

ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ПОКАЗАТЕЛЕЙ СРАВНИТЕЛЬНОГО МЕДИКО-БИОЛОГИЧЕСКОГО ИССЛЕДОВАНИЯ

Обеснюк В.Ф., Хромов-Борисов Н.Н.

Южно-Уральский институт биофизики (г.Озерск)

Желаемое направление в телеконференции: Секция 9. Проблемы биоинформатики и системной биологии

Тел: 8-35130-75236

E-mail: v-f-o@newmail.ru

456784, Россия, г.Озерск

пр. Ленина 23, кв.11

Обеснюку В.Ф.

ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ПОКАЗАТЕЛЕЙ СРАВНИТЕЛЬНОГО МЕДИКО-БИОЛОГИЧЕСКОГО ИССЛЕДОВАНИЯ

Обеснюк В.Ф.¹, Хромов-Борисов Н.Н.²

¹Южно-Уральский институт биофизики (г. Озерск)

²Санкт-Петербургский государственный медицинский университет им. акад. И.П. Павлова; Российский научно-исследовательский институт травматологии и ортопедии им. Р.Р. Вредена; НИИ акушерства и гинекологии им. Д.О. Отта СЗО РАМН, (г. Санкт-Петербург)

Результаты простейшего сравнительного исследования по методу “case-control” [1], либо “case-base” [2, 3], группируются в так называемую четырехпольную таблицу сопряженности (contingency epidemiological table), которая позволяет выполнить оценку показателей отношения шансов (OR , odds ratio) либо относительного риска (RR , relative risk). Так как всякая выборка является случайной даже при работе с некоторыми однородными большими группами, эти показатели будут случайными, обладающими некоторой неопределенностью, которая дополнительно маскируется дискретным характером наблюдений (шум дискретизации). Однако вероятность наблюдения эффекта или относительного эффекта непрерывна, поэтому неопределенность величин OR или RR должна получить непрерывную оценку.

Разрешить противоречие между дискретностью наблюдений и непрерывностью показателей OR и RR можно лишь приближенно, выбирая частотный аппроксимационный метод [1, 4 – 9] и, соответственно, доверительное оценивание (confidence interval), либо байесовский подход с использованием некоторой доступной априорной информации о возможных распределениях выборочных и измеряемых величин (credible interval). Каждая из таких задач в конкретном исследовании всегда может быть решена численно, благодаря чему те или иные методики оценки нашли реализацию в ряде известных приближенных компьютерных алгоритмов, например, в открытой среде программирования “R” [10], пакете “LePAC” [11] и др. Как оказалось, байесовский подход допускает также точное аналитическое представление непрерывной плотности распределения величин OR и RR , а значит, возможность получения сравнительно строгих оценок, как для интервалов неопределенности и их центров, так и для связанных с этими распределениями вторичных характеристик (показатель согласованности “каппа Коуэна” и т.п.) при самых общих (неинформативных) априорных

предположениях в рамках биномиального описания распределения дискретных событий в каждой независимой выборке.

Например, пусть при сравнении двух случайных выборок с объемами N (экспонированные) и L (интактные) наблюдалось M и K специфических “случаев”, соответственно. Тогда при априорном (до выборки) предположении, что каждый индивидум одной выборки имеет одинаковые шансы попасть в свою группу “случаев” этой же выборки, возможно строгое интегральное представление плотности вероятности наблюдения неизвестных показателей OR и RR , которое может быть выражено через известные специальные функции. Априорное предположение о наличии равных шансов соответствует одинаковой вероятности наблюдения целочисленных M среди $M = 0; 1; \dots N$ и K среди $K = 0; 1; \dots L$, а также – равномерному распределению ожидаемых частот непрерывных параметров доли “случаев” в каждой из двух выборок до их формирования (группирование событий по методу “case-base”).

В частности можно показать, что апостериорная условная плотность вероятности распределения относительного риска дается выражением

$$g(RR | M, N, K, L) = \begin{cases} \frac{B(M+K+2, L-K+1) \cdot RR^M}{B(M+1, N-M+1) \cdot B(K+1, L-K+1)} F(M-N, K+M+2; M+L+3; RR), & RR < 1; \\ \frac{B(M+K+2, L+N-K-M+1)}{B(M+1, N-M+1) \cdot B(K+1, L-K+1)}, & RR = 1; \\ \frac{B(M+K+2, N-M+1)}{B(M+1, N-M+1) \cdot B(K+1, L-K+1) \cdot RR^{K+2}} F\left(K-L, K+M+2; K+N+3; \frac{1}{RR}\right), & RR > 1. \end{cases}$$

Здесь числа M, N, K, L , однократно наблюдаемые эмпирически, выступают в роли параметров распределения; $F(a, b; c; z) = {}_2F_1(a, b; c; z)$ – гипергеометрическая функция Гаусса; $B(a, b)$ – бета-функция Эйлера. В области малых значений относительного риска плотность нарастает пропорционально $\sim RR^M$, а в области больших значений – убывает пропорционально $\sim 1/RR^{K+2}$. Иными словами, асимптотическое поведение распределения зависит в большей мере от наблюдаемого числа “случаев”, чем от объемов выборок. Обнаруженный характер асимптотик противоречит ранее найденным приближениям [1, 5, 6, 8], имевшим приемлемую пригодность только для описания центральной части распределения.

Условная плотность вероятности распределения отношения шансов дается выражением

$$f(OR | M, N, K, L) = \frac{B(M+K+2, N-M+L-K+2) \cdot OR^M}{B(M+1, N-M+1) \cdot B(K+1, L-K+1)} F(N+2, K+M+2; N+L+4; 1-OR), \quad OR > 0.$$

Однако оно шире распределения отношения рисков, и в области сравнительно частых специфических событий имеет хорошо различимое систематическое смещение. Таким образом, оценка OR , используемая в качестве приближения к искомой величине показателя RR во многих исследованиях, менее предпочтительна для применения. Оказалось также, что в ряде вычислительных пакетов программ гипергеометрическая функция неверно табулируется в области отрицательного аргумента, поэтому для организации вычислений за основу было принято ее интегральное представление. Примеры наблюдаемой формы распределений $g(RR | M, N, K, L)$ и $f(OR | M, N, K, L)$ приведены на рисунке, из которого следует, что попытки использования гауссова приближения для их описания [4] всегда будут приводить к неконтролируемым систематическим погрешностям как центральных, так и интервальных оценок. Кроме того, применение принципа максимального правдоподобия к оценкам относительного риска по однократному измерению будет приводить к его систематической недооценке. В качестве центра распределения более уместно оценивание медианы.

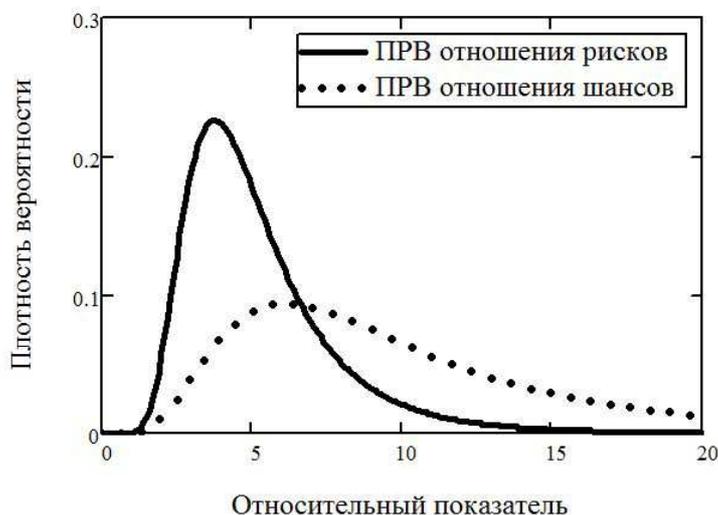


Рисунок. Пример плотности распределения вероятности относительных оценок RR и OR при сравнении двух случайных выборок. Интактные: 5 “случаев” из 50. Экспонированные: 13 “случаев” из 25.

Заметим, что сопоставление байесовских интервальных оценок отношения рисков с частотными [9] показало, что байесовский подход дает более узкие интервалы не-

определенности для RR при малом количестве наблюдаемых случаев M и K , что представляется понятным в силу использования дополнительной информации об априорных распределениях, а также в силу большей математической строгости. Для исследований с большой статистической мощностью оценки практически совпадают.

Отметим, что рассматривая задачу интервальной оценки показателя RR в рамках пуассоновской статистики редких событий, нетрудно получить иное, более простое апостериорное приближение для плотности вероятности, обладающее “правильным” асимптотическим поведением:

$$\tilde{g}(RR | M, N, K, L) = \frac{N/L}{B(M+1, K+1)} \cdot \frac{(RR \cdot N/L)^M}{(1 + RR \cdot N/L)^{K+M+2}}.$$

Комбинация параметров распределения такова, что при заданном соотношении объемов выборок (обычно $N/L \sim 1$) его можно считать двухпараметрическим, что позволяет удобно табулировать функцию в виде обычной таблицы подобно другим двухпараметрическим распределениям.

Отмеченное обстоятельство частично оправдывает практику применения двухпараметрического логнормального приближения в качестве аппроксимации центральной части распределения Бреслоу и Деэм [1], а также Альбомом и Норрелом ([5], стр.87–89; оценка логит-границ Вульфа). Однако логнормальное приближение и найденные распределения относительного риска имеют существенно различное поведение на “хвостах”, что не позволяет производить интервальное оценивание со слишком большой величиной доверительной вероятности, пользуясь границами Вульфа.

Известно, что выборочный метод “case–control” широко используется не только в эпидемиологических исследованиях. Величины RR по своему математическому смыслу аналогичны отношениям правдоподобий в диагностических тестах, широко применяемых в фармакологии, психологии, педагогике, генетике и т.п., как при описании их положительных результатов применения ($LR[+]$), так и при описании отрицательных результатов ($LR[-]$). Известно также, что в отдельных ситуациях не всякий алгоритм способен правильно давать точечную оценку указанным показателям, не говоря уже об интервальных. Как правило, это происходит при анализе четырехпольных таблиц с нулями или очень малыми значениями при исчислении “случаев” или “контролей”. Большие отклонения в оценках могут наблюдаться также при слишком высоком принятом доверительном уровне.

Для демонстрации работоспособности найденных распределений в нижеприведенной таблице показано сравнение наших результатов доверительного оценивания с лучшими известными нам алгоритмами других разработчиков. Таблицы соответствуют исследованию, в котором в группе из 34 больных результаты применения некоторого диагностического теста оказались положительными для всей выборки, в то время как в группе сравнения для здоровых наблюдалось 4 ложно-положительных “случая” и 4 811 истинно-отрицательных “контролей”. Заметим, что “эмпирическая” точечная оценка для $LR[-]$ здесь бесконечна, а нижняя интервальная – конечна.

Таблица 1А. Интервальные оценки отношения правдоподобий для положительных результатов диагностического теста $LR[+]$

Программа (алгоритм)	Доверительная вероятность					
	95%		99%		99,9%	
	Границы доверительных интервалов					
	Нижняя	Верхняя	Нижняя	Верхняя	Нижняя	Верхняя
R-скрипт [10]	455,9	2 883,6	370,4	4 343,3	296,5	7 403,3
LePAC [11]	456,2	2 883,0	370,7	4 347,1	296,7	7 409,4
MOVER [12]	455,0	2 883,9	368,0	4 342,9	291,9	7 401,2
Наст. работа	456,3	2 885,9	370,7	4 347,2	296,7	7 409,5

Таблица 1Б. Интервальные оценки отношения правдоподобий для отрицательных результатов диагностического теста $LR[-]$

Программа (алгоритм)	Доверительная вероятность					
	95%		99%		99,9%	
	Границы доверительных интервалов					
	Нижняя	Верхняя	Нижняя	Верхняя	Нижняя	Верхняя
R-скрипт [10]	10,0	1 381,3	7,1	6 974,8	5,1	69 990,0
LePAC [11]	10,0	1 384,5	7,1	6 996,8	5,1	66 481,2
MOVER [12]	10,0	1 381,5	7,1	6 976,0	5,1	69 905,5
Наст. работа	10,0	1 384,2	7,1	6 975,7	5,1	69 987,4

Заключение. Найдено явное непрерывное выражение для распределения эпидемиологических оценок отношения рисков в четырехпольных методах, которое можно считать точным при довольно общих априорных предположениях. При сопоставлении однородных групп распределение работоспособно в широких условиях при изучении как редких, так и частых событий. Рекомендуемые интервальные оценки несколько отличаются от логит-интервалов Вульфа. Найденное распределение может быть использовано в современном программном обеспечении эпидемиологических, медико-статистических и биологических исследований. Инструментальная ценность формулы распределения будет выше, прежде всего, в ситуациях с малым количеством наблюдаемых “случаев”.

ЛИТЕРАТУРА

1. *Breslow N.E., Day N.E.* Statistical Methods in Cancer Research, Vol. 1, The Analysis of Case-Control studies. — London etc., 1980. — 350 p.
2. *Kupper I.L., McMichael A.J., Spirtas R.* A Hybrid epidemiologic study design useful in estimating relative risk // *Journal of the American Statistical Association.* — 1975. — V.70. — P.524–528.
3. Эпидемиологический словарь. /Под ред. Дж.М.Ласта для Международной эпидемиологической ассоциации. — М.: ОИЗ, 2009. — 316 с.
4. *Кабак С.Л., Кабак Ю.С.* Какому методу статистической обработки результатов следует отдавать предпочтение при оценке эффективности нового метода лечения? // *Медицинский журнал* (Минск: БГМУ). — 2003. — №3. URL: <http://www.bsmu.by/files/mj/3-2003/31.pdf>, дата обращения 07.02.2013.
5. *Альбом А., Норелл С.* Введение в современную эпидемиологию. — Таллинн: RNE, 1996. — 122 с.
6. *Clayton D., Hills M.* Statistical Models in Epidemiology. — Oxford, New York, Tokyo, 1993. — 368 p.
7. *Флетчер Р., Флетчер С., Вагнер Э.* Клиническая эпидемиология. Основы доказательной медицины // Пер. с англ. — М: Медиа-Сфера, 1998. — 352 с.
8. *Гланц С.* Медико-биологическая статистика // Пер. с англ. — М.: Практика, 1998. — 459 с.
9. *Obesnyuk V.F.* Approximation of relative risk estimate distribution in a two-sample study using “case-base” reasoning. — In: *Radioactive Sources and Radiation Exposure Effects on Mayak PA Workers and Population Living in the Area of Nuclear Facility Influence.* Pt. 4. — Chelyabinsk: СРН, 2012. — P. 113–121.
10. *Agresti A.* URL: <http://www.stat.ufl.edu/~aa/cda/R/bayes/index.html>, дата обращения 07.02.2013.
11. *Lecoutre B., Poitevineau J.* LePAC version 2.0.41 (Program for the Analysis of Comparisons) URL: <http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/PAC.htm>, дата обращения 07.02.2013.
12. *Zou G.Y., Donner A.* Construction of confidence limits about effects measures: a general approach.// *Statistics in Medicine*, 2008, **27**, P.1693–1702.