

Ю. Г. ИВАННИКОВ, В. А. КОНДРАТЬЕВ

ОЦЕНКА ЭФФЕКТИВНОСТИ МАССОВЫХ
ПРОТИВОЭПИДЕМИЧЕСКИХ МЕРОПРИЯТИЙ

(Представлено академиком В. М. Глушковым 12 XI 1971)

Противоэпидемические мероприятия при целом ряде инфекционных заболеваний носят массовый характер, охватывают большие контингенты людей, иногда население всей страны, и оценка их эффективности — важная и актуальная задача. Эта оценка строится на основании отношения вероятности p_1 заболеть в контрольной группе к вероятности p_2 заболеть в опытной группе: $I = p_1 / p_2$, называемом индексом эффективности. Поскольку точечная оценка не дает представления о возможном отклонении индекса эффективности от среднего значения, ищутся доверительные границы для него. В литературе даются правила построения доверительного интервала для I только в случае, когда распределение частот приближается к пуассоновскому. Результаты, изложенные в ⁽²⁾, также не решают задачи, так как они обеспечивают построение доверительного интервала с уровнем доверия $\alpha < \alpha'$ (где α' — избранный). Здесь рассматривается построение доверительных границ с уровнем доверия $\alpha = \alpha'$ в случае нормального и биномиального распределений, а также некоторые свойства индекса эффективности. В пп. 1–3 p_1 и p_2 понимаются как вероятности некоторых событий A_1 и A_2 ; в пп. 1, 2 поставленная задача решается исходя из испытаний по схеме Бернулли, причем в п. 2 используется нормальная аппроксимация; в п. 3 используется критерий χ^2 .

1. Пусть проведено n_1 испытаний для события A_1 и n_2 испытаний для события A_2 , причем все испытания независимы. Результат этих испытаний изобразится точкой (m_1, m_2) целочисленной решетки

$$D = \{(m_1, m_2): m_1 = 0, 1, \dots, n_1, m_2 = 0, 1, \dots, n_2\};$$

здесь m_1 и m_2 — число реализаций событий A_1 и A_2 соответственно. Пусть $\{D_i\}$, $i = 1, \dots, k$, — конечная последовательность областей решетки D , $D_i \subset D$, удовлетворяющая следующим условиям:

- 1°) $\{D_i\}$ сужающаяся, $D_{i+1} \subset D_i$, $i = 1, 2, \dots, k-1$;
- 2°) $D_1 = D$;
- 3°) D_2 не содержит точек (m_1, m_2) , у которых $m_1 = 0$;
- 4°) для любой области D_i из этой последовательности, если $(m_1^0, m_2^0) \in D_i$ и $m_1 \geq m_1^0$, а $m_2 \leq m_2^0$, то $(m_1^0, m_2) \in D_i$.

Определим функции $f_i(I)$ следующим образом:

$$f_i(I) = \sup_{p_1^0/p_1^0=1} P \{(m_1, m_2) \in D_i | p_1 = p_1^0, p_2 = p_2^0\},$$
$$i = 1, 2, \dots, k, \quad I > 0.$$

Лемма. Функция f_i непрерывные и неубывающие на промежутке $(0, +\infty)$, причем для $i \geq 2$ $\lim_{I \rightarrow 0} f_i(I) = 0$, $\lim_{I \rightarrow +\infty} f_i(I) = 1$. При фиксированном I последовательность $f_i(I)$, $i = 1, 2, \dots, k$, невозрастающая.

Зафиксируем α из промежутка $(0, 1)$. Тогда из леммы следует, что при $i \geq 2$ уравнение

$$f_i(I) = \alpha \tag{1}$$

разрешимо.

Пусть для $i \geq 2$ I_i — наибольшее из решений уравнения (1). I_1 положим равным нулю. Последовательность I_i , $i = 1, 2, \dots, k$, неубывающая.

На самом деле, применяя определение I_i и лемму, имеем при $i_1 < i_2$

$$f_{i_2}(I_{i_2}) = \alpha = f_{i_1}(I_{i_1}) \geq f_{i_2}(I_{i_1})$$

1. Следовательно, $I_{i_2} \geq I_{i_1}$.

Пусть $(m_1, m_2) \in D_i \setminus D_{i+1}$ (при $i = k$ положим $D_{k+1} = \emptyset$). Положим $f_i(m_1, m_2) = I_i$. Тем самым на решетке D определена функция I_n (однозначно, поскольку последовательность D_i сужается).

Теорема 1. Если m_1 и m_2 — число реализаций событий A_1 и A_2 в n_1 и n_2 испытаниях соответственно, то

$$\sup_{p_1^0, p_2^0} P \left\{ I_n(m_1, m_2) > \frac{p_1}{p_2} \mid p_1 = p_1^0, p_2 = p_2^0 \right\} = \alpha,$$

где $I_n(m_1, m_2)$ — нижняя доверительная граница для отношения $I = p_1/p_2$ с уровнем доверия α . При этом I_n не убывает по m_1 и не возрастает по m_2 .

Таким образом, для нахождения нижней доверительной границы для I следует задать последовательность областей D_i с указанными свойствами.

Этот процесс обратим в следующем смысле. Пусть I_n' — нижняя доверительная граница для I с уровнем доверия, не превосходящим α (I_n' — функция результатов испытаний $(m_1, m_2) \in D$). Пусть I_n' не убывает по m_1 и не возрастает по m_2 . Перенумеруем все значения I_n' на D в порядке возрастания: $I_1' < I_2' < \dots < I_k'$; здесь k — количество различных значений функции I_n' , I_i' — ее i -е по величине значение.

Пусть $D_i = \{(m_1, m_2) \in D: I_n'(m_1, m_2) \geq I_i'\}$, $i = 1, 2, \dots, k$. Тогда D_i удовлетворяют условиям 1°) — 4°). Построим по областям D_i указанным выше способом нижнюю доверительную границу I_n . Тогда справедлива

Теорема 2. 1°) $I_n'(m_1, m_2) \leq I_n(m_1, m_2)$ при всех $(m_1, m_2) \in D$.

2°) Если уровень доверия α достигается при некоторых значениях p_1 и p_2 для границы I_n' , то он при тех же p_1 и p_2 достигается и для границы I_n .

Будем говорить, что уровень доверия α для нижней границы I_n достигается в обобщенном смысле в точке $I_0 \in (0, +\infty)$, если

$$\lim_{I \rightarrow I_0} \sup_{p_1^0/p_2^0 = I} P \{ I_n(m_1, m_2) > I \mid p_1 = p_1^0, p_2 = p_2^0 \} = \alpha.$$

Ясно, что если уровень доверия достигается, то он достигается и в обобщенном смысле.

Теорема 3. 1°) Если уровень доверия α достигается в обобщенном смысле для границы I_n' , в точке I_0 , то в этой точке он достигается в том же смысле и для границы I_n .

2°) Если I_0 — какое-либо значение функции I_n , то уровень доверия α для границы I_n достигается в обобщенном смысле в точке I_0 .

Из всего изложенного ясно, что задача построения нижней доверительной границы не решается однозначно. Именно, для приведенного алгоритма любой выбор последовательности областей D_i со свойствами 1°) — 4°) определяет некоторую нижнюю границу, причем для различных таких последовательностей полученные границы различны.

Теоремы 2 и 3 показывают, что построенные таким образом границы в определенном смысле оптимальнее любых других возможных границ. Кроме того, теорема 2 дает некоторую возможность разумным образом ликвидировать упомянутую неоднозначность. А именно, можно взять в качестве нижней границы I_n' значение, полученное с помощью какого-либо простого алгоритма, и, «поправив» его описанным выше способом, получить более оптимальную границу I_n . При этом первое утверждение теоремы 2 может помочь при установлении каких-либо полезных свойств границы I_n (типа сходимости при $n_1, n_2 \rightarrow \infty$), если таковыми обладает грани-

ца I_n' . Например, можно положить $I_n' = p_{1n} / p_{2n}$, где p_{1n} — нижняя граница для вероятности p_1 , p_{2n} — верхняя граница для вероятности p_2 (вычисляются обычным способом). I_n' — доверительная граница с уровнем доверия $\alpha \leq \alpha'$, $1/2\alpha$ — уровень доверия для доверительных границ p_{1n} , p_{2n} .

2. Пусть n_1 , n_2 , m_1 , m_2 имеют тот же смысл, что и в п. 1. Обозначим через p_1^* и p_2^* частоты событий A_1 и A_2 :

$$p_1^* = m_1/n, \quad p_2^* = m_2/n,$$

а через I^* — наблюдаемый индекс эффективности

$$I^* = p_1^*/p_2^*.$$

Если n_1 и n_2 достаточно велики (а p_2 отделена от нуля, $p_2 \geq a$, a — известное положительное число), то, применив нормальную аппроксимацию биномиального закона, можно получить закон распределения I^* и, исходя из него, найти доверительные границы для I . Таким образом, получим

$$I_B^* = \left[I^* + \frac{t_\alpha^2}{2an_1} - t_\alpha \left(I^{*2} \left(\frac{1}{a} - 1 \right) \frac{1}{n_2} + I^* \left(\frac{1}{a} - I^* \right) \frac{1}{n_1} + \frac{t_\alpha^2}{4a^2 n_1^2} + \left(\frac{1}{a} - 1 \right) \frac{1}{n_1 n_2} \right)^{1/2} \right] \frac{1}{1 + t_\alpha^2/n_1}, \quad \text{если } I^* < I_{\max},$$

$$I_B = 1/a \quad (\text{и достоверна}), \quad \text{если } I^* > I_{\max},$$

$$I_{\max} = \frac{1}{a} \left(1 - t_\alpha \left(\left(\frac{1}{a} - 1 \right) \frac{1}{n_2} \right)^{1/2} \right)^{-1};$$

здесь α — уровень доверия, t_α — квантиль нормального распределения.

В пп. 1, 2 речь шла только об одной границе. Если известен способ нахождения нижней (верхней) границы, то верхнюю (нижнюю) границу можно найти через нижнюю (верхнюю) границу для обратного отношения $I' = 1/I = p_2/p_1$. При этом условие $p_2 \geq a$ в п. 2 заменится на условие $p_1 \geq a$. Для нахождения доверительного интервала по формулам п. 2 потребуется условие: $p_1 \geq a$, $p_2 \geq b$, a и b — известные положительные числа.

3. Пусть теперь проведено n испытаний, причем каждое испытание есть пара испытаний — одно для события A_1 и другое для события A_2 . В результате этих испытаний получим n_1 — число реализаций события $A_1 A_2$, n_2 — то же для $\bar{A}_1 A_2$, n_3 — для $A_1 \bar{A}_2$, n_4 — для $\bar{A}_1 \bar{A}_2$ (чертой сверху обозначено противоположное событие). Пусть π_1 , π_2 , π_3 , π_4 — вероятности перечисленных выше событий соответственно. Из соотношений $p_1 = \pi_2 + \pi_4$, $p_2 = \pi_3 + \pi_4$, $I = p_1/p_2$ получим

$$\pi_1 = 1 - I p_2 - \pi_3, \quad \pi_2 = I p_2 - \pi_4, \quad \pi_3 = \pi_3, \quad \pi_4 = \pi_4. \quad (3)$$

Таким образом, при заданном I вероятности π_i являются функциями параметров π_3 и π_4 . Тогда, согласно (1), стр. 342, случайная величина

$$\chi^2 = \sum_{i=1}^4 \frac{(n_i - \hat{\pi}_i n)^2}{n \hat{\pi}_i}$$

распределена асимптотически как χ^2 с одной степенью свободы. (Здесь $\hat{\pi}_i$ — оценки для π_i , вычисленные по формулам (3) с использованием в качестве значений параметров их оценки $\pi_3^* = n_3/n$, $\pi_4^* = n_4/n$.) Таким образом, при заданном I можно проверить по критерию χ^2 зависимость (3). Выбирая те значения I , для которых гипотеза о наличии зависимости (3) не отвергается, получим доверительное множество для I . Оно оказывается интервалом. Его границы

$$I_{B, n} = \frac{1}{1+r} \left[I^* + \frac{1}{2} \frac{1 - \pi_3^* + \pi_4^*}{p_2^*} r \pm \left(\frac{1}{4} \frac{(1 - p_2^*)^2}{p_2^{*2}} r^2 + \frac{\pi_1^* \pi_2^*}{p_2^{*2}} r \right)^{1/2} \right];$$

здесь $I^* = p_1^*/p_2^*$, $p_1^* = \pi_2^* + \pi_4^*$, $p_2^* = \pi_3^* + \pi_4^*$, $\pi_i^* = n_i/n$, $r = t_\alpha^2 \times \times (n(1 - p_2^*))^{-1}$, t_α — квантиль нормального распределения. При большом n ,

пренебрегая малыми членами, получим

$$I_{в, н} = I^* \pm \frac{t_{\alpha}}{p_2^*} \left(\frac{\pi_1^* \pi_2^*}{n(1 - p_2^*)} \right)^{1/2}.$$

4. Если исследуется влияние k независимых факторов с эффективностями $I_1, \dots, I_k$ на вероятность событий A_1 и A_2 , то индекс эффективности суммарного влияния всех факторов выразится следующим образом:

$$I = \prod_{i=1}^k I_i. \quad (4)$$

В практике эпидемиологических исследований часто приходится решать вопрос о доверительной границе отношения

$$I = p_1 p_1' / (p_2 p_2'), \quad (5)$$

где p_1 и p_2 трактуются как вероятности заболеть лицу из контрольной и опытной группы соответственно, а p_1' и p_2' — как вероятности иметь для него заданный диагноз в группе сходных заболеваний.

Если здесь предположить, что $P\{p_1' = p_1^0 | p_1 = p_1^0\} = 1$, то нижняя доверительная граница индекса эффективности может быть вычислена по формуле $I_n = p_{1н} p_{1н}' / (p_{2н} p_{2н}')$.

Оценки для $p_{1н}'$ и $p_{2н}'$ даются через частоты подтверждения диагноза у заболевших с помощью лабораторных методов. Из (5) видно, что лабораторное обследование проводится в обеих группах — контрольной и опытной. Однако можно вычислить значение индекса эффективности, совпадающее с получаемым по формуле (5) на основании лабораторного обследования больных только в опытной группе:

$$I = (p_1/p_2 + p_2' - 1)/p_2'. \quad (6)$$

Выражение (6) верно тогда, когда фактор влияет только на изучаемое заболевание. Если фактор защиты влияет также и на смежные заболевания, то для индекса эффективности устанавливается следующая зависимость:

$$I = [p_1/p_2 - I^*(1 - p_2')]/p_2',$$

где I^* — индекс эффективности неспецифического влияния исследуемого фактора.

Для установления ценности исследованного фактора защиты могут быть использованы экономические критерии. В частности, можно вычислить наименьшее допустимое значение индекса эффективности, при котором экономический выигрыш от применения данной меры будет сбалансирован с затратами по ее осуществлению: $I = p_1 c_1 / (p_1 c_1 - c_2)$, где p_1 — заболеваемость людей, защищаемых данным мероприятием, c_1 — потери на одного заболевшего, c_2 — затраты на одного защищенного мероприятием в тех же единицах, что и c_1 .

Всесоюзный научно-исследовательский институт гриппа
Министерства здравоохранения СССР
Москва

Поступило
4 XI 1971

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

- ¹ С. Р. Рао, Линейные статистические методы и их применения, М., 1968.
- ² Ю. К. Беляев, ДАН, 169, № 4, 755 (1966).