

Н. Н. Хромов-Борисов

## КОРРЕКТНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ ЧАСТОТ ГЕНОТИПОВ ПО ГЕНАМ СЕМЕЙСТВ *UCP* И *PPAR* В ГРУППЕ ЖИТЕЛЕЙ БЛОКАДНОГО ЛЕНИНГРАДА И КОНТРОЛЬНОЙ ГРУППЕ

Российский научно-исследовательский институт травматологии и ортопедии им. Р. Р. Вредена,  
195427 Санкт-Петербург, ул. Акад. Байкова, 8; e-mail: Nikita.KhromovBorisov@gmail.com

Корректный статистический анализ данных, опубликованных в журнале «Успехи геронтологии» И. В. Поляковой и соавт. (2014), убедительно свидетельствует, что, вопреки мнению авторов, распределения генотипов у жителей блокадного Ленинграда и у жителей Северо-Западного региона России статистически неразличимы по всем пяти изученным генам. Основными причинами ошибочных выводов авторов являются пренебрежение проблемой множественных сравнений и невозможность сформировать адекватную контрольную группу. Представлена схема гармонизированного статистического анализа подобных данных, которая включает не только частотные, но и бейзовские точечные и интервальные оценки для долей генотипов и их разностей, для индекса фиксации (коэффициента инбридинга)  $F_{IS}$ , для размера эффекта  $\phi$  (коэффициента сопряженности Пирсона) и для достигнутой мощности  $(1 - \beta)$ , а также оценки апостериорных вероятностей для нулевой гипотезы  $P(H_0|D)$ , бейзовы факторы  $BF_{01}$ , значения  $p_{obs}$  с границами предсказательного интервала и значения  $P_s$  с поправкой на множественность проверяемых нулевых гипотез.

**Ключевые слова:** жители блокадного Ленинграда, частоты генотипов, статистический анализ, индекс фиксации, размер эффекта, статистическая мощность, апостериорные вероятности, предсказательные интервалы, поправки на множественность сравнений, бейзов фактор

В 2014 г. в этом журнале (т. 27, № 3) опубликована статья «Анализ частот генотипов и аллелей генов семейств *UCP* и *PPAR* в группе жителей блокадного Ленинграда и контрольной группе» [8]. Работа получила широкую огласку примерно в 30 отечественных радио-, теле-, электронных и печатных СМИ, где она преподносится как крупное достижение отечественной науки с броскими слоганами: «Блокадная ДНК», «Блокадный ген», «Гены блокадника», «Гены жизни», «Гены стойкости», «Код героев», «Гены выживания», «Ген "блокадного Ленинграда"». Краткая новостная заметка об этой работе опубликована российскими

журналистами в журнале «Science» [34], которая процитирована в письме отечественных специалистов по геномной биоинформатике в этом же журнале [30].

Основные выводы работы гласят: «У жителей блокадного города установлено достоверное увеличение частоты аллеля *G* и генотипа *G/C* гена *PPARA*, что в условиях длительного голода способствовало выживанию. А генотип *C/C* гена *UCP3* способствовал выживанию только у женщин». Однако эти и все другие выводы этой работы вызывают серьезные сомнения, прежде всего из-за применения авторами неадекватного статистического анализа.

Фактически, единственным инструментом статистического анализа полученных данных в этой работе является статистика критерия  $\chi^2$ , соответствующие значения  $p$  и один (самый снисходительный) уровень значимости  $\alpha=0,05$ . Однако в научной литературе опубликовано несколько методических работ, в которых сформулированы требования к статистическому анализу результатов исследований связи генотип—фенотип [11–14, 35]. Обязательной должна быть проверка согласия наблюдаемых частот генотипов по каждому гену (маркеру) с равновесием Харди—Вейнберга (РХВ). Кроме приемлемого уровня значимости ( $\alpha$ ), исследователям следует заранее выбрать приемлемую мощность статистического критерия  $(1 - \beta)$  и приемлемое значение размера эффекта (ES — effect size). Объемы выборок должны быть достаточно большими для того, чтобы обеспечить желаемую мощность критерия. Критерий статистической значимости должен быть достаточно строгим, то есть пороговое значение  $\alpha$  должно быть малым. Должна быть введена поправка на множественность сравнений по всему спектру возможных статистических проверок. Вместо приближенных асимптотических методов (типа  $\chi^2$ )

рекомендуется использовать рандомизационные и перестановочные процедуры, основанные на алгоритмах Монте-Карло и бутстрепа. Наряду с традиционными (ортодоксальными частотными) процедурами рекомендуется использовать методы бейзовской статистики. В частности, наряду со значениями  $\rho$ , или даже вместо них, следует приводить значения бейзова фактора ( $BF$ ) как практически более приемлемого показателя значимости наблюдаемых эффектов [11–14, 35].

Недостаточно ограничиваться точечными оценками значений  $\rho$  и/или значениями бейзова фактора. Необходимо для большинства перечисленных статистических показателей вычислять интервальные оценки типа традиционных доверительных интервалов (ДИ) или бейзовских правдоподобных (credible) интервалов. В первую очередь, это касается оценок размера эффекта.

Важнейшим требованием является проведение независимых повторных экспериментов и наблюдений, поскольку достаточно часто результаты одиночного исследования связи генотип—фенотип впоследствии не воспроизводятся — эффект, известный под названием «проклятье победителя» (Winner's course). Только если результаты таких исследований воспроизводятся в независимых повторностях, выводы из них будут убедительными [13].

К перечисленным требованиям следует добавить процедуры статистических предсказаний, то есть вычисления предсказательных вероятностей и предсказательных интервалов (ПИ), в частности для значений  $\rho$  [23].

#### **Несостоятельность значений $\rho < 0,05$ как основного показателя статистической значимости наблюдаемых эффектов**

Как известно, значения  $\rho$  ничего не говорят ни о вероятности проверяемой нулевой гипотезы, ни о размере эффекта. Традиционно принято считать, что значение  $\rho$  есть мера свидетельства против нулевой гипотезы: чем меньше значение  $\rho$ , тем, якобы, сильнее основание для сомнения в нулевой гипотезе. Отсюда возникает распространенное заблуждение интерпретировать значение  $\rho$  как вероятность нулевой гипотезы  $H_0$ . Однако, по определению, значение  $\rho$  является вероятностью получить наблюдаемые данные и все те, которые еще менее вероятны (или еще более экстремальные), при условии, что справедлива нулевая гипотеза:  $\rho = Pr(D \geq D_{obs} | H_0)$ .

Здесь неравенство  $D \geq D_{obs}$  обозначает подмножество, состоящее из наблюдаемых данных  $D_{obs}$  и других (ненаблюдаемых) более редких или более экстремальных. Чаще всего в их роли выступают соответствующие значения той или иной критериальной статистики. Очевидно, что оно никак не может быть одновременно (апостериорной) вероятностью нулевой гипотезы:  $Pr(H_0 | D)$ .

Но именно вероятность нулевой и/или альтернативной гипотез интересуют исследователей. Взаимоотношения значений  $\rho$  и апостериорной вероятности нулевой гипотезы  $Pr(H_0 | D)$  представлены в табл. 1 в виде неравенства. Действительно, можно видеть, что только начиная со значения  $\rho = 0,001$  нижняя граница для вероятности нулевой гипотезы может достигать значения 0,019 (~2%). Но может быть и больше.

Кроме того, легко понять, что при повторении эксперимента новые данные будут отличаться от полученных ранее. Соответственно, и значения  $\rho$  тоже будут варьировать, а это означает, что каждое из них является реализацией случайной переменной  $\tilde{\rho}$ . Известно, что при справедливости нулевой гипотезы переменная  $\tilde{\rho}$  распределена равномерно на отрезке от 0 до 1. Это, в свою очередь, означает, что в одиночном эксперименте даже при справедливости нулевой гипотезы переменная  $\tilde{\rho}$  случайно может принимать любое значение на этом отрезке — как очень малое (близкое к нулю), так и очень большое (близкое к 1). Отсюда с неизбежностью следует, что любой эксперимент должен быть повторен многократно. Только при таком очевидном условии научные выводы могут быть надежными и убедительными.

Более того, принятие в подавляющем большинстве публикаций значения  $\alpha = 0,05$  в качестве критического уровня значимости является проявлением синдрома статистической снисходительности и доверчивости. Еще в 50-е гг. прошедшего столетия А. Н. Колмогоров призывал рассматривать это значение малопригодным для серьезных статистических выводов и рекомендовал ориентироваться на значение  $\alpha = 0,001$ . Эти его рекомендации воспроизведены в пяти энциклопедиях [1–7], две из которых переведены на английский [20–22]. Сейчас это становится общепринятым [15, 19, 25–28, 36]<sup>1</sup>. Обоснование этой рекомендации наглядно представлено в табл. 1.

<sup>1</sup> Об исключительной важности этих работ свидетельствует тот факт, что одна из них опубликована одновременно в четырех медицинских журналах [25–28].

Калибровка значений  $p$ 

Наблюдаемое $p$ -значение, $p_{obs}$	Верхняя граница 95% предсказательного интервала для $p$ [16, 23]	Нижняя граница для вероятности нулевой гипотезы $P(H_0 D)$ [32, 33, 37]	Верхняя граница для вероятности воспроизведения $P_{rep}$ [18]
0,05	0,79	0,41	0,50
0,01	0,58	0,13	0,67
0,001	0,30	0,019	0,83
$10^{-4}$	0,13	0,0025	0,91
$10^{-5}$	<b>0,050</b>	$3,1 \cdot 10^{-4}$	0,96
$10^{-6}$	0,017	$3,8 \cdot 10^{-5}$	0,98

Примечание. Жирным шрифтом выделены значения, которые можно считать более или менее убедительными свидетельствами против нулевой гипотезы  $H_0$ . Использованы программы: P-value Prediction Interval Calculator [23] — <http://www.nature.com/mp/journal/v19/n12/extref/mp2013184x1.xls>; LePrep — Probabilities of replication — <http://lmrs.univ-rouen.fr/Persopage/Lecoutre/PAC.htm>.

### Проблема множественных сравнений

Справедливость выводов и интерпретация практически всех представленных в обсуждаемой работе результатов вызывают сомнение, прежде всего из-за того, что в ней проигнорирована известная проблема множественных сравнений при статистическом анализе данных.

Логика традиционной процедуры проверки статистических (нулевых) гипотез состоит в том, что при проверке каждой статистической гипотезы закладывается возможность ошибочного отклонения верной нулевой гипотезы (ошибка первого рода). Очевидно, что чем больше гипотез мы проверяем на одном и том же наборе данных, тем больше будет возрастать вероятность допустить хотя бы одну такую ошибку. Это явление называют эффектом множественных сравнений, и его обязательно следует учитывать при статистическом анализе. Это одна из сложнейших проблем прикладной статистики, которая не имеет однозначного решения. Строго говоря, поправки на множественность сравнений следует применять к выбранному исследователями уровню значимости ( $\alpha$ ), но на практике удобнее корректировать получаемые значения  $\rho_{obs}$ . При относительно небольшом числе ( $k$ ) проверяемых гипотез (сравнений) достаточно использовать известные поправки для значений  $\rho_{obs}$  по Шидаку или по Бонферрони. По Шидаку скорректированное значение  $\rho_s$  вычисляется по формуле:

$$\rho_s = 1 - (1 - \rho_{obs})^k.$$

При малых значениях  $\rho_{obs}$  эта поправка упрощается до поправки по Бонферрони, то есть сводится к простому умножению полученного значения  $\rho_{obs}$  на число сравнений ( $k$ ):  $\rho_B = \rho_{obs} \cdot k$ .

В обсуждаемой работе сравнивают две группы — жителей блокадного Ленинграда и кон-

трольную группу жителей Северо-Западного региона России, не находившихся в Ленинграде в годы блокады. У всех них определены частоты аллелей и генотипов по пяти генам —  $UCP2$ ,  $UCP3$ ,  $PPARA$ ,  $PPARD$  и  $PPARG$ , то есть  $k = 5$ . Без поправок на множественность, статистически значимое различие между сравниваемыми группами обнаружено лишь для одного гена  $PPARA$  из пяти изученных:  $\rho_{obs} = 0,035$ . Для всех остальных четырех генов значения  $\rho_{obs}$  превышают уровень значимости  $\alpha = 0,05$ . Поэтому достаточно произвести поправку на множественность только для этого значения:  $\rho_s = 1 - (1 - 0,035)^5 = 0,163$ ;  $\rho_B = 0,035 \cdot 5 = 0,175$ .

Это означает, что все значения  $\rho_{obs}$ , полученные для любого из пяти генов, являются незначимыми, и даже меньшее из них, нескорректированное исходное  $\rho_{obs} = 0,035$ , оказывается ложно значимым. Таким образом, для всех пяти генов распределения генотипов между сравниваемыми группами оказываются статистически неразличимыми (однородными). В итоге, ни один вывод работы не соответствует истине, и на этом полагается ставить точку. Поэтому нет никаких оснований для того, чтобы пытаться выискивать какие-либо частные различия между частотами отдельных генотипов (или аллелей) в разных вариантах дробления исходных двух групп на подгруппы по полу, по возрастным категориям, по длительности пребывания в блокадном Ленинграде, социальному статусу блокадников (дети, иждивенцы), причем некоторые из этих подгрупп чрезвычайно малочисленны — 4 долгожителя, 2 иждивенца мужского пола.

Всего в обсуждаемой работе приведены значения  $\rho_{obs}$  для 28 различных комбинаций генотипов и аллелей в разнообразных подгруппах, 19 из которых значимы на традиционном критическом  $\alpha = 0,05$ . Наименьшим из них является  $\rho_{obs} = 0,01$ ,

которое, по мнению авторов, свидетельствует о якобы более низкой частоте генотипа G/C по гену PPARA у жителей блокадного Ленинграда по сравнению с таковой в контрольной группе (два сообщенных в работе значения  $\rho_{obs} = 0,005$  и  $\rho_{obs} = 0,009$  на поверку оказались вычисленными неверно). Однако после поправки на множественность реальное значение  $\rho_s$  становится незначимым:  $\rho_s = 1 - (1 - 0,01)^{28} = 0,25$ .

Соответственно, вопреки мнению авторов, все — без единого исключения — 28 значений  $\rho_{obs}$  оказываются незначимыми, что и следовало ожидать.

### Гармонизированный статистический анализ данных

В связи с вышесказанным, современный статистический анализ не ограничивается сообщением одного-единственного показателя согласия с нулевой гипотезой, в качестве которого чаще всего выступает значение  $\rho_{obs}$ , даже если оно скорректировано на множественность сравнений. Анализ данных должен быть комплексным, гармонизированным. В руководствах и методических статьях по статистике давно рекомендуется, наряду со значениями  $\rho_{obs}$  или даже вместо них, указывать ДИ для размера эффекта, например для разности сравниваемых параметров.

### Размер эффекта

Распространенной мерой силы связи генотип—фенотип является отношение шансов (OR), когда исходные данные представлены в виде таблицы сопряженности 2×2 [12, 14, 35]. Считается, что практического внимания заслуживают те связи, для которых нижняя граница 95 % ДИ  $OR_L > 2,2$  [12], и теоретический анализ подтверждает это [9, 31]. Когда в анализируемой таблице сопряженности более двух строк и/или столбцов, тогда интегральным показателем размера эффекта может выступать обычный коэффициент сопряженности Пирсона:

$$\varphi = \sqrt{\frac{\chi^2}{N}}$$

На практике значения  $\varphi < 0,2$  считаются пренебрежимо малыми.

### Мощность статистического критерия

Не менее важным показателем надежности статистических выводов является мощность применяемого статистического критерия ( $1 - \beta$ ), которая есть вероятность правильного принятия верной альтернативной гипотезы  $H_1$ . При наблюдаемом размере эффекта  $\varphi$  и заданном уровне значимости  $\alpha$  можно оценить так называемую достигнутую мощность. Наименьшим приемлемым (снисходительным) значением мощности обычно принято считать значение  $(1 - \beta) = 0,8$ .

### Бейзов фактор

Современные руководства и методические статьи рекомендуют дополнять результаты традиционного (ортодоксального частотного) статистического анализа результатами бейзовского подхода [11–14, 35]. В таком случае вычисляют значения бейзова фактора  $BF_{01}$ , который является отношением вероятности (правдоподобия) получить наблюдаемые данные при справедливости нулевой гипотезы к вероятности их получения при альтернативной гипотезе. Одновременно этот показатель показывает, во сколько раз апостериорные шансы (odds) в пользу нулевой гипотезы (против альтернативной) отличаются от ее априорных шансов:

$$BF_{01} = \frac{P(D|H_0)}{P(D|H_1)} = \frac{P(H_0|D)}{P(H_1|D)} \cdot \frac{P(H_0)}{P(H_1)}$$

Значения  $BF_{01} > 1$  свидетельствуют в пользу нулевой гипотезы против альтернативной, значения  $BF_{01} < 1$  свидетельствуют в пользу альтернативной гипотезы против нулевой. На практике значения  $BF_{01} < 3$  принято считать пренебрежимо малыми.

### Предсказательные вероятности и предсказательные интервалы

Кроме ДИ, очень важной характеристикой изучаемых вероятностно-статистических явлений являются предсказательные интервалы (ПИ) для оцениваемых параметров и предсказательные (апостериорные) вероятности гипотез. В табл. 1 приведены верхние границы таких 95 % интервалов для значений  $\rho$  [23] и нижняя граница апостериорной вероятности нулевой гипотезы.

Результаты описанного выше гармонизированного статистического анализа для данных о распределении генотипов по диморфизму +2528G/C в гене PPARA у жителей блокадного Ленинграда

и в контрольной группе приведены в табл. 2. Для вычисления большинства показателей использованы точные методы. В верхней части таблицы результаты авторов дополнены интервальными оценками для долей генотипов и их разностей, точечными и интервальными (бейзовскими) оценками индекса фиксации (коэффициента инбридинга)  $F_{IS}$ , значениями бейзовых факторов  $BF_{01}$ , оценками апостериорных вероятностей для нулевой гипотезы о согласии с РХВ. В нижней части таблицы приведено наблюдаемое значение  $p_{obs}$  для точного критерия проверки однородности с границами 95 % ПИ, а также значение  $P_s$  с поправкой на пять проверяемых нулевых гипотез, бейзов фактор  $BF_{01}$ , размер эффекта  $\phi$  на основе статистики  $\chi^2$  с 95 % ДИ, нижняя граница для апостериорной вероятности нулевой гипотезы  $P(H_0|D)$  и достигнутая мощность  $(1 - \beta)$  также с 95 % ДИ.

С одной стороны, полученное значение  $p = 0,035$  меньше традиционного уровня  $\alpha=0,05$ . И оно согласуется с тем, что 95 % ДИ для раз-

ности долей для генотипа G/C не покрывает безразличное значение 0. Аналогичным образом 95 % ДИ для коэффициента сопряженности  $\phi$  тоже не покрывает безразличное значение 0. Формально (без поправки на множественность сравнений) можно было бы говорить о статистически значимом различии сравниваемых распределений генотипов. Однако различие значимо лишь на самом снисходительном уровне  $\alpha = 0,05$ . Если же вычислить 99 % ДИ, то он покроет 0. С другой стороны, полученное значение бейзова фактора  $BF_{01}=1,59$  превышает безразличное значение 1 и формально свидетельствует в пользу нулевой гипотезы  $H_0$  против альтернативной гипотезы  $H_1$  и противоречит выводу на основе нескорректированного значения  $p_{obs}$ . В таких случаях предпочтение следует отдавать выводам, основанным на бейзовском подходе.

С третьей стороны, 95 % ПИ для значения  $p$  получается чрезвычайно широким — от  $10^{-6}$  до 0,83. Это означает, что если повторить это иссле-

Таблица 2

Гармонизированный статистический анализ распределений генотипов по диморфизму +2528G/C в гене PPARA у жителей блокадного Ленинграда и в контрольной группе

PPARA		Блокадники		Контрольная группа		Разность
Генотип		$n_{1i}$	$f_{1i}$ с 95 % ДИ	$n_{2i}$	$f_{2i}$ с 95 % ДИ	$\Delta$ с 95 % ДИ
GG		145	0,62 0,70 0,78	80	0,47 0,58 0,68	-0,01 0,12 0,26
GC		53	0,19 0,26 0,34	54	0,29 0,39 0,43	-0,21 -0,13 -0,002
CC		8	0,016 0,039 0,081	5	0,011 0,036 0,092	-0,058 0,003 0,052
Всего		206	1,00	139	1,00	
РХВ	$p_{obs}$	0,31		0,34		
	$F_{IS}$	-0,04	0,09 0,25	-0,21	-0,08 0,08	
	$BF_{01}$	5,1		3,9		
	$P_{post}$	0,77		0,73		
<i>Проверка однородности распределений и статистические предсказания</i>						
Точное $p_{obs}$ с 95 % ПИ		Значение $p$ с поправкой на множественность, $P_s$	Бейзов фактор, $BF_{01}$	Коэффициент сопряженности $\phi$ с 95 % ДИ	Нижняя граница для апостериорной вероятности $P(H_0 D)$	Достигнутая мощность, $(1 - \beta)$ с 95 % ДИ
$10^{-6}$ 0,035 0,83		0,16	1,59	0,05 0,14 0,25	> 0,32	0,13 0,74 0,99

Примечание. Использована компактная форма записи, при которой слева и справа от точечных оценок параметров в виде подстрочных индексов указаны границы 95 % ДИ [24]. Жирным шрифтом выделены значения, которые обычно (без поправки на множественность сравнений) принято считать статистически значимыми на уровне значимости  $\alpha = 0,05$ . Использованы программы: точные критерии для проверки согласия с РХВ и для проверки однородности в таблицах сопряженности 2x3. <https://www.cog-genomics.org/software/stats>; HWMET — бейзовское оценивание популяционного коэффициента инбридинга [10]. <http://www.reading.ac.uk/Statistics/genetics/software.html>; HWEBayes — бейзов фактор  $BF_{01}$  для РХВ. <https://cran.r-project.org/web/packages/HWEBayes/index.html>; HardyWeinberg — апостериорная вероятность гипотезы о РХВ. <https://cran.r-project.org/web/packages/HardyWeinberg/index.html>; JASP — бейзовский анализ таблиц сопряженности. <https://jasp-stats.org/>; StatXact -8 — доверительные интервалы для коэффициента сопряженности. <http://www.cytel.com/software/statxact>; G\*Power — анализ статистической мощности [17]. <http://www.gpower.hhu.de/>; pwr — основные функции для анализа мощности. <https://cran.r-project.org/web/packages/pwr/index.html>; MOVER-D.xls — ДИ для разности долей [29]. <http://medicine.cf.ac.uk/primary-care-public-health/resources/>; калькулятор предсказательных интервалов для P-значений [23]. <http://www.nature.com/mp/journal/v19/n12/extref/mp2013184x1.xls>; LePrep — вероятности воспроизведения. <http://lmrs.univ-rouen.fr/Persopage/Lecoutre/PAC.htm>.

дование с такими же объемами выборок, то новое значение  $\rho_{obs}$  может получиться как очень малым ( $10^{-6}$ ), так и очень большим (0,83). Об отсутствии сколько-нибудь существенного различия между сравниваемыми группами свидетельствуют также оценки нижней границы для апостериорной вероятности нулевой гипотезы  $P(H_0|D) > 0,32$  и интервальная оценка достигнутой мощности использованного критерия. Нижняя граница мощности  $(1-\beta)_L = 0,18$  далека от приемлемого для мощности значения 0,8.

Таким образом, все эффекты, сообщаемые в обсуждаемой работе, оказываются статистически незначимыми.

К этому следует добавить, что выборка людей, не находившихся в Ленинграде в годы блокады, не может служить («чистой») контрольной группой. Если бы они подверглись воздействию условий блокады, то невозможно заведомо сказать, кто из них и с какими генотипами выживет. Это означает, что схема (дизайн) эксперимента не является схемой «случай—контроли» (case—control), а схемой, которую принято называть «случай—когорта» (case—cohort, синонимы: case—base, case—population). По существу, эта когорта является смесью субъектов, потенциально выживших и погибших, и их соотношение в такой смеси невозможно оценить, поскольку для их различения нужен некий «диагностический тест». Поэтому, строго говоря, контролем должны были быть образцы ДНК ленинградцев, скончавшихся от голода в блокаду. Возникает неразрывный порочный круг, и поставленная авторами задача фактически заведомо не имеет решения.

Корректный гармонизированный статистический анализ результатов работы И.В.Поляковой и соавт. убедительно свидетельствует, что, вопреки мнению авторов, распределения генотипов у жителей блокадного Ленинграда и у жителей Северо-Западного региона России статистически неразличимы по всем пяти изученным генам. Основными причинами ошибочных выводов авторов являются пренебрежение проблемой множественных сравнений и принципиальная невозможность сформировать адекватную контрольную группу.

## Литература

1. Колмогоров А.Н. Вероятность // В кн.: Большая советская энциклопедия (в 51-м т.). Т. 7 / Под ред. С.И.Вавилова (до 1951 г.), Б.А.Введенского (с 1951 г.), 2-е изд. (1949–1968). М.: БСЭ, 1951.

2. Колмогоров А.Н. Вероятность // В кн.: Большая советская энциклопедия (в 30-ти т.). Т. 5 / Под ред. А.М.Прохорова, 3-е изд. (1969–1978). М.: Сов. энцикл., 1971.

3. Колмогоров А.Н. Вероятность // В кн.: Математическая энциклопедия (в 5-ти т.) Т. 1 / Под ред. И.М.Виноградова (1977–1978). М.: Сов. энцикл., 1977.

4. Колмогоров А.Н. Вероятность // В кн.: Вероятность и математическая статистика: Энциклопедия. М.: БСЭ, 1999; М.: Дрофа, 2003. С. 97.

5. Колмогоров А.Н. Вероятность // В кн.: Большая российская энциклопедия (в 35-ти т.) Т. 5 / Отв. ред. С.Л.Кравец (2004–2015). М.: БРЭ, 2006.

6. Колмогоров А.Н. Теория вероятностей: математика, её содержание, методы и значение (в 3-х т.) / Под ред. А.Д.Александрова и др. М., 1956. Т. 2. Гл. 11.

7. Колмогоров А.Н. Теория вероятностей: вероятность и математическая статистика: Энциклопедия / Под ред. Ю.В.Прохорова. М.: БРЭ, 1999; М.: Дрофа, 2003. С. 874–875.

8. Полякова И.В., Готов О.С., Хорошинина Л.П. и др. Анализ частот генотипов и аллелей генов семейств UCP и PPAR в группе жителей блокадного Ленинграда и контрольной группе // Успехи геронтол. 2014. Т. 27. № 3. С. 418–424.

9. Рубанович А.В., Хромов-Борисов Н.Н. Теоретический анализ показателей предсказательной эффективности бинарных генетических тестов // Экол. генетика. 2013. Т. 11. № 1. С. 77–90.

10. Ayres K.L., Balding D.J. Measuring departures from Hardy-Weinberg: a Markov chain Monte Carlo method for estimating the inbreeding coefficient // Heredity. 1998. Vol. 80. Pt. 6. P. 769–777.

11. Balding D.J. A tutorial on statistical methods for population association studies // Nat. Rev. Genet. 2006. Vol. 7. № 10. P. 781–791.

12. Burton P.R., Clayton D.G., Cardo L.R. et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls // Nature. 2007. Vol. 447. № 7145. P. 661–678.

13. Chanoock S.J., Manolio T., Boehnke M. et al. Replicating genotype-phenotype associations // Nature. 2007. Vol. 447. № 7145. P. 655–660.

14. Clarke G.M., Anderson C.A., Pettersson F.H. et al. Basic statistical analysis in genetic case-control studies // Nat. Protoc. 2011. Vol. 6. № 2. P. 121–133.

15. Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values // R. Soc. Open Sci. 2014. Vol. 1. № 1. P. 140216.

16. Cumming G. Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better // Perspect. Psychol. Sci. 2008. Vol. 3. № 4. P. 186–300.

17. Faul F., Erdfelder E., Lang A.-G., Buchner A. G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences // Behav. Res. Methods. 2007. Vol. 39. № 2. P. 175–191.

18. Goodman S.N. A comment on replication, p-values and evidence // Stat. Med. 1992. Vol. 11. № 7. P. 875–879.

19. Johnson V.E. Revised standards for statistical evidence // PNAS. 2013. Vol. 110. № 48. P. 19313–19317.

20. Kolmogorov A.N. The theory of probability: Mathematics: Its Content, Methods, and Meaning / Ed. by A.D.Aleksandrov et al. The MIT Press. Massachusetts Institute of Technology, Cambridge, Massachusetts. 1963. Vol. 2. Ch. 11. P. 240–241. Dover Publications, Inc. Mineola, New York, 1999.

21. Kolmogorov A.N. Probability: Great Soviet encyclopedia (31 vols.) / Ed. by A.M. Prokhorov, J. Paradise. New York, London: Macmillan, Collier-Macmillan, 1973–1983.

22. Kolmogorov A.N. Probability: Encyclopedia of Mathematics wiki. Online: [https://www.encyclopediaofmath.org/index.php/Main\\_Page](https://www.encyclopediaofmath.org/index.php/Main_Page)

23. *Lazzeroni L.C., Lu Y., Belitskaya-Lévy I.* P-values in genomics: Apparent precision masks high uncertainty // *Mol. Psychiat.* 2014. Vol. 19. № 12. P. 1336–1340.
24. *Louis T.A., Zeger S.L.* Effective communication of standard errors and confidence intervals // *Biostatistics.* 2009. Vol. 10. № 1. P. 1–2.
25. *Motulsky H.J.* Common misconceptions about data analysis and statistics // *J. Pharmacol. Exp. Ther.* 2014. Vol. 3. № 1. P. 200–205.
26. *Motulsky H.J.* Common misconceptions about data analysis and statistics // *Naunyn Schmiedeberg's Arch. Pharmacol.* 2014. Vol. 387. № 11. P. 1017–1023.
27. *Motulsky H.J.* Common misconceptions about data analysis and statistics // *Pharmacol. Res. Perspect.* 2015. Vol. 3. № 1. P. e00093.
28. *Motulsky H.J.* Common misconceptions about data analysis and statistics // *Brit. J. Pharmacol.* 2015. Vol. 172. № 8. P. 2126–2132.
29. *Newcombe R.G.* Confidence intervals for proportions and related measures of effect size. Boca Raton, Florida: CRC Press/Taylor & Francis. 2012.
30. *Oleksyk T.K., Brukhin V., O'Brien S.J.* Putting Russia on the genome map // *Science.* 2015. Vol. 350. № 6262. P. 747.
31. *Rubanovich A.V., Khromov-Borisov N.N.* Theoretical analysis of the predictability indices of the binary genetic tests // *Rus. J. Genet: Appl. Res.* 2014. Vol. 4. № 2. P. 146–158.
32. *Sellke T., Bayarri M.J., Berger J.O.* Calibration of  $p$  values for testing precise null hypotheses // *Amer. Stat.* 2001. Vol. 55. № 1. P. 62–71.
33. *Shafer G., Shen A., Vereshchagin N., Vovk V.* Test martingales, Bayes factors and  $p$ -values // *Stat. Sci.* 2011. Vol. 26. № 1. P. 84–101.
34. *Smirnova Y., Davydova A.* Did good genes help people outlast brutal Leningrad siege? // *Science.* 2015. Vol. 348. № 6239. P. 1068.
35. *Stephens M., Balding D.J.* Bayesian statistical methods for genetic association studies // *Nat. Rev. Genet.* 2009. Vol. 10. № 10. P. 681–690.
36. *Sterne J.A.C., Davey Smith G.* Sifting the evidence—what's wrong with significance tests? // *Brit. med. J.* 2001. Vol. 322. № 7280. P. 226–31.
37. *Vovk V.G.* A logic of probability, with application to the foundations of statistics // *J.R. Stat. Soc. Series B «Stat. Methodol.».* 1993. Vol. 55. № 2. P. 317–351.

Adv. geront. 2016. Vol. 29. № 3. P. 454–460

*N.N. Khromov-Borisov*

**CORRECT STATISTICAL ANALYSIS OF GENOTYPE FREQUENCIES FOR UCP AND PPAR GENE FAMILIES IN RESIDENTS OF BESIEGED LENINGRAD AND THE CONTROL GROUP**

R. R. Vreden Russian Research Institute of Traumatology and Orthopedics, 8, ul. akad. Baykova, St. Petersburg 195427; e-mail: Nikita.KhromovBorisov@gmail.com

Correct harmonized statistical re-analysis of the data published in this Journal by I. V. Polyakova et al. (2014) clearly shows that, contrary to the authors' opinion, the distribution of genotypes among residents of besieged Leningrad and the residents of the North-West region of Russia appeared to be statistically indistinguishable in all five genes studied. The main causes of the erroneous conclusions of the authors are neglecting the problem of multiple comparisons and fundamental impossibility of sampling adequate control group. A scheme for harmonized statistical analysis of such data is presented. It implies not only frequentist but Bayesian point and interval estimates for genotype proportions and their differences, for fixation index (coefficient of inbreeding)  $F_{IS}$ , for the effect size  $\phi$  based on  $\chi^2$  statistic (contingency coefficient) and for the achieved power  $(1 - \beta)$ , as well as estimates of posterior probabilities for the null hypothesis  $P(H_0|D)$ , Bayes factors  $BF_{01}$ , observed  $p$ -values,  $p_{obs}$ , with the prediction intervals, and  $p$ -values adjusted for the multiplicity of null hypotheses tested ( $P_g$ ).

**Key words:** survivors of Leningrad siege, genotype frequencies, statistical analysis, fixation index, effect size, statistical power, posterior probabilities, prediction intervals, adjustment for the multiplicity of comparisons, Bayes factor