

Верифицируемость данных в исследовании

Балыхин Михаил Григорьевич

*ФГБОУ ВО «Московский государственный университет пищевых производств»
Адрес: 125080, Москва, Волоколамское шоссе, дом 11
E-mail: rektor@mgupp.ru*

Косычева Марина Александровна

*ФГБОУ ВО «Московский государственный университет пищевых производств»
Адрес: 125080, Москва, Волоколамское шоссе, дом 11
E-mail: kosychevama@mgupp.ru*

Рассматривается принцип верифицируемости как основа построения научного исследования. Описываются возможности установления истинности научных утверждений в результате эмпирической проверки. Дается понятие публикационного искажения (publication bias). Приводятся примеры манипуляций с научными данными. Проанализировано создание научных репозиториях как средств длительного хранения, накопления и обеспечения открытого доступа к результатам научных исследований.

Ключевые слова: верифицируемость, публикационное искажение, эмпирическая проверка, проверка гипотез, манипуляции научных данных

Одним из ключевых принципов, на которых строится современная наука, является принцип верифицируемости научных исследований – то есть воспроизводимость исследования, представленного в статье с целью апробации выдвинутых гипотез. Иными словами, верифицируемость как феномен методологии науки делает возможным установление истинности научных исследований в результате их эмпирической проверки¹.

Гипотезы, связанные с данными наблюдения или эксперимента, могут быть проверены напрямую в случае непосредственной верифицируемости, или могут быть косвенно верифицируемы, если речь идет об установлении логических отношений между косвенно верифицируемыми и прямо верифицируемыми утверждениями. В современных научных концепциях верифицируемость является результатом многопланового взаимоотношения между соперничающими теориями и данными их экспериментальных проверок. Однако, не все научные теории могут быть верифицированы. В науке есть примеры того, что некоторые теории, например, теория относительности, не требуют проверки, люди полагаются на нее, потому что она является общепризнанной. В связи с этим, К. Поппер утверждал, что теория не может считаться истинной, если она не поддается фальсификации. То есть теория может быть принята не

потому, что она включает корректные данные, а потому, что она исключает неточности и не демонстрирует искусственности (Поппер, 1983). Таким образом, формулируемые гипотезы должны соответствовать двум принципам – фальсифицируемости (опровержению) и верифицируемости (подтверждению).

Для современного ученого формулирование разного рода предположений в форме гипотез должно сопровождаться доказательством их истинности путем проведения эмпирического исследования. Для этого требуется четкая формулировка гипотез и ясная и структурируемая методология проверки.

Проблема верифицируемости научных данных заключается в том, что практически любая гипотеза, может быть сформулирована таким образом, что она может быть подтверждена. Многие исследователи поддаются искушению привести те результаты исследования, которые ее подтверждают и опустить те, которые будут ее опровергать. В данном случае мы сталкиваемся с таким явлением как публикационное искажение (publication bias) или предвзятость исследователя, когда ученые намеренно выбирают для опубликования положительные или значимые результаты исследования, а отрицательные результаты намеренно умалчивают. Эта проблема особенно актуальна при написании

¹ Швырев. В.С. Верифицируемость // Электронная библиотека ИФ РАН «Новая философская энциклопедия». <https://iphlib.ru/library/el=&a=d&c=newphilenc&d=&rl=1&href=http:%2f%2f0592.html>

обзоров литературы, систематических обзоров и мета-анализов, так как эти виды исследований основываются на уже опубликованных работах и рассматриваются как изначально предвзятые. Следовательно, при описании исследования настоятельно рекомендуют указывать все полученные в его результате данные, независимо от их интерпретации.

Публикационное искажение имеет определенные предпосылки. Рецензируемые научные журналы стремятся представлять исключительно весомые результаты исследований, те, которые помогут журналу получить наибольшее количество цитат на опубликованные статьи, повышая его наукометрические показатели. Как следствие, исследования, результаты которых выглядят противоречиво, могут быть исключены из портфеля журнала. Отсюда и склонность ученых к однозначности представляемых ими данных, их стремление замалчивать «неудобные» результаты.

Наряду с публикационным искажением имеют место и другие виды искажения или предвзятости. Искажение цитирования, например, характеризуется тем, что при отборе литературы для анализа, авторы статьи пользуются лишь списком цитируемых работ из уже опубликованных статей, менее цитируемые работы исключаются из анализа. В этой связи также рассматривают предвзятость распространения результатов исследования: исследователи уделяют внимание преимущественно значимым положительным и легкодоступным результатам, игнорируя источники, которую сложнее найти (например, отчеты правительства или результаты клинических исследований) (Song et al., 2010). Подтасованное, фальсифицированное исследование может иметь высокую статистическую значимость, быть репрезентативным, без методологических недостатков, но, по своей сути, оно окажется бессмысленным мусором, засоряющим поле научного знания.

Последствия таких исследований могут быть разными, но для медицинских исследований они имеют самый негативный эффект, так как напрямую связаны со здоровьем человека. Худшим сценарием может быть случай, когда необоснованное медицинское вмешательство ложно считается эффективным из-за публикационного искажения, что может привести к назначению пациенту неэффективного или вредоносного курса лечения (Song et al., 2010).

Манипуляция научными данными остается актуальной проблемой, имеющей колоссальные

последствия. Широкое распространение получили манипуляции с графическим представлением информации. Визуализация информации упрощает её восприятие, вызывает к ней доверие, усиливая персуазивность ее воздействия на адресата. Например, неверное масштабирование, диаграммы с урезанными столбцами, объемные диаграммы, представление зависимостей между несравнимаемыми величинами могут привести к искажению восприятия даже при наличии верных данных (Шестов, 2017).

Некорректно представленные статистические данные или представление неполного набора данных также являются способом манипуляции данными. Например, когда выборка изначально необъективна, авторы намеренно выбирают среднюю величину, которая поможет представить результаты в выгодном свете или наоборот, скрыть нежелательные.

Верифицируемость данных, отсюда, получает особое звучание в контексте исследования. Рецензенты должны получить возможность убедиться в достоверности представления данных в тексте статьи. Возникает необходимость того, чтобы доступ рецензента и заинтересованных читателей к данным исследований был свободным, гарантировал неизменность их содержания. Исследователи, соответственно, должны получить инструменты оперативного размещения результатов своих научных исследований, их накопления, сохранения и распространения их в научном сообществе.

Цифровые технологии предоставили ученым такой инструментарий: появилась возможность размещения данных исследований в открытом доступе по ссылке на личных сайтах в сети Интернет, на сайтах учебных заведений и исследовательских институтов с целью обеспечения публичности и верифицируемости данных. Но наиболее популярным инструментом представления результатов исследования являются репозитории. Под последними принято понимать «публично доступные открытые архивы информации научных, исследовательских и образовательных организаций, в которых члены сообщества размещают свои опубликованные и подготовленные к печати статьи и другие материалы научно-исследовательской и научно-организационной деятельности» (Рождественская, 2015, с. 86). Основная задача репозиториев – оперативно и широко представлять новые актуальные научные документы и данные.

Представляется, что глобальное распространение сети репозиториев будет способствовать созданию

более стабильной системы для обмена результатами научных исследований. Репозитории создают возможность доступа к результатам исследований по всему миру, и более того, дают возможность учреждению или отдельному ученому участвовать в глобальном обмене данными, таким образом поддерживая функционирование системы научной коммуникации².

Репозиторий обеспечивает доступ к разнообразным ресурсам, среди которых опубликованные статьи, препринты, наборы данных, рабочие документы, изображения, программное обеспечение и т. д. Пользователи репозитория позволяют каждому ученому добавлять информацию к уже имеющимся. Как следствие, появляется возможность объединения данных из разных репозиториях, усиливая, тем самым, доказательную базу проведенных исследований.

Некоторые репозитории позволяют отслеживать статистику всех событий в режиме реального времени - это информация об изменениях, дополнениях, комментировании, аннотировании, рецензировании, доступе к ресурсу, количестве скачиваний и т. д., относительного любого объекта, размещенного в репозитории.

Ряд репозиториях (например, Mendeley Data), позволяет присваивать подгруженным данным doi. Таким образом, владелец данных получает на них зафиксированные цитирования, в случае их использования другими исследователями в своих статьях. Еще один репозиторий, который присваивает цифровой идентификатор объекта, Zenodo является открытым репозиторием, позволяющий исследователям всех дисциплин обмениваться и сохранять результаты своих исследований независимо от их размера и формата. Благодаря бесплатной загрузке и бесплатному доступу, результаты исследований, хранящиеся на Zenodo, будут доступны для обнаружения и использования в течение длительного времени. Репозиторий для результатов исследований figshare – это структурированный архив, в котором ученые могут делать доступными все результаты своих исследований – цитировать, делиться и находить новое. На бесплатном аккаунте можно хранить любое количество публичных данных. Репозитории помогают

оценить авторам свой собственный вклад в науку, накапливая метрические показатели, что является немаловажным фактом оценки доверия того или иного ресурса. COAR (Confederation of Open Access Repositories) - это международная ассоциация, в которую входят 157 членов и партнеров со всего мира, представляющих библиотеки, университеты, исследовательские институты, правительственные фонды и др. COAR объединяет отдельные репозитории и сети репозиториях для наращивания потенциала, согласования политик и практик и выступает в качестве флага для сообщества репозиториях.³ Для российских ученых Сибирское отделение РАН разработало веб-навигатор SciGuide, который помогает вести поиск качественных научных ресурсов⁴.

Таким образом, создание репозиториях, их доступность и рейтингование активностей исследователей в пространстве репозиториях повысят объективность исследований, позволят отслеживать попытки искажения данных. Открытый доступ к репозиториям поможет молодым ученым проводить более тщательный отбор цитируемого материала, а возможность открытой коммуникации в режиме реального времени сделает возможным комментировать результаты исследований, снижая вероятность фальсификации и манипулирования данными исследований.

Литература

- Поппер, К. Р. (1983). *Логика и рост научного знания*. Прогресс.
- Рождественская, М. Ю. (2015). Репозиторий как реализация идей открытого доступа к научным публикациям: подходы к классификации. *Библиосфера*, 2, 86-94.
- Шестов, Б. Н. (2017). Способы манипулятивного воздействия посредством визуализации данных в современных СМИ. *Медиаискон*, 4. Режим доступа: <http://www.mediascope.ru/2377>
- Song, F., Parekh, S., Hooper, L., Loke, Y. K., Ryder, J., Sutton, A. J., Hing, C., Kwok, C. S, Pang, C., & Harvey, I. (2010). Dissemination and publication of research findings: an updated review of related biases. *Health Technology Assessment*. <http://doi.org/10.3310/hta14080>

² Методические рекомендации по разработке репозиториях /под ред. М. Е. Шварцман, 2018. https://www.openrepository.ru/images/docs/Metod_Schwarzman.pdf

³ COAR (Ассоциация репозиториях открытого доступа) начала свою деятельность в 2009 году в рамках проекта по подключению европейских хранилищ, DRIVER. Этот проект выявил необходимость более глобального международного подхода, который бы поддерживал создание сетей для всех хранилищ со всего мира. Таким образом, COAR была основана в сентябре 2009 года с 28 членами. С тех пор COAR значительно расширила свою международную деятельность с членами и партнерами из более чем 50 стран на всех 5 континентах.

⁴ <http://prometeus.nsc.ru/sciguide/page02.ssi>

Research Data Verifiability

Mikhail G. Balykhin

*Moscow State University of Food Production
11 Volokolamskoe highway, Moscow, 125080, Russian Federation
E-mail: rektor@mgupp.ru*

Marina A. Kosycheva

*Moscow State University of Food Production
11 Volokolamskoe highway, Moscow, 125080, Russian Federation
E-mail: kosychevama@mgupp.ru*

The principle of verifiability as the basis for the construction of scientific research is considered. The possibilities of establishing the truth of scientific statements as a result of empirical verification are described. The concept of publication bias is presented. Examples of manipulations with scientific data are given. The creation of scientific repositories as a means of long-term storage, accumulation and ensuring open access to the results of scientific researches is analyzed.

Keywords: verifiability, publication bias, empirical verification, hypothesis testing, manipulation of scientific data

References

- Popper, K. R. (1983). *Logika i rost nauchnogo znaniya* [Logic and the growth of scientific knowledge]. Progress.
- Rozhdestvenskaya, M. Yu. (2015). Repository as an implementation of the ideas of open access to scientific publications: approaches to classification. *Bibliosfera* [Bibliosphere], 2, 86-94.
- Shestov, B. N. (2017). Methods of manipulative influence through data visualization in modern media. *Mediaskop* [Mediascope], 4. Access mode: <http://www.mediascope.ru/2377>
- Song, F., Parekh, S., Hooper, L., Loke, Y. K., Ryder, J., Sutton, A. J., Hing, C., Kwok, C. S, Pang, C., & Harvey, I. (2010). Dissemination and publication of research findings: an updated review of related biases. *Health Technology Assessment*. <http://doi.org/10.3310/hta14080>