

Барникова, И.Э. Использование информационных технологий для оценки размера эффекта в биомеханических исследованиях // Труды кафедры биомеханики Университета имени П.Ф. Лесгафта, 2017, Вып. XI. – С. 6-11.

УДК 004+51-37+577.35

## ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ ДЛЯ ОЦЕНКИ РАЗМЕРА ЭФФЕКТА В БИОМЕХАНИЧЕСКИХ ИССЛЕДОВАНИЯХ

**Барникова Ирина Эдуардовна**, к.п.н., доцент  
Национальный государственный Университет физической культуры,  
спорта и здоровья имени П.Ф. Лесгафта, Санкт-Петербург, кафедра био-  
механики

**Аннотация.** Значение статистики *p-value*, являющееся одним из результатов проверки статистических гипотез, часто интерпретируется как показатель практической значимости и ценности проведенного исследования. Как свидетельствуют результаты обсуждений специалистов в области статистики в иностранной научной литературе, вместе со значением *p-value* необходимо приводить также размер эффекта (ES) в качестве показателя практической значимости. Для биомеханики важность представления оценки размера эффекта объясняется малыми объемами выборок проводимых исследований, в связи с чем показатели статистических тестов могут искажаться. Применение информационных технологий в виде онлайн калькуляторов помогает в расчете наиболее часто используемых индексов определения размера эффекта.

**Ключевые слова:** размер эффекта, проверка статистических гипотез, информационные технологии, *p-value*, индекс *d* – Козна, биомеханические исследования.

## USES OF INFORMATION TECHNOLOGIES FOR EFFECT SIZE ESTIMATION IN BIOMECHANICAL RESEARCH

**Irina E. Barnikova**, PhD, Associate Professor  
Lesgaft National State University of Physical Culture, Sports and Health, St. Petersburg, Department of Biomechanics

**Abstract.** The *p-value* statistics, as one of the statistical hypotheses testing results, often interpret as the practical significance indicator and value of the study. As evidenced by the results of the statisticians' discussions in foreign scientific literature it is also should be present the effect size (ES) as an indicator of practical significance. The importance of effect size estimation in the field of biomechanics explained by the small samples size of the conducted studies, and, therefore, the indicators of statistical tests can be distort. Uses of information technology in the form of online calculators helps in calculating the most commonly used the effect size indices determining.

**Keywords:** effect size, statistical hypothesis testing, information technology, *p-value*, Cohen's *d*, biomechanical research.

## ВВЕДЕНИЕ

Для подтверждения теоретической и практической значимости в большинстве научных исследований на протяжении многих лет используется метод проверки статистических гипотез с применением специальных критериев. Очень часто к «решающим» факторам, разделяющим результаты исследований на «значимые» и «не значимые», относится значение статистики *p-value*. Статистическая значимость при этом рассматривается как эквивалент важности полученных результатов и силы научных доказательств в пользу практического и теоретического эффекта исследования. Рост конкуренции в научных кругах привел к тому, что в литературе значительно вырос процент публикаций с положительными статистически значимыми результатами. Это стало следствием почти полного исчезновения отрицательных результатов в целом ряде стран и научных направлений [10, 11, 21]. Частое некорректное использование и злоупотребление применением статистики *p-value*, невозможность воспроизведения полученных в исследованиях результатов в отдельных случаях поставило под сомнение применение статистики вообще. Дело дошло даже до того, что некоторые научные журналы не принимали к изданию статьи, в которых фигурировали значения *p-value* [4, 27, 29]. С целью исправления сложившейся ситуации, а также для повышения качества и переоценки статистической практики, применяемой в анализе научных работ, был предпринят ряд мер по ее улучшению.

### **Проблемы *p-value***

Обсуждение проблем применения метода проверки статистических гипотез и статистики *p-value* в зарубежных научных кругах продолжается уже многие годы [6, 14-16, 24, 25].

К наиболее частым проблемам статистики *p-value* относят ее зависимость от размера выборки и разброса исследуемых данных, а также некорректное применение статистического значения *p* в качестве оценки практической значимости результатов исследований. Невозможность сравнения величины полученных значений *p* между собой также является одной из проблем, поскольку разница между существенным значением *p-value* и явно несущественным значением сама по себе не обязательно статистически значима [8, 23, 24, 26].

7 марта 2016 года впервые за 177 лет своего существования Американская статистическая ассоциация (ASA) опубликовала заявление по результатам обсуждений и работы специальной комиссии, касающееся использования *p-value* в научных исследованиях. В заключительных замечаниях было, в частности, отмечено, что статистическая значимость, представленная значением *p-value*, не измеряет размер практического эффекта или его важность, и что само по себе значение *p* не дает достаточных доказательств относительно модели или гипотезы [28].

В последнее время многие зарубежные научные журналы в качестве подтверждения практической ценности полученных результатов предлагают авторам приводить не только результаты стандартной проверки статистической гипотезы (среднее, ошибку среднего или стандартное отклонение, значение *p-value*), но и оценку *размера эффекта* в абсолютных или стандартизированных единицах, доверительные интервалы, а также мощность критерия.

Следует отметить, что в публикациях российских авторов в области биомеханики спорта оценка размера эффекта практически не встречается. Это во многом связано с тем, что российские ученые не имеют информации об этом показателе и не знают, как его определять.

В связи с этим целью настоящей статьи является обзор зарубежных публикаций по использованию размера эффекта.

### Размер эффекта и его значение

В иностранных источниках *размер эффекта* определяется как количественное отражение величины (степени проявления) некоторого явления, которое используется для решения интересующего вопроса [17]. Размер эффекта (ES – effect size) обозначается аббревиатурой ES.

Следует отметить, что размер эффекта (ES) – обозначение целого семейства индексов, которые оценивают значение эффекта воздействия [4]. В качестве размера эффекта может быть использована абсолютная и стандартизированная разница между средними значениями в исследуемых группах, степень взаимосвязи между признаками, оценка шансов и рисков, оценка пропорций, соответствие или несоответствие определенной модели и т. д. [17].

О необходимости применения размера эффекта для указания практической значимости экспериментальных исследований упоминается в целом ряде научных работ [10, 15, 18, 22, 25], в том числе и в биомеханике спорта [17, 20, 21].

В качестве аргумента в пользу необходимости дополнительного определения размера эффекта авторы называют следующие причины:

1. В отличие от статистической значимости значений *p-value* размер эффекта показывает практическую значимость результатов исследования [25].

2. Представленные в стандартизированном виде результаты исследования позволяют сравнивать между собой признаки, измеренные в различных статистических шкалах [18].

3. Размер эффекта является очень полезным в проведении метаанализа, позволяя сравнивать между собой стандартизированные значения, полученные в исследованиях разных авторов [11, 18].

4. ES предыдущих исследований может быть использован для определения объема выборки и мощности критерия планируемых исследований [4, 11, 18].

5. Представленный вместе с доверительным интервалом ES, также, как и другие числовые характеристики, может использоваться для оценивания генеральных параметров совокупности [11].

6. В биомеханике определение размера эффекта необходимо из-за малых объемов выборок [20, 21].

Выбор типа индексов размера эффекта зависит от статистической шкалы, в которой измерены признаки, использованных в исследовании статистических критериев и дизайна эксперимента [6]. К основным группам (семействам) ES относятся:

- *d* – *семейство*, основанное на различии между средними значениями выборок (*d* – Коэна (Cohen's *d*); *g* – Хеджесса (Hedges's *g*);  $\Delta_s$  – Гласса (Glass's  $\Delta_s$  (delta));
- *r* – *семейство*, основанное на определении взаимосвязи;
- *c* – *семейство*, оценивающее размер эффекта для категориальных данных (оценка шансов, рисков, пропорций).

В свою очередь среди этих групп можно выделить параметрические и непараметрические индексы, стандартизированные и нестандартизированные.

В отдельных случаях возможен альтернативный выбор между ES индексами, поскольку существует возможность конвертации одних индексов в другие.

Наибольшее распространение в качестве оценки размера эффекта получил расчет значения *d* – Коэна (Cohen's *d*). Формула его достаточно проста и значение *d* может быть легко рассчитано без применения сложных

компьютерных статистических программ при известных значениях средних и стандартных отклонений. Так, например, для критерия  $t$  – Стьюдента в случае независимых выборок одинакового объема применяется следующая формула расчета размера эффекта  $d$  – Козна (1):

$$d = \frac{M_1 - M_2}{\sqrt{\frac{S_1^2 + S_2^2}{2}}}, \quad (1)$$

где  $M_1$  – среднее значение первой выборки,  $M_2$  – среднее значение второй выборки,  $S_1$  – стандартное отклонение первой выборки,  $S_2$  – стандартное отклонение второй выборки.

Выражение, стоящее в знаменателе, носит название совместного стандартного отклонения  $S_{pooled}$ .

Следует отметить, что для малых объемов выборок формула расчета  $d$  – Козна применяется с поправкой во избежание переоценивания размера эффекта.

В то же время стоит отметить, что в компьютерных статистических программах, таких, как SPSS и Statgraphics, возможность расчета этого индекса не реализована. В дополнение к этому можно добавить, что существуют онлайн калькуляторы, с помощью которых можно рассчитать значение размера эффекта  $d$  – Козна (Cohen's  $d$ ). Например, такую возможность на русском языке предоставляет ресурс: <https://www.easycalculation.com/ru/statistics/effect-size.php>. [1].

До настоящего времени не существует определенных стандартов, как оценивать величину размера эффекта. J. Cohen (1988) в качестве методических рекомендаций были предложены следующие значения  $d$  для определения величины размера эффекта:  $d=0,2$  считается малым эффектом,  $d=0,5$  – средним,  $d=0,8$  – большим.

Эти значения можно принимать только в качестве ориентиров, так как для каждого отдельного исследования должен применяться свой подход к оценке. Необходимо учитывать теоретическую, экономическую, этическую и практическую сторону в оценке размера эффекта [20]. К примеру, в биомеханических исследованиях элитных спортсменов даже совсем небольшой размер эффекта может быть значительным [21].

## ЗАКЛЮЧЕНИЕ

Необходимость оценки величины размера эффекта (ES) совместно со значением  $p$ -value оправдана с практической и теоретической точки зрения, так как позволит повысить качество представления статистических данных исследований в биомеханике.

## ЛИТЕРАТУРА

1. Калькулятор размера эффекта [Электронный ресурс]. – Режим доступа: <https://www.easycalculation.com/ru/statistics/effect-size.php>. (дата обращения 20.11.2017).
2. Baguley T. Standardized or simple effect size: what should be reported? // British Journal of Psychology. – 2009. – 100 (3). – P. 603-617.
3. Becker L.A. Effect size (ES). 2000. [Digital resource]. – Access mode: <https://www.uccs.edu/lbecker/effect-size.html> (date of the address 19.11.2017).

4. Cohen J. Statistical power analysis for the behavioral sciences. – New York: Lawrence Earlbaum Associates, 1988 – P. 568.
5. Cohen J. The earth Is round ( $p < 05$ ) // American Psychologist. – 1994. – Vol. 49 (12). – P. 997-1003.
6. Ialongo C. Understanding the effect size and its measures // Biochem. Med. (Zagreb). – 2016. – 26 (2). – P. 150-163.
7. Dahiru T. P-value, a true test of statistical significance? // A cautionary note Ann Ib Postgrad Med. – 2008. – Vol. 6 (1). – P. 21-26.
8. Fanelli D. Do pressures to publish increase scientists' bias? An Empirical Support from US States Data // PLoS ONE. – 2010. – Vol. 5 (4). – [Digital resource]. – Access mode: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2858206/> (date of the address 20.11.2017).
9. Fanelli D. Negative results are disappearing from most disciplines and countries // Scientometrics. – 2012. – No. 90. – P. 891-904.
10. Ferguson C.J. An effect size primer: a guide for clinicians and researchers // Professional Psychology: Research & Practice. – 2009. – No. 40. – P. 532-538.
11. Fritz C.O., Morris P.E., Richler J.J. Effect size estimates: current use, calculations, and Interpretation // J. Exp. Psychol. Gen. – 2012. – Vol. 141 (1). – P. 2-18.
12. Gelman A., Loken E. The statistical crisis in science // American Scientist. – 2014. – Vol. 102. – P. 460-465.
13. Goodman S. A dirty dozen: twelve p-value misconceptions // Semin Hematol. – 2008. – Vol. 45 (3). – P.135-140.
14. Goodman S.N. STATISTICS. Aligning statistical and scientific reasoning // Science. – 2016. – Vol. 352 (6290). – P.1180-1181.
15. Kelley K., Preacher K.J. On effect size // Psychol. Methods. – 2012. – Vol.17 (2). – P.137-52.
16. Kline R.B. Beyond significance testing: Reforming data analysis methods in behavioral research. Washington, DC: American Psychological Association, 2004. – P. 325.
17. Knudson D. Significant and meaningful effects in sports biomechanics research // Journal Sports Biomechanics. – 2009. – Vol. 8, Issue 1. – P. 96-104.
18. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs // Front Psychol. – 2013. – No. 4. – P. 863.
19. Matosin N, Frank E., Engel M., Lum J.S., Newell K.A. Negativity towards negative results: a discussion of the disconnect between scientific worth and scientific culture // Disease Models & Mechanisms. – 2014. – Vol. 7 (2). – P. 171-173.
20. Mullineaux D.R., Bartlett R.M., Bennett S. Research design and statistics in biomechanics and motor control // Journal of Sports Sciences – 2001. – Vol. 19, Issue 10. – P. 739-760.
21. Mullineaux D.R Research methods: sample size and variability effects on statistical power / In the book: Biomechanical evaluation of movement in sport and exercise: the British Association of Sport and Exercise Science guide / ed. by C. Payton and R. Bartlett, London: RoutledgeTaylor & Francis Group, 2008 – P. 153-176.
22. Nakagawa S, Cuthill I.C. Effect size, confidence intervals and statistical significance: a practical guide for biologists // Biol. Rev. Camb. Philos. Soc. – 2007. – Vol.

82, No.4. – P. 591-605.

23. Nickerson R.S. Null hypothesis significance testing: a review of an old and continuing controversy // Psychological Methods. – 2000. – Vol. 5 (2). – P. 241-301.

24. Simmons J.P., Nelson L.D., Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant // Psychol. Sci. – 2011. – Vol. 22 (11). – P.1359-1366.

25. Sullivan G.M., Feinn R. Using effect size – or why the P value is not enough // Journal of Graduate Medical Education. – 2012. – Vol. 4. (3). – P. 279-282.

26. Trafimow D., Marks M. Editorial // Basic and Applied Social Psychology. – 2015. – Vol. 37 (1). – P. 1-2.

27. Wasserstein R.L., Lazar N.A. The ASA's statement on p-values: context, process, and purpose // The American Statistician. – 2016 – Vol. 70, Issue 2. – P. 129-133.

28. Woolston C. Psychology journal bans P values. 2015. [Digital resource]. – Access mode: <http://www.nature.com/news/psychology-journal-bans-p-values-1.17001> (date of the address 20.11.2017).