Sozial- und Präventivmedizin Médecine sociale et préventive 29, 251 – 252 (1984)

Die Trennschärfe von Tests, die auf dem standardisierten Sterblichkeits-Quotienten beruhen

Ch. E. Minder

Institut für Sozial- und Präventivmedizin, Universität Bern, Finkenhubelweg 11, CH-3012 Bern.

1. Einführung

Der standardisierte Sterblichkeits-Quotient (SSQ; in der angelsächsischen Literatur SMR, standardised mortality ratio) wird im allgemeinen verwendet, um die Mortalitäts-Erfahrung einer Risiko-Bevölkerung mit derjenigen einer Standard-Bevölkerung zu vergleichen. In diesem Beitrag soll gezeigt werden, dass die Möglichkeit besteht, vor Beginn einer Studie die Chance abzuschätzen, einen bedeutsamen Unterschied zwischen den zwei Bevölkerungen zu entdecken. In epidemiologischer Terminologie untersuchen wir die Sensitivität*, in statistischer die Trennschärfe des verwendeten Tests. Das Verfahren wird anhand eines Beispieles illustriert.

2. Annahmen und Methode

Wir nehmen an, dass sowohl die Risiko-Bevölkerung wie die Standard-Bevölkerung in gleiche Altersgruppen eingeteilt sind. Die Standard-Bevölkerung soll sehr gross sein, mit demzufolge stabiler Mortalitäts-Erfahrung; die Sterbewahrscheinlichkeit, berechnet für die vorgesehene Beobachtungsperiode, in Altersgruppe i der Standard-Bevölkerung sei P_i.

Für die Risiko-Bevölkerung sollen für jede Altersklasse i r_i die beobachtete Anzahl Todesfälle, n_i die Grösse der Altersklasse und p_i die (unbekannte) Sterbewahrscheinlichkeit in der Beobachtungsperiode bedeuten.

Damit kann der SSQ definiert werden als das 100-fache des Verhältnisses

(1) $$M: = \frac{\Sigma r_i}{\Sigma n_i P_i} = \frac{B}{E}.$$

(Armitage, S. 388). Hier erfolgt die Summation über alle Altersklassen; $B = \Sigma r_i$ ist die totale Anzahl der beobachteten Todesfälle, und $E = \Sigma n_i P_i$ ist die aufgrund der Mortalitäts-Erfahrung der Standard-Bevölkerung zu erwartende Anzahl Todesfälle.

In der Folge wird ausserdem die Annahme gemacht, dass die Sterbewahrscheinlichkeiten P_i und p_i allesamt klein sind.

Im allgemeinen möchte man nun testen, ob sich die Mortalitäts-Erfahrung der untersuchten Bevölkerung von derjenigen der Standard-Bevölkerung unterscheidet. In mathematisch-statistischer Terminologie möchte man die

(2) Null-Hypothese gleicher Sterbewahrscheinlichkeit, H_o: $p_i = P_i$ für alle Schichten i, gegen die Alternativ-Hypothese ungleicher Sterbewahrscheinlichkeit, H_A: $p_i \neq P_i$ für mindestens eine Schicht i testen.

Aus (2) und den Verteilungs-Eigenschaften von B ergibt sich ein Test wie folgt:

(3) H_o wird verworfen, wenn $\frac{(B - E)^2}{E} > \chi^2_{1,\alpha}$ ist, sonst wird H_o akzeptiert.

Das Symbol $\chi^2_{1,\alpha}$ bezeichnet den α-Punkt der χ^2_1-Verteilung (siehe Tabelle 1).

Tabelle 1: Prozent-Punkte der χ^2_1 - und Standard-Gauss-Verteilung

α	0,05	0,01	0,05	0,001
$\chi^2_{1,\alpha}$	3,84	6,63	7,88	10,83
z_α	1,96	2,57	2,81	3,29

3. Die Trennschärfe des SSQ-Tests

Die Trennschärfe (power) eines Tests ist definiert als die Wahrscheinlichkeit, die Nullhypothese zu verwerfen, wenn die Alternativ-Hypothese zutrifft;

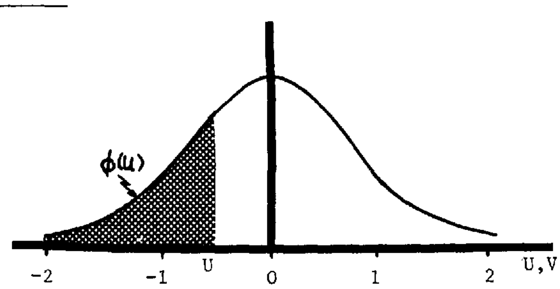
(4) $$\text{Trennschärfe} = W\,(H_o \text{ wird verworfen } / \ H_A \text{ ist wahr}) = W\left\{\frac{(B - E)^2}{E} > \chi^2_{1,\alpha} \ / \ H_A\right\}.$$

Der statistische Begriff der Trennschärfe entspricht dem epidemiologischen Begriff der Sensitivität eines diagnostischen Tests. Offensichtlich hängt die Trennschärfe von α, sowie von der exakten Alternative, die zutrifft, ab. Eine einfache Rechnung zeigt, dass in unserem Fall:

(5) $$\text{Trennschärfe} \doteq 1 + \phi(U) - \phi(V).$$

Dabei bedeuten $\phi(U)$ und $\phi(V)$ die Fläche unter der Standard-Gauss-Kurve rechts und U bzw. V. (Figur 1); die Funktion ϕ ist tabelliert (z.B. Armitage, S.458-9). Die Grössen U und V werden folgendermassen bestimmt:

Figur 1: Gauss-Kurve und Fläche ϕ



(6)
6.1 $E = \Sigma n_i P_i$
6.2 $\lambda = \Sigma n_i (p_i - P_i)$
6.3 z_α festlegen.
6.4 $$U = \frac{-E(\lambda + z_\alpha E)}{E + \lambda}.$$
6.5 $$V = \frac{E(-\lambda + z_\alpha E)}{E + \lambda}$$

* der Begriff der Sensitivität wird hier, zur Illustration des statistischen Konzeptes der Trennschärfe, ausserhalb seiner eigentlichen Domäne, den diagnostischen Tests, verwendet.

4. Ein Beispiel

Es ist beabsichtigt, in gewissen Abteilungen einer Firma mit ausschliesslich weiblichen Beschäftigten eine einjährige Studie über das Krebs-Risiko durchzuführen. Was ist die Chance, eine Verdoppelung bzw. eine Verdreifachung der Krebssterblichkeit gegenüber der gesamt-schweizerischen weiblichen Bevölkerung gleichen Alters zu entdecken, wenn ein Test mit SSQ auf dem 5%-Niveau ($\alpha = 0,05$) gebraucht werden soll?

Die vorhandenen Daten sind in Tabelle 2 zusammengestellt.

Tabelle 2: Mortalität von Schweizer Frauen an Krebs (Codes A 45 bis A 60), Sterblichkeitsdaten BfS 1980, ausgewählte Altersklassen (persönliche Mitteilung, F. Paccaud).

Altersklasse	P_i	Beschäftigte n_i (fiktive Zahlen)
45-49	0,00111	203
50-54	0,00173	351
55-59	0,00256	339
60-64	0,00374	285

(a) Verdoppelung

6.1: $E = \Sigma n_i P_i = 203 \times 0,00111 + 351 \times 0,00173 + 339 \times 0,00256 + 285 \times 0,00374 = 2,7663$

6.2: $\lambda = \Sigma n_i (p_i - P_i) = \Sigma n_i (2P_i - P_i) = \Sigma n_i P_i = 2,7663$

(Verdoppelung in der Risiko-Bevölkerung: $p_i = 2P_i$).

6.3: $z_\alpha = 1,96$, da $\alpha = 0,05$ (siehe Tabelle 1).

6.4: $U = \frac{-2,7663}{\sqrt{5,5326}} (2,7663 + 1,96 \sqrt{2,7663}) = -7,087$

6.5: $V = \frac{2,7663}{\sqrt{5,5326}} (-2,7663 + 1,96 \sqrt{2,7663}) = 0,5809$

Aus der Tabelle der Gauss-Verteilung erhält man:
$\phi(-7,087) = 0,000$; $\phi(0,5805) = 0,719$.

Daraus ergibt sich die
$T = 1 + 0,000 - 0,719 = 0,281$.

Das heisst: bei einer Verdoppelung des Krebsrisikos in den untersuchten Abteilungen gegenüber dem schweizerischen Mittel besteht eine 28-prozentige Chance, diesen Sachverhalt durch die Studie zu entdecken.

(b) Verdreifachung

Die Rechnung verläuft analog. Wie vorher hat man

$E = 2,7663$

$\lambda = \Sigma n_i (p_i - P_i) = \Sigma n_i (3 P_i - P_i) = 2 \Sigma n_i P_i = 2 E = 5,5326$

$z_\alpha = 1,96$

$U = \frac{-2,7663}{\sqrt{8,2989}} (5,5326 + 1,96 \sqrt{2,7663}) = -8,44$

$V = \frac{2.7663}{\sqrt{8,2989}} (-5,5326 + 1,96 \sqrt{2,7663}) = -2,18$

$\phi(U) = 0,000$; $\phi(V) = 0,0146$

$T = 1 + 0,000 - 0,0146 = 0,985$

Das heisst: es besteht eine 98,5-prozentige Chance, eine Verdreifachung der Krebssterblichkeit zu entdecken.

Es ist klar, dass eine Vergrösserung der Risiko-Bevölkerung (bei gleichbleibendem Risiko), zum Beispiel durch Einschluss von weiteren Firmen derselben Branche, die Trennschärfe des Tests erhöht. In diesem Sinne lassen sich verschiedene Varianten durchspielen, und es lässt sich vor Studienbeginn abschätzen, welche Erfolgsmöglichkeiten die Studie hat.

Nach Durchführung der Studie kann eine Rechnung wie die obige gebraucht werden, um bei nicht signifikantem Resultat eine obere Grenze für das erhöhte Risiko zu geben. In unserem Beispiel ist es bei einem nicht signifikanten Resultat unwahrscheinlich, dass das Risiko verdreifacht oder höher ist; dagegen ist eine Verdoppelung noch durchaus denkbar: die Zahlen sind zu klein um mit Sicherheit eine Verdoppelung zu entdecken.

5. Dank

Der Autor möchte Herrn Dr. F. Paccaud dafür danken, dass er seine Aufmerksamkeit auf das Problem gelenkt, und auch passende Daten bereitgestellt hat; Frau G. Thompson dafür, dass sie die Arbeit so sorgfältig geschrieben hat.

Literatur

Armitage, P.: Statistical Methods in Medical Research. London: Blackwell, 1980.

Zusammenfassung

In diesem Artikel wird eine Anleitung zur Berechnung der Trennschärfe eines auf dem standardisierten Sterblichkeits-Quotienten beruhenden Tests gegeben. Anhand eines Beispiels wird der Nutzen dieser Grösse bei der Planung und Analyse von epidemiologischen Studien illustriert. Vor Studienbeginn kann unter der Voraussetzung eines Effektes bekannter Grösse die Erfolgs-Chance einer Studie abgeschätzt werden, was eine problemgerechte Planung ermöglicht. Nach Studien-Ende kann, bei nicht signifikantem Resultat, eine obere Grenze für den gemessenen Effekt gegeben werden.

Résumé

Cet article présente une façon de calculer la puissance d'un test basé sur le taux de mortalité standardisé. A l'aide d'un exemple, on montre l'utilité de cette quantité pour la planification et l'analyse des études épidémiologiques. Avant l'exécution d'une étude, on peut déterminer la chance d'un succès , supposant l'existence d'un effet d'une certaine magnitude. Après la collection des données et quand le résultat n'est pas significatif , on peut donner une limite supérieure à l'effet postulé.

Summary

This paper presents a simple way of calculating the (approximate) power of a test based on the SMR. By means of an example, the use of this quantity, both for planning a study and for analysis is illustrated. Postulating an effect of a given magnitude on mortality, one can calculate the chances of success of a planned study. After data collection, one can, in the case of a nonsignificant result, at least give an upper limit to the effect: larger effects would likely have produced a significant result.