ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ ОЦЕНОК ИНТЕНСИВНОСТИ РИСКОВАННОГО ПОВЕДЕНИЯ НА ОСНОВЕ НЕРАВЕНСТВА ЧЕБЫШЕВА

Суворова А.В., Пащенко А.Е., Тулупьева Т.В., Тулупьев А.Л.

УДК 311.2 + 616-036.22

Суворова А.В., Пащенко А.Е., Тулупьева Т.В., Тулупьев А.Л. **Построение доверительных интервалов оценок интенсивности рискованного поведения на основе неравенства Чебышева.**

Аннотация. В статье выводятся интервальные оценки интенсивности рискованного поведения и риска, с ним связанного. Предложен подход к проверке согласованности оценок указанных показателей, полученных различными способами по различным исходным данным.

Ключевые слова: интервальные оценки, интенсивность поведения, согласованность оценок, неравенство Чебышева.

Suvorova A.V., Paschenko A.E., Tulupyeva T.V., Tulupyev A.L, **Risky behavior rate estimates calculation based on Chebychev inequality.**

Abstract. The paper presents estimations of intensity of risky behavior and risk, with it connected are deduced. The approach to check of a coordination of estimations of the specified indicators received in the various ways under the various initial data is offered.

Keywords: Interval estimates, rate of behavior, estimates consistency, Chebyshev inequality.

1. Введение. В современной эпидемиологии остро стоит вопрос об оценке риска передачи и приобретения опасных неизлечимых инфекций (например, ВИЧ). Наиболее точно такой риск характеризуется инсиденс-показателем (число заразившихся за определенный период среди лиц, подвергавшихся риску заражения, отнесенное к человеко×месяцам наблюдения). Инсиденс-показатель можно оценить, зная индивидуальный риск заражения за заданный период времени каждого отдельного респондента. Модель Белла—Тревино [7] увязывает оценку риска с числом эпизодов рискованного поведения. Число же эпизодов можно оценить, если в свою очередь известна оценка интенсивности рискованного поведения, рассмотренного как случайный процесс определенного класса.

Точечные оценки интенсивности рискованного поведения и риска, с ним связанного, получены в работах [1–3]. Заметим, что указанные оценки строились 2 раза — по данным двух типов. Сначала оценки были получены по сведениям о нескольких последних эпизодах рискованного поведения, а затем — по сведениям о максимальном, минимальном и обычном интервалах между эпизодами рискованного поведения.

Целями данной работы являются, во-первых, построение интервальных оценок интенсивности рискованного поведения и риска, с ним связанного; а во-вторых — проверка согласованности оценок, полученных по данным разных типов.

2. Доверительные интервалы. Для построения интервальных оценок интенсивности рискованного поведения и риска, с ним связанного, построим для точечных оценок, полученных в работах [1–3], доверительные интервалы, т. е. интервалы, накрывающие истинное значение с заданной вероятностью. Таким образом, для каждой оценки $\hat{\theta}$ построим интервал вида

$$I = \left(\hat{\theta} - a, \hat{\theta} + b\right),$$

такой, что выполняется условие $P\left(\theta \in I\right) = \alpha$, где α — заранее заданная вероятность попадания в интервал I (т. е. I — доверительный интервал с уровнем доверия α).

При построении доверительных интервалов будем использовать неравенство Чебышева:

$$P\left(\left|X - \mathsf{E}\left(X\right)\right| > \varepsilon\right) \le \frac{\mathsf{D}\left(X\right)}{\varepsilon^2},$$

где X — некоторая случайная величина с дисперсией $\mathsf{D}\left(X\right)$ и математическим ожиданием $\mathsf{E}\left(X\right)$.

Используя заданное значение α, вычислим ε:

$$\varepsilon = \sqrt{\frac{\mathsf{D}\left(X\right)}{\alpha}}.$$

Таким образом,

$$P\left(\left|X - \mathsf{E}\left(X\right)\right| > \sqrt{\frac{\mathsf{D}\left(X\right)}{\alpha}}\right) \le \alpha.$$

Переходя к противоположному событию, получим неравенство

$$P\left(\left|X - \mathsf{E}\left(X\right)\right| \le \sqrt{\frac{\mathsf{D}\left(X\right)}{\alpha}}\right) \ge 1 - \alpha.$$

Тогда доверительный интервал с уровнем доверия $1-\alpha$ имеет вид

$$\left(\mathsf{E}\left(X\right) - \sqrt{\frac{\mathsf{D}\left(X\right)}{\alpha}},\ \mathsf{E}\left(X\right) + \sqrt{\frac{\mathsf{D}\left(X\right)}{\alpha}}\right).$$

Рассмотрим полученные в предыдущих статьях оценки.

1. Оценка интенсивности $\hat{\lambda} = \frac{\nu}{\tau}$, рассчитанная по данным о ν последних эпизодах рискованного поведения респондента. Рискованное поведение рассматривается как пуассоновский процесс, т. е. плотность распределения случайной величины λ имеет вид

$$g(\lambda) = \tau \frac{(\tau\lambda)^{\nu}}{\nu!} e^{-\tau\lambda}.$$

Проверим, что такая функция $g(\lambda)$ удовлетворяет условию $\int\limits_{-\infty}^{+\infty} g(t)\,dt = 1$:

$$\int\limits_{-\infty}^{+\infty} g(t)\,dt = \int\limits_{-\infty}^{+\infty} \tau \frac{(\tau t)^{\nu}}{\nu!} e^{-\tau t} dt = \int\limits_{0}^{+\infty} \tau \frac{(\tau t)^{\nu}}{\nu!} e^{-\tau t} dt = [y = \tau t] = \tau \frac{1}{\tau} \int\limits_{0}^{+\infty} \frac{y^{\nu}}{\nu!} e^{-y} dy =$$

$$= -e^{-y} \frac{y^{\nu}}{\nu!}\bigg|_{0}^{+\infty} + \int\limits_{0}^{+\infty} \frac{y^{\nu-1}}{(\nu-1)!} e^{-y} dy = \int\limits_{0}^{+\infty} e^{-y} dy = -e^{-y}\Big|_{0}^{+\infty} = 1.$$

Вычислим характеристики этой случайной величины:

$$\mathsf{E}(\lambda) = \int\limits_{0}^{\infty} t \frac{(\tau t)^{\nu}}{\nu!} e^{-\tau t} dt = \frac{\nu+1}{\tau} \int\limits_{0}^{\infty} \frac{(\tau t)^{\nu+1}}{(\nu+1)!} e^{-\tau t} dt = \frac{\nu+1}{\tau},$$

$$\mathsf{D}(\lambda) = \int\limits_{0}^{\infty} \left(t - \frac{\nu+1}{\tau}\right)^{2} \frac{(\tau t)^{\nu}}{\nu!} e^{-\tau t} dt = \int\limits_{0}^{\infty} \left(t^{2} - \frac{2t(\nu+1)}{\tau} + \left(\frac{\nu+1}{\tau}\right)^{2}\right) \frac{(\tau t)^{\nu}}{\nu!} e^{-\tau t} dt =$$

$$= \int\limits_{0}^{\infty} t^{2} \frac{(\tau t)^{\nu}}{\nu!} e^{-\tau t} dt - 2\frac{\nu+1}{\tau} \int\limits_{0}^{\infty} t \frac{(\tau t)^{\nu}}{\nu!} e^{-\tau t} dt + \left(\frac{\nu+1}{\tau}\right)^{2} \int\limits_{0}^{\infty} \frac{(\tau t)^{\nu}}{\nu!} e^{-\tau t} dt =$$

$$= \frac{(\nu+1)(\nu+2)}{\tau^{2}} - 2\frac{(\nu+1)^{2}}{\tau^{2}} + \frac{(\nu+1)^{2}}{\tau^{2}} = \frac{\nu+1}{\tau^{2}}(\nu+2-\nu-1) = \frac{\nu+1}{\tau^{2}}.$$

Доверительный интервал с уровнем доверия $1-\alpha$ имеет вид:

$$\left(\frac{\nu+1}{\tau} - \frac{1}{\tau}\sqrt{\frac{\nu+1}{\alpha}}, \frac{\nu+1}{\tau} + \frac{1}{\tau}\sqrt{\frac{\nu+1}{\alpha}}\right).$$

2. Оценка кумулятивного риска, связанного с рискованным поведением, полученная по данным о последних эпизодах рискованного поведения. Зависимость между кумулятивным риском и интенсивностью поведения выражается формулой [1, 7]

$$\Pr = 1 - e^{-\lambda pT},$$

где λ — интенсивность; p — риск заражения за один эпизод; T — период, ~~в~~ днях; $\Pr$ — кумулятивный риск.

Оценка интенсивности λ рассматривается как случайная величина, поэтому кумулятивный риск, как функцию от λ, тоже будем рассматривать как случайную величину. Вычислим характеристики этой случайной величины:

$$\mathsf{E}(\Pr) = \mathsf{E}\left(1 - e^{-\lambda pT}\right) = \int_0^\infty \left(1 - e^{-tpT}\right) g(t)\,dt = \int_0^\infty g(t)\,dt - \int_0^\infty g(t) e^{-tpT}\,dt =$$

$$= 1 - \int_0^\infty \tau \frac{(\tau t)^\nu}{\nu!} e^{-\tau t} e^{-tpT}\,dt = [y = \tau t] = 1 - \int_0^\infty \frac{y^\nu}{\nu!} e^{-y\left(1+\frac{pT}{\tau}\right)}\,dy = 1 - \frac{1}{1 + \frac{pT}{\tau}} = \frac{pT}{\tau + pT},$$

$$\mathsf{D}(\Pr) = \mathsf{D}\left(1 - e^{-\lambda pT}\right) = \int_0^\infty \left(1 - e^{-tpT}\right)^2 g(t)\,dt - \left(\mathsf{E}(\Pr)\right)^2 =$$

$$= \int_0^\infty g(t)\,dt - 2\int_0^\infty g(t) e^{-tpT}\,dt + \int_0^\infty g(t) e^{-2tpT}\,dt - \left(\mathsf{E}(\Pr)\right)^2 =$$

$$= 1 - 2\frac{\tau}{\tau + pT} + \frac{\tau}{\tau + 2pT} - \left(\frac{pT}{\tau + pT}\right)^2 = \frac{\tau (pT)^2}{(\tau + 2pT)(\tau + pT)^2}.$$

Доверительный интервал с уровнем доверия $1-\alpha$ имеет вид:

$$\left(\frac{pT}{\tau + pT} - \frac{pT}{\tau + pT}\sqrt{\frac{\tau}{\alpha(\tau + 2pT)}},\ \frac{pT}{\tau + pT} + \frac{pT}{\tau + pT}\sqrt{\frac{\tau}{\alpha(\tau + 2pT)}}\right).$$

3. Оценка n^* числа эпизодов, произошедших за заданный промежуток времени T. Оценка получена по данным об обычном интервале между эпизодами рискованного поведения и о «рекордных» интервалах — максимальном и минимальном. Отметим, что ответы могут содержать сведения как обо всем наборе указанных величин, так и о каком-то его подмножестве. При построении оценки каждому значению n_i из интервала $[0, 5T]$ была сопоставлена вероятность

$$p_i = \tilde{f}_i \Big/ \sum_{l=0}^{5T} \tilde{f}_l,$$

где $\tilde{f}_i = \tilde{f}_{\min,\mathrm{med},\max}\left(t_{\min}, t_{\mathrm{med}}, t_{\max}, n_i\right)$.

Характеристики этой случайной величины:

$$\mathsf{E}(n)=\sum_{i=0}^{5T} i p_i=\sum_{i=0}^{5T} i\left(\tilde{f}_i \Big/ \sum_{l=0}^{5T}\tilde{f}_l\right)= \sum_{i=0}^{5T} i\tilde{f}_i \Big/ \sum_{l=0}^{5T}\tilde{f}_l=a,$$

$$\mathsf{D}(n)=\sum_{i=0}^{5T} i^2 p_i-\mathsf{E}(n)^2=b.$$

Доверительный интервал с уровнем доверия $1-\alpha$ имеет вид:

$$\left(a-\sqrt{\frac{b}{\alpha}},a+\sqrt{\frac{b}{\alpha}}\right).$$

Пусть $T_{\max}$ — максимальный интервал между эпизодами рискованного поведения респондента, $T_{\min}$ — минимальный и T_{med} — обычный интервал. Как отмечалось, в результате интервью нам могут быть известны различные сочетания данных: только $T_{\max}$, только $T_{\min}$, только T_{med}, $T_{\max}$ и $T_{\min}$, $T_{\max}$ и T_{med}, T_{med} и $T_{\min}$, одновременно $T_{\max}$, T_{med}, $T_{\min}$. Для всех этих вариантов были выведены формулы плотности распределения соответствующих порядковых статистик и их сочетаний. Отметим, что формулы, использующие T_{med}, были получены только для случая нечетного *n*, а при вычислении математического ожидания и дисперсии используются все значения из интервала $[0,5T]$ последовательно, поэтому сейчас рассмотрим те варианты, в которых используются только значения $T_{\max}$ и $T_{\min}$.

Выпишем выражения для $\tilde{f}_i$ для каждого случая (T — исследуемый промежуток времени):

а) известно только значение $T_{\min}$, тогда

$$f_{\min}(t)=n\lambda e^{-n\lambda t},\quad -\infty<t<+\infty,\ \tilde{f}(n)=\frac{n^2}{T}e^{-\frac{n^2}{T}T_{\min}},\ \tilde{f}_i=\frac{i^2}{T}e^{-\frac{i^2}{T}T_{\min}};$$

б) известно только значение $T_{\max}$, тогда

$$f_{\max}(t)=n\lambda e^{-\lambda t}\left(1-e^{-\lambda t}\right)^{n-1},\quad -\infty<t<+\infty,\ \tilde{f}_i=\frac{i^2}{T}e^{-\frac{i}{T}T_{\max}}\left(1-e^{-\frac{i}{T}T_{\max}}\right)^{i-1};$$

в) известны значения $T_{\max}$ и $T_{\min}$, тогда,

$$f_{\min,\max}(x,y)=\frac{n!}{(n-2)!}\lambda^2 e^{-\lambda(x+y)}\left(e^{-\lambda x}-e^{-\lambda y}\right)^{n-2},$$

если $-\infty < x < y < +\infty$, и $f_{\min,\max}(x,y) = 0$ в остальных случаях;

$$\tilde{f}_i = \frac{i!}{(i-2)!}\frac{i^2}{T^2} e^{-\frac{i}{T}(T_{\min}+T_{\max})}\left(e^{-\frac{i}{T}T_{\min}} - e^{-\frac{i}{T}T_{\max}}\right)^{i-2}.$$

Ранее отмечалось, что все формулы для плотностей $\tilde{f}_i$, в которых используется значение T_{med}, справедливы только для нечетных значений i. Поэтому в этих случаях при вычислении математического ожидания и дисперсии суммирование будет проводиться только по нечетным значениям:

$$\mathsf{E}(n) = \sum_{i=0}^{\left[\frac{5T}{2}\right]} (2i+1)\, p_{2i+1} = \frac{\sum_{i=0}^{\left[\frac{5T}{2}\right]} (2i+1)\, \tilde{f}_{2i+1}}{\sum_{l=0}^{\left[\frac{5T}{2}\right]} \tilde{f}_{2l+1}} = a,$$

$$\mathsf{D}(n) = \sum_{i=0}^{\left[\frac{5T}{2}\right]} (2i+1)^2\, p_{2i+1} - \mathsf{E}(n)^2 = b.$$

Выражения для $\tilde{f}_i$ для оставшихся случаев:

г) известно только значение T_{med}:

$$f_{\text{med}}(t) = \frac{n!}{\left(\left(\frac{n-1}{2}\right)!\right)^2} \lambda e^{-\lambda t(n-1)/2}\left(1-e^{-\lambda t}\right)^{0.5(n-1)}, \quad -\infty < t < +\infty,$$

$$\tilde{f}_i = \frac{i!}{\left(\left(\frac{i-1}{2}\right)!\right)^2}\frac{i}{T} e^{-(i/T)T_{\text{med}}(i+1)/2}\left(1-e^{-(i/T)T_{\text{med}}}\right)^{(i-1)/2};$$

д) известны значения T_{med} и $T_{\min}$:

$$f_{\min,\text{med}}(x,y) = \frac{n!}{(0.5(n-3))!(0.5(n-1))!}\lambda^2 e^{-\lambda\left(x+\frac{n+1}{2}y\right)}\left(e^{-\lambda x}-e^{-\lambda y}\right)^{\frac{n-3}{2}},$$

если $-\infty < x < y < +\infty$, и $f_{\min,\text{med}}(x,y) = 0$ в остальных случаях,

$$\tilde{f}_i = \frac{i!}{(0.5(i-3))!(0.5(i-1))!}\left(\frac{i}{T}\right)^2 e^{-\frac{i}{T}\left(T_{\min}+\frac{i+1}{2}T_{\text{med}}\right)}\left(e^{-\frac{i}{T}T_{\min}} - e^{-\frac{i}{T}T_{\text{med}}}\right)^{\frac{i-3}{2}};;$$

е) известны значения $T_{\max}$ и T_{med}:

$$f_{\mathrm{med},\max}(x,y)=$$

$$=\frac{n!}{(0.5(n-1))!(0.5(n-3))!}\lambda^2 e^{-\lambda(x+y)}\left(1-e^{-\lambda x}\right)^{\frac{n-1}{2}}\left(e^{-\lambda x}-e^{-\lambda y}\right)^{\frac{n-3}{2}},$$

если $-\infty<x<y<+\infty$, и $f_{\mathrm{med},\max}(x,y)=0$ в остальных случаях,

$$\tilde{f}_i=\frac{i!}{(0.5(i-1))!(0.5(i-3))!}\left(\frac{i}{T}\right)^2 e^{-\frac{i}{T}(T_{\mathrm{med}}+T_{\max})}\times$$

$$\times\left(1-e^{-\frac{i}{T}T_{\mathrm{med}}}\right)^{\frac{i-1}{2}}\left(e^{-\frac{i}{T}T_{\mathrm{med}}}-e^{-\frac{i}{T}T_{\max}}\right)^{\frac{i-3}{2}};$$

ж) известны значения $T_{\max}$, $T_{\min}$, T_{med}:

$$f_{\min,\mathrm{med},\max}(x,y,z)=\frac{n!}{\left(\left(\frac{n-3}{2}\right)!\right)^2}\lambda^3 e^{-\lambda(x+y+z)}\times$$

$$\times\left(\left(e^{-\lambda x}-e^{-\lambda y}\right)\left(e^{-\lambda y}-e^{-\lambda z}\right)\right)^{\frac{n-3}{2}},$$

если $-\infty<x<y<z<+\infty$, и $f_{\min,\mathrm{med},\max}(x,y,z)=0$ в остальных случаях,

$$\tilde{f}_i=\frac{i!}{\left(\left(\frac{i-3}{2}\right)!\right)^2}\left(\frac{i}{T}\right)^3 e^{-\frac{i}{T}(T_{\min}+T_{\mathrm{med}}+T_{\max})}\times$$

$$\times\left(\left(e^{-\frac{i}{T}T_{\min}}-e^{-\frac{i}{T}T_{\mathrm{med}}}\right)\left(e^{-\frac{i}{T}T_{\mathrm{med}}}-e^{-\frac{i}{T}T_{\max}}\right)\right)^{\frac{i-3}{2}}.$$

4. Оценка $\Pr^*$ кумулятивного риска, связанного с рискованным поведением, полученная по данным о максимальном и минимальном интервалах между эпизодами рискованного поведения. Преобразуя рассмотренную ранее формулу [1, 7], получим

$$\Pr=1-e^{-np},$$

где p — риск заражения за один эпизод; n — число эпизодов, произошедших за период времени T; $\Pr$ — кумулятивный риск.

Оценка $\Pr^*$ вычисляется следующим образом [9]:

$$\mathrm{Pr}^* = \mathsf{E}\,\mathrm{Pr} = \sum_{i=0}^{5T} p_i \,\mathrm{Pr}_i = \sum_{i=0}^{5T} p_i \left(1 - e^{-n_i p}\right).$$

Тогда для случайной величины Pr имеем:

$$\mathsf{E}\,\mathrm{Pr} = \mathrm{Pr}^*,\ \mathsf{D}\,\mathrm{Pr} = \sum_{i=0}^{5T} p_i \left(1 - e^{-n_i p}\right)^2 - \left(\mathrm{Pr}^*\right)^2 .$$

Доверительный интервал

$$\left(\mathrm{Pr}^* - \sqrt{\frac{\mathsf{D}\,\mathrm{Pr}}{\alpha}},\ \mathrm{Pr}^* + \sqrt{\frac{\mathsf{D}\,\mathrm{Pr}}{\alpha}}\right).$$

3. Согласованность данных. В результате интервью нам становятся известными сведения о нескольких (до 3–4) последних эпизодах поведения, о максимальном, минимальном и обычном интервале между эпизодами. Ранее нами получены оценки интенсивности рискованного поведения и кумулятивного риска, с ним связанного [1–3]. Заметим, однако, что в одном случае эти оценки получены только по данным о последних эпизодах рискованного поведения, а в другом — только по данным о «рекордных» (максимальном и минимальном) интервалах между эпизодами указанного поведения и об обычном интервале, т. е. по ответам одного респондента рассчитываются несколько не связанных между собой оценок интенсивности поведения. Таким образом, необходимо оценить согласованность различных оценок между собой.

Дадим формальную постановку задачи. Пусть получена оценка интенсивности $\tilde{\lambda}$ одним из методов. Тогда вероятность $P\left(\lambda \le \tilde{\lambda}\right)$ того, что оценка интенсивности, вычисленная другим методом, будет не больше $\tilde{\lambda}$, будет отражать согласованность этих оценок: чем больше эта вероятность, тем лучше согласованы оценки.

Рассмотрим согласованность оценок, полученных в работах [1–3]. Пусть $\hat{\lambda} = \nu/\tau$ — оценка интенсивности, рассчитанная по данным о ν последних эпизодах рискованного поведения респондента. Пусть также нам известны данные о «рекордных» интервалах между эпизодами рискованного поведения, причем эти данные могут быть представлены в любом сочетании (например, только о минимальном интервале, ~~о~~ максимальном, ~~о~~ минимальном и обычном и т. д.). Вычислим вероятность $P\left(\lambda \le \hat{\lambda}\right)$. Перейдем от интенсивности λ к числу эпизодов n

рискованного поведения за интервал времени T, используя формулу $\lambda = n/T$. Тогда

$$P\left(\lambda \le \hat{\lambda}\right) = P\left(n \le \frac{\nu}{\tau}T\right) = \sum_{i \le \frac{\nu}{\tau}T} p_i = \sum_{i \le \frac{\nu}{\tau}T} \frac{\tilde{f}_i}{\sum_{k=0}^{5T} \tilde{f}_k}.$$

Рассмотрим обратное соотношение, т. е. пусть нам известно значение λ^* — оценка интенсивности, полученная по данным о «рекордных» интервалах между эпизодами. Тогда вероятность того, что оценка интенсивности, вычисленная по данным о ν последних эпизодах, не больше, чем λ^*, выражается следующим образом:

$$P\left(\lambda \le \lambda^*\right) = \int_0^{\lambda^*} g(t)dt,$$

где $g(\lambda)$ — плотность распределения случайной величины λ.

Как уже отмечалось, рискованное поведение рассматривается как пуассоновский процесс с основным уравнением

$$\Pr(t,k,\lambda) = \frac{(\lambda t)^k}{k!} e^{-\lambda t},$$

где λ — интенсивность рискованного поведения; t — промежуток времени наблюдения за поведением респондента; k — число эпизодов рискованного поведения, случившихся в этот промежуток; $\Pr(t,k,\lambda)$ — вероятность того, что за промежуток t при поведении с интенсивностью λ случится k эпизодов указанного поведения.

Тогда плотность распределения имеет вид

$$g(\lambda) = \alpha \Pr(\tau,\nu,\lambda),$$

где α — коэффициент, нормирующий $\Pr(\tau,\nu,\lambda)$, т. е.

$$\int_{-\infty}^{+\infty} g(t)dt = 1.$$

Ранее мы рассматривали общий случай, когда значение интенсивности принадлежало промежутку $[0,+\infty]$, поэтому выполнялось $\alpha = \tau$. Однако отметим, что, как правило, происходит не более пяти эпизодов рискованного поведения в день, т. е. $0 \le \lambda \le 5$. Вычислим коэффициент α для этого случая:

$$\int_{-\infty}^{+\infty} g(t)\,dt = \int_0^5 \alpha \frac{(\tau t)^{\nu}}{\nu!} e^{-\tau t} dt = [y = \tau t] = \frac{\alpha}{\tau} \int_0^{5\tau} \frac{y^{\nu}}{\nu!} e^{-y} dy =$$

$$= \frac{\alpha}{\tau} \left(-e^{-y} \frac{y^{\nu}}{\nu!} \Bigg|_0^{5\tau} + \int_0^{5\tau} \frac{y^{\nu-1}}{(\nu-1)!} e^{-y} dy \right) = \frac{\alpha}{\tau} \left(1 - e^{-5\tau} \sum_{i=0}^{\nu} \frac{(5\tau)^i}{i!} \right),$$

$$\alpha = \frac{\tau}{1 - e^{-5\tau} \sum\limits_{i=0}^{\nu} \frac{(5\tau)^i}{i!}}.$$

Тогда искомая вероятность имеет вид:

$$P(\lambda \le \lambda^*) = \int_0^{\lambda^*} g(t) dt = \alpha \int_0^{\lambda^*} \frac{(\tau t)^{\nu}}{\nu!} e^{-\tau t} dt = \frac{\alpha}{\tau} \left(1 - e^{-\tau\lambda^*} \sum_{i=0}^{\nu} \frac{(\tau\lambda^*)^i}{i!} \right) =$$

$$= \left(1 - e^{-\tau\lambda^*} \sum_{i=0}^{\nu} \frac{(\tau\lambda^*)^i}{i!} \right) \Bigg/ \left(1 - e^{-5\tau} \sum_{k=0}^{\nu} \frac{(5\tau)^k}{k!} \right).$$

Заметим, что по данным о «рекордных» интервалах между эпизодами рискованного поведения мы получаем не одну, а несколько оценок (при различных сочетаниях этих данных). Таким образом, необходимо проверить согласованность этих оценок между собой.

Пусть нам известно значение n^* — оценка числа эпизодов рискованного поведения, произошедших за время T, причем эта оценка получена по одному из возможных наборов исходных данных (например, только по минимальному интервалу между эпизодами). Тогда вероятность того, что значение, полученное по другим данным (например, по максимальному и минимальному интервалам в совокупности), будет не больше n^*, вычисляется следующим образом:

$$P(n \le n^*) = \sum_{i=0}^{n^*} p_i = \sum_{i=0}^{n^*} \frac{\tilde{f}_i}{\sum\limits_{k=0}^{5T} \tilde{f}_k}.$$

4. Заключение. На основе неравенства Чебышева получены интервальные оценки интенсивности рискованного поведения и риска, с ним связанного. Также описан подход к проверке согласованности оценок, полученных различными способами, по различным исходным данным.

Представленные результаты могут использоваться не только в эпидемиологии, они могут быть полезны в любой области знаний, изучающей поведение человека. В частности, в психологии особенно важной бывает оценка рискованного поведения при изучении дезадаптивного поведения и разработке психокорректирующих воздействий.

Следует заметить, что неравенство Чебышева позволило дать достаточно «грубые» оценки; в дальнейших исследованиях их планируется уточнить, воспользовавшись особенностями распределений вероятности, лежащих в основе выбранной математической модели рискованного поведения.

Литература

1. *Пащенко А. Е., Тулупьев А. Л., Тулупьева Т. В.* Оценка интенсивности поведения респондента в условиях информационного дефицита // Тр. СПИИРАН. 2008. Вып. 7. С. 239–254.
2. *Пащенко А.Е., Суворова А.В.* Программный комплекс для экспертного оценивания интенсивности поведения респондента в условиях дефицита информации // Науч. докл. науч.-практ. конф. студентов, аспирантов, молодых ученых и специалистов «Интегрированные модели, мягкие вычисления, вероятностные системы и комплексы программ в искусственном интеллекте» Коломна, 26–27 мая 2009 г. Т. 2. М., 2009. С. 220–241.
3. *Тулупьев А. Л., Суворова А. В., Тулупьева Т. В., Пащенко А. Е.* Косвенные оценки и сравнение параметров угрозообразующего поведения в разных группах по неполным и неточным данным // Тр. СПИИРАН. 2009. Вып. 9. С. 252–261.
4. *Невзоров В. Б.* Рекорды. Математическая теория. М.: ФАЗИС, 2000. 244 с.
5. *Тулупьева Т. В., Тулупьев А. Л., Пащенко А. Е., Красносельских Т. В.* Приверженность ВААРТ и рискованное поведение среди пациентов Санкт-Петербургского Центра-СПИД: статистические модели, психологические и социо-демографические факторы // Тр. СПИИРАН. 2008. Вып. 6. С. 207–237.
6. *Тулупьева Т. В., Пащенко А. Е. , Тулупьев А. Л., Красносельских Т. В., Казакова О. С.* Модели ВИЧ-рискованного поведения в контексте психологической защиты и других адаптивных стилей. СПб.: Наука, 2008. 140 с.
7. *Bell D. C., Trevino R. A.* Modeling HIV Risk [Epidemiology] // J. Acquir Immune Defic Syndr. 1999. Vol. 22, N 3. P. 280–287.
8. *Хованов Н. В*. Анализ и синтез показателей при информационном дефиците // СПб., 1996. 196 с.
9. *Пащенко А.Е., Тулупьев А.Л., Суворова А.В., Тулупьева Т.В.* Сравнение параметров угрозообразующего поведения в разных группах на основе неполных и неточных данных // Тр. СПИИРАН. 2009. Вып. 9. С. 252–261.

Суворова Алена Владимировна — студент математико-механического факультета Санкт-Петербургского государственного университета (СПбГУ). Область научных интересов: математическая статистика, теория вероятности. Число научных публикаций — 6. SUVALV@mail.ru; СПбГУ, математико-механический факультет, Университетский пр., д.28, Петродворец, Санкт-Петербург, 198504, РФ. Научный руководитель А.Л. Тулупьев.

Suvorova Alena Vladimirovna — student of the Faculty of Mathematics and Mechanics of the Saint Petersburg State University. Research interests: mathematical statistics, probability theory. The number of publications — 6. SUVALV@mail.ru; St.Petersburg State University, Faculty of Mathematics and Mechanics, 28, Universitetsky prospekt, 198504, Peterhof, St. Petersburg, Russia. Supervisor — A.L. Tulupiev.

Тулупьева Татьяна Валентиновна — канд. психол. наук, доцент; старший научный сотрудник научно-исследовательской группы междисциплинарных проблем информатики Учреждения Российской академии наук Санкт-Петербургский институт информатики и автоматизации РАН (СПИИРАН), доцент кафедры информатики математико-механического факультета Санкт-Петербургского государственного университета (СПбГУ), доцент кафедры психологии управления и педагогики Северо-Западной академии государственной службы (СЗАГС). Область научных интересов: применение методов математики и информатики в гуманитарных исследованиях, информатизация организации и проведения психологических исследований, применение методов биостатистики в эпидемиологии, психология личности, психология управления. Число научных публикаций — 45. TVT@iias.spb.su, www.tulupyev.spb.ru; СПИИРАН, 14-я линия В.О., д.39, Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450.

Tulupyeva Tatiana Valentinovna — Ph.D. in Psychology, associate professor; senior researcher, Interdisciplinary Computer Science Research and Development Group, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), associate professor, Computer Science Department, Faculty of Mathematics and Mechanics, St. Petersburg State University (SPbSU), associate professor, Management Psychology and Pedagogic Department, North-West Academy of Public Administration (NWAPA). Research interests: application of mathematics and computer science in humanities, informatization of psychological studies, application of biostatistics in epidemiology, psychology of personality, management psychology. The number of publications — 45. TVT@iias.spb.su, www.tulupyev.spb.ru; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450.

Тулупьев Александр Львович — канд. физ.-мат. наук, доцент; ведущий научный сотрудник научно-исследовательской группы междисциплинарных проблем информатики Учреждения Российской академии наук Санкт-Петербургский институт информатики и автоматизации РАН (СПИИРАН), доцент кафедры информатики математико-механического факультета С.-Петербургского государственного университета (СПбГУ). Область научных интересов: представление и обработка данных и знаний с неопределенностью, применение методов математики и информатики в социокультурных исследованиях, применение методов биостатистики и математического моделирования в эпидемиологии, технология разработки программных комплексов с СУБД. Число научных публикаций — 140. ALT@iias.spb.su, www.tulupyev.spb.ru; СПИИРАН, 14-я линия В.О., д.39, Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450.

Tulupyev Alexander Lvovich — Ph.D. in Appl. Math. and CS, associate professor; leading researcher, Interdisciplinary Computer Science Research and Development Group, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), associate professor, Computer Science Department, Faculty of Mathematics and Mechanics, St. Petersburg State University (SPbSU). Research interests: uncertain knowledge and data representation and processing, application of mathematics and computer science in sociocultural studies, applications of biostatistics and mathematical modeling in modern epidemiology, software technologies and development of information systems with databases. The number of publications — 140. ALT@iias.spb.su, www.tulupyev.spb.ru; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450.

Пащенко Антон Евгеньевич — младший научный сотрудник научно-исследовательской группы междисциплинарных проблем информатики Учреждения Российской академии наук Санкт-Петербургский институт информатики и автоматизации РАН (СПИИРАН). Область научных интересов: математическая статистика, статистическое моделирование, применение методов биостатистики и математического моделирования в эпидемиологии. Число научных публикаций — 35. AEP@iias.spb.su, www.tulupyev.spb.ru; СПИИРАН, 14-я линия В.О., д.39, Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450.

Paschenko Anton Evgen'evich — junior researcher, Interdisciplinary Computer Science Research and Development Group, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: mathematical statistics, statistical modeling, application of biostatistics and mathematical modeling in epidemiology. The number of publications — 35. AEP@iias.spb.su, www.tulupyev.spb.ru; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450.

Рекомендовано ЛПИ СПИИРАН, зав. лаб. Р.М. Юсупов , чл.-корр. РАН.
Статья поступила в редакцию 20.12.2009.

РЕФЕРАТ

Суворова А.В., Пащенко А.Е., Тулупьева Т.В., Тулупьев А.Л. **Построение доверительных интервалов оценок интенсивности рискованного поведения на основе неравенства Чебышева.**

В статье строятся интервальные оценки (доверительные интервалы) интенсивности рискованного поведения и риска, с ним связанного. Их построение основано на данных о последних эпизодах рискованного поведения респондента, которые используются для оценки интенсивности поведения, а также для оценки кумулятивного риска, связанного с рискованным поведением. Кроме того, даны интервальные оценки риска для случая с произвольным числом эпизодов поведения, произошедших за заданный промежуток времени. Наконец, получена оценка кумулятивного риска, связанного с рискованным поведением, по данным о максимальном и минимальном интервалах между эпизодами рискованного поведения.

Описан подход к проверке согласованности оценок, полученных различными способами, по различным исходным данным.

SUMMARY

Syvorov A.V., Paschenko A.E., Tulupyev A.L., Tulupyeva T.V. **Risky behavior rate estimates calculation based on Chebychev inequality.**

The paper presents interval estimates (confidence intervals) of rate of risky behavior and related risk. They are based on the data about last episodes of risky behavior of the respondent which are used for an estimate of intensity of behavior, and also for an estimate of the cumulative risk related to risky behavior. Besides, interval estimate of risk in case of arbitrary number of episodes of behavior are given. Finally, the estimate of the cumulative risk related to risky behavior is done with data about the maximum and minimum intervals between episodes of risky behavior.

The approach to consistency verification of the estimates based on the various initial data is described.