls in dem durch Auslese entstandenen Kollektiv eine positive Korrelation zu beobachten ist (und umgekehrt).

Unter der Voraussetzung $w = x$, unter der die Formel (1) gilt, ist $\frac{h' l'}{j' k'} = \frac{hx\, lz}{jy\, kx} = \frac{h}{k} \frac{lz}{jy}$. Dies ist kleiner als $\frac{h l}{j k}$, wenn $z < y$ ist. Auf Grund der Voraussetzung $w = x$ besteht somit im Ausgangskollektiv eine positive Korrelation zwischen den beiden Merkmalen, wenn in dem durch Auslese entstandenen Kollektiv die Korrelation Null vorhanden ist, und im Ausgangskollektiv kann eine positive Korrelation oder die Korrelation Null bestehen, wenn in dem durch Auslese entstandenen Kollektiv eine negative Korrelation zu beobachten ist.

Würde man neben der Voraussetzung $w=x$ noch die Voraussetzung $y=z$ machen, dann wäre $\frac{h'\,l'}{j'\,k'}=\frac{h\,l}{j\,k}$, d.h. es bestünde in beiden Kollektiven dieselbe Art von Korrelation.

Unter der Voraussetzung $x=y$, unter der die Formel (2) gilt, ist $\frac{h'\,l'}{j'\,k'}=\frac{hw\;lz}{jy\;ky}$. Dies kann größer oder kleiner oder gleich $\frac{h\,l}{j\,k}$ sein, je nachdem $\frac{w\,z}{y^2}$ größer oder kleiner oder gleich 1 ist. Ein Rückschluß auf die Korrelation zwischen den beiden Merkmalen im Ausgangskollektiv ist also nicht möglich, wenn man nur die Voraussetzung $x=y$ macht.

Unter Voraussetzung des Urnenmodells, das zur Herleitung von Formel (3) diente, ist es außerdordentlich leicht zu sagen, wie es mit einem Rückschluß auf die Art der Korrelation im Ausgangskollektiv bestellt ist. Denn nach Formel (3) ist $\frac{h'\,l'}{j'\,k'}=\frac{h\,l}{j\,k}$, d.h. es ist hier ein umkehrbar eindeutiger Rückschluß bezüglich der Art der Korrelation durchführbar.

Praktisch weiß man gewöhnlich nicht, ob die für die Auslese gemachten Voraussetzungen auch tatsächlich in der lebendigen Wirklichkeit gegeben sind. Um die sich daraus ergebende Fragwürdigkeit zu klären, kann man die Methode der Material-Variation heranziehen. Man berechnet die Korrelation zwischen den beiden Merkmalen nicht nur an Hand eines einzigen Materials, sondern wiederholt die Berechnung an Hand eines zweiten Materials oder mehrerer anderer Materialien, die alle ganz verschiedenartige Beschaffenheiten aufweisen. Wenn sich an Hand jedes der verschiedenen Materialien eine Vereinbarkeit mit der Korrelation Null herausstellt, so darf die gefundene Unabhängigkeit als ein Ergebnis angesehen werden, das offenbar nicht durch die Auslese zustande gekommen ist. Wenn sich andererseits in jedem der verschiedenen Materialien irgendeine positive Korrelation finden läßt, dann darf das Bestehen einer positiven Korrelation als ein Ergebnis angesehen werden, das nicht durch die Auslese hervorgerufen worden ist.

Leider verfügt nicht jeder Forscher über ein Super-Material, das sich aus mehreren verschieden gearteten Materialien zusammensetzt. Es ist daher begrüßenswert, daß GROSSE sich darum bemüht, in der Frage der Auslesevorgänge voranzukommen.

Dr. O. MITTMANN, Universitäts-Hautklinik, 53 Bonn-Venusberg