

ПОДХОДЫ К ФОРМИРОВАНИЮ ВЫБОРКИ ДЛЯ КОНТРОЛЯ КАЧЕСТВА РАБОТЫ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

DOI: 10.17691/stm2023.15.2.02

УДК 004.891.3:311

Поступила 02.12.2022 г.

© **С.Ф. Четвериков**, к.т.н., начальник сектора разработки систем внедрения медицинских интеллектуальных технологий¹;
К.М. Арзамасов, к.м.н., руководитель отдела медицинской информатики, радиомики и радиогеномики¹;
А.Е. Андрейченко, к.ф.-м.н., ведущий научный сотрудник отдела медицинской информатики, радиомики и радиогеномики¹; руководитель направления искусственного интеллекта²; ведущий научный сотрудник³;
В.П. Новик, научный сотрудник сектора разработки систем внедрения медицинских интеллектуальных технологий¹;
Т.М. Бобровская, младший научный сотрудник сектора разработки систем внедрения медицинских интеллектуальных технологий¹;
А.В. Владзимирский, д.м.н., зам. директора по научной работе¹; профессор кафедры информационных и интернет-технологий⁴

¹Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, ул. Петровка, 24, стр. 1, Москва, 127051;

²ООО «К-Скай», набережная Варкауса, 17, Петрозаводск, Республика Карелия, 185031;

³Национальный исследовательский университет ИТМО, Кронверкский пр., 49, Санкт-Петербург, 197101;

⁴Первый Московский государственный медицинский университет им. И.М. Сеченова (Сеченовский университет), ул. Трубецкая, 8/2, Москва, 119991

Цель исследования — оценка эффективности подходов формирования выборки при проведении периодического контроля качества результатов работы систем искусственного интеллекта (СИИ), применяемых в медико-биологической практике.

Материалы и методы. Проанализированы подходы формирования выборок, основанные на точечной статистической оценке, проверке статистических гипотез, на использовании готовых статистических таблиц, а также варианты подходов, представленных в ГОСТ Р ИСО 2859-1-2007 «Статистические методы. Процедуры выборочного контроля по альтернативному признаку». Рассмотрены варианты формирования выборок разного размера для генеральных совокупностей от 1000 до 100 000 исследований.

Анализ подходов к формированию выборки проводили в рамках эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и их дальнейшего применения в системе здравоохранения г. Москвы.

Результаты. Готовые таблицы имеют конкретные статистические входные данные, что не позволяет считать их универсальным вариантом для медико-биологических исследований. С помощью точечной статистической оценки можно рассчитать выборку исходя из заданных статистических параметров с определенным доверительным интервалом. Данный подход перспективен в случае, когда для исследователя важна лишь ошибка первого рода, а ошибка второго рода является неприоритетной. Использование подхода, основанного на проверке статистических гипотез, дает возможность учитывать ошибки первого и второго рода исходя из заданных статистических параметров. Применение ГОСТ Р ИСО 2859-1-2007 для формирования выборки позволяет использовать готовые значения в зависимости от заданных статистических параметров.

При оценке эффективности изучаемых подходов установлено, что для наших целей оптимальное количество исследований при проведении контроля качества СИИ для анализа медицинских изображений составляет 80 единиц. Это удовлетворяет требованиям репрезентативности, баланса рисков потребителя и поставщика услуг СИИ, а также оптимизации трудозатрат сотрудников, вовлеченных в процесс контроля качества результатов работы СИИ.

Ключевые слова: искусственный интеллект; статистические методы; формирование выборки; контроль качества СИИ.

Как цитировать: Chetverikov S.F., Arzamasov K.M., Andreichenko A.E., Novik V.P., Bobrovskaya T.M., Vladzimirsky A.V. Approaches to sampling for quality control of artificial intelligence in biomedical research. *Sovremennye tehnologii v medicine* 2023; 15(2): 19, <https://doi.org/10.17691/stm2023.15.2.02>

Для контактов: Арзамасов Кирилл Михайлович, e-mail: ArzamasovKM@zdrav.mos.ru

Approaches to Sampling for Quality Control of Artificial Intelligence in Biomedical Research

S.F. Chetverikov, PhD, Head of the Sector of the Development of Systems for the Implementation of Intelligent Medical Technologies¹;

K.M. Arzamasov, MD, PhD, Head of the Department of Medical Informatics, Radiomics and Radiogenomics¹;

A.E. Andreichenko, PhD, Leading Researcher, Department of Medical Informatics, Radiomics and Radiogenomics¹; Head of the Group of Artificial Intelligence²; Leading Researcher³;

V.P. Novik, Researcher, Sector of the Development of Systems for the Implementation of Intelligent Medical Technologies¹;

T.M. Bobrovskaya, Junior Researcher, Sector of the Development of Systems for the Implementation of Intelligent Medical Technologies¹;

A.V. Vladzimirsky, MD, DSc, Deputy Director for Research¹; Professor, Department of Information and Internet Technologies⁴

¹Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, 24/1 Petrovka St., Moscow, 127051, Russia;

²K-SkAI LLC, 17 Naberezhnaya Varkausa, Petrozavodsk, the Republic of Karelia, 185031, Russia;

³ITMO National Research University, 49 Kronverksky Pr., Saint Petersburg, 197101, Russia;

⁴First Moscow State Medical University named after I.M. Sechenov (Sechenov University), 8/2 Trubetskaya St., Moscow, 119991, Russia

The aim of the study is to evaluate the efficacy of approaches to sampling during periodic quality control of the artificial intelligence (AI) results in biomedical practice.

Materials and Methods. The approaches to sampling based on point statistical estimation, statistical hypothesis testing, employing ready-made statistical tables, as well as options of the approaches presented in GOST R ISO 2859-1-2007 "Statistical methods. Sampling procedures for inspection by attributes" have been analyzed. We have considered variants of sampling of different sizes for general populations from 1000 to 100,000 studies.

The analysis of the approaches to sampling was carried out as part of an experiment on the use of innovative technologies in computer vision for the analysis of medical images and their further application in the healthcare system of Moscow (Russia).

Results. Ready-made tables have specific statistical input data, which does not make them a universal option for biomedical research. Point statistical estimation helps to calculate a sample based on given statistical parameters with a certain confidence interval. This approach is promising in the case when only a type I error is important for the researcher, and a type II error is not a priority. Using the approach based on statistical hypothesis testing makes it possible to take account of type I and II errors based on the given statistical parameters. The application of GOST R ISO 2859-1-2007 for sampling allows using ready-made values depending on the given statistical parameters.

When evaluating the efficacy of the studied approaches, it was found that for our purposes, the optimal number of studies during AI quality control for the analysis of medical images is 80 items. This meets the requirements of representativeness, balance of the risks to the consumer and the AI service provider, as well as optimization of labor costs of employees involved in the process of quality control of the AI results.

Key words: artificial intelligence; statistical methods; sampling; AI quality control.

Введение

Системы искусственного интеллекта (СИИ) в медицине — это инструменты, позволяющие автоматизировать такие рутинные процессы, как заполнение электронных карт, анализ и диагностика медицинских изображений, аналитика и составление планов лечения пациентов. СИИ помогают уменьшить трудозатраты в процессе медико-биологической деятельности, а также повысить точность рекомендаций, диагностики, назначений и т.д. [1–7].

При внедрении СИИ в клиническую практику важной частью является контроль качества результатов работы данных систем [8–10] с целью подтверждения их безопасности и эффективности [11–20]. Для автономной и адаптивной СИИ особую ценность представляет оперативный периодический контроль качества, который позволяет с минимальными рисками и малым временем реагирования скорректировать ее работу.

Все виды дефектов в работе СИИ условно можно разделить по способу обнаружения на две группы: выявляемые при автоматизированной и ручной

проверке. Автоматизированная проверка позволяет проводить контроль качества всей генеральной совокупности за определенный период. Ручная проверка осуществляется на ограниченной выборке и требует существенных ресурсных затрат, так как подразумевает под собой открытие и просмотр изображений, находящихся в исследовании; клинический анализ оригинального изображения; оценку результата работы СИИ; ведение документации по контролю качества и т.д. В связи с этим вопрос формирования выборки для проведения контроля качества результатов работы СИИ является актуальным и должен решать как минимум две важные задачи:

1) репрезентативность и правильность долей распределения (содержания) исследуемого признака [21–24];

2) учет трудозатрат сотрудников, вовлеченных в проведение ручного контроля качества результатов работы СИИ.

Предложено множество подходов для расчета объемов выборок в различных областях науки [25–43], однако осуществить аргументированный выбор того или иного подхода при планировании медико-биологического исследования или при внедрении СИИ для практического применения не представляется возможным. Так, в работе [37] сообщается, что в зависимости от используемого подхода получаются различные объемы выборок.

Цель исследования — оценка эффективности подходов формирования выборки при проведении периодического контроля качества результатов работы систем искусственного интеллекта, применяемых в медико-биологической практике.

Материалы и методы

Рассмотрены подходы формирования выборки, основанные на точечной статистической оценке [31, 44–47]; на проверке статистических гипотез — вариант 1 [48–51] и вариант 2 [52, 53]; на применении ГОСТ Р ИСО 2859-1-2007 «Статистические методы. Процедуры выборочного контроля по альтернативному признаку» [54–58].

Анализ подходов к формированию выборки проводили в рамках эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и их дальнейшего применения в системе здравоохранения г. Москвы [59, 60]. Ранее были обнаружены различные виды дефектов в результатах обработки медицинских изображений СИИ [59], которые снижают клиническую и диагностическую ценность изучаемых нами систем.

В данной статье набор медицинских исследований за определенный период, которые обработаны СИИ, мы будем называть партией. Каждая партия состоит из единиц продукции одного вида, класса, типоразмера и состава, обработанных в практически одинаковых

условиях в один и тот же период времени. Единица продукции — это исследование, обработанное СИИ (подключенное к радиологической информационной системе «Единый радиологический информационный сервис города Москвы» (ЕРИС)) за определенный отчетный период.

Каждое исследование содержит в себе следующие сведения:

популяционные параметры (гендерно-возрастные показатели, этнический состав, регионы проживания и т.д.); сведения о деперсонализации; сведения о медицинских организациях, послуживших источниками для формирования набора данных;

характеристику исследований (анатомическая область(и); модальность; проекции; типы медицинских изделий — диагностических приборов; виды и характеристики протоколов исследований);

целевую патологию в соответствии с международной классификацией болезней [61];

случаи наличия/отсутствия патологических находок.

Контроль качества результатов работы СИИ проводится периодически. При контроле качества результатов работы изучаемых нами СИИ осуществляется ручной просмотр исследований экспертами. По причине большого количества (более 1000) исследований в партии отсутствует возможность контроля качества в полном объеме из-за ограничений во времени, а также в связи с небольшим количеством экспертов. В рамках эксперимента [59] было установлено, что не более 10% дефектных единиц продукции содержится в генеральной совокупности. Это означает, что вся партия за отчетный период содержит не более 10% единиц продукции с технологическими дефектами [62]. Таким образом, в статье описываются подходы, соответствующие качественному признаку, где известна доля случаев, в которых встречается изучаемый признак. Объем генеральной совокупности находится в пределах от 1000 до 100 000 единиц продукции.

Решение задачи контроля качества партии с результатами работы СИИ проводили на основе выборочного наблюдения, которое базировалось на понятиях генеральной и выборочной совокупности (выборка).

В данной статье описана серийная бесповторная выборка, которая характеризовалась тем, что выбранная единица продукции отбиралась из всего объема генеральной совокупности и не возвращалась обратно, таким образом вероятность попадания остальных единиц продукции возрастала.

Расчеты выполняли при помощи программного обеспечения PASS 15 Update – NCSS (www.ncss.com) и LibreOffice Calc (www.libreoffice.org).

Результаты и обсуждение

Подход, основанный на точечной статистической оценке, учитывает отклонение результатов выборочного исследования от генеральных значений.

При данном подходе объем выборки рассчитывается по формуле:

$$n = \frac{Nt^2wq}{N\Delta^2 + t^2wq},$$

где n — объем выборки; N — объем генеральной совокупности; t — коэффициент, показывающий, с какой вероятностью (надежностью) можно гарантировать достоверность полученного результата или критическое значение критерия Стьюдента при соответствующем уровне значимости (для уровня значимости 0,05 коэффициент $t=1,96$); Δ — предельная ошибка показателя; w — доля изучаемого признака; $q=(1-w)$ — доля, где изучаемый признак отсутствует.

Таким образом, при доле изучаемого признака (w) 0,9, уровне статистической значимости 0,95 и предельно допустимой ошибке (Δ) 0,05 мы получаем объем выборки (n), равный 138 единицам продукции.

Подход, основанный на проверке статистических гипотез (вариант 1), предполагает проверку статистической гипотезы H_0 (качество партии соответствует предъявляемым требованиям) при наличии альтернативной гипотезы H_1 (качество партии не соответствует предъявляемым требованиям). Если среди n -единиц продукции число дефектных (m) не превышает приемочное число c ($m \leq c$) (максимально допустимое количество забракованных единиц выборки, позволяющее принять решение о приемке товарной партии по качеству), то партия принимается; в противном случае — бракуется.

Для выбора плана контроля (определения выборки) используется формула:

$$p_n(m \leq c) = \sum_{m=0}^c p_n(m),$$

где m — число дефектных единиц продукции в выборке n ; $p_n(m)$ — вероятность появления дефектных единиц продукции m в выборке n ; c — приемочное число.

Так как в рамках эксперимента [59] объем всей партии превышал объем выборки более чем на 10% [56], то оперативные характеристики определяли по формуле:

$$p_n = C_n^m q^m (1-q)^{n-m},$$

где C_n^m — количество сочетаний появления дефектных единиц продукции m в выборке n :

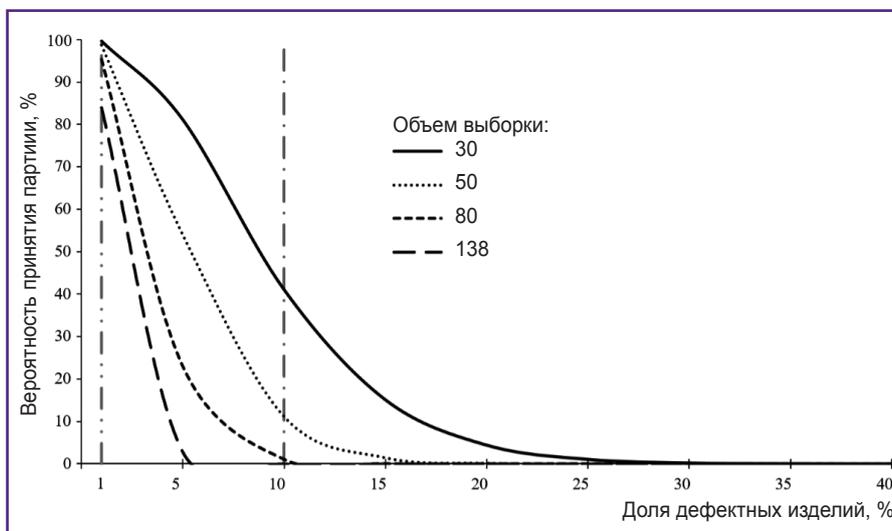
$$C_n^m = \frac{n!}{m!(n-m)!}.$$

На примере эксперимента [59] было использовано приемочное число, равное двум единицам продукции, произведены расчеты и построены кривые для выборок в 30, 50, 80, 138 единиц продукции (см. рисунок).

На рисунке обозначены риски потребителя и поставщика. Риск поставщика — вероятность отклонить партию хорошего качества (т.е. в генеральной совокупности удельный вес дефектных единиц продукции менее 10%). Учитывая долю заявленных дефектных изделий от поставщика, риск принимаем равным 1%. Риск потребителя — вероятность принять некачественную партию. Учитывая долю дефектных изделий, определенных потребителем, риск принимаем равным 10%.

Анализируя данные табл. 1 и учитывая заданные риски потребителя и поставщика на уровне не более 10% и не более 5% соответственно, мы установили, что объем выборки, равный 80 единицам продукции, удовлетворяет требованиям как со стороны потребителя, так и поставщика.

Подход, основанный на проверке статистических гипотез (вариант 2), базируется на принципах вероятности отклонения нулевой гипотезы; учитывает риски поставщика и потребителя. Нулевая гипотеза H_0 предполагает, что если в генеральной совокупности содержится более 10% дефектных единиц продукции, то вся партия за отчетный период содержит более 10% единиц продукции с технологическими дефек-



Оперативная характеристика для различных объемов выборки: вертикальная штрихпунктирная линия с двумя точками — риск поставщика; вертикальная штрихпунктирная линия с одной точкой — риск потребителя

Таблица 1
Зависимость рисков от объема выборки

Объем выборки	Риск потребителя, %	Риск поставщика, %
30	41	0
50	11	1
80	1	5
138	0	16

Таблица 2
Зависимость рисков от объема выборки и приемочного числа

Объем выборки	Приемочное число/доля дефектов, %	Вероятность отклонения нулевой гипотезы, %	Риск поставщика, %	Риск потребителя, %
30	0/0	100,0	0	8
30	1,0/3,3	36,2	63,8	37
50	0/0	100,0	0	1
50	1/2	36,4	63,6	7
80	0/0	100,0	0	<1
80	1,0/1,3	92,1	7,9	<1
80	2,0/2,5	67,7	32,3	2
120	0/0	100,0	0	<1
120	1,0/0,8	100,0	0	<1
120	2,0/1,7	98,4	1,6	<1
120	3,0/2,5	91,9	8,1	<1
120	4,0/3,3	78,8	21,2	2

тами. Соответственно, при альтернативной гипотезе H_1 — менее 10% единиц продукции с технологическими дефектами. Вероятность отклонения нулевой гипотезы — не менее 80%.

Выполнены расчеты (табл. 2) для выборок в 30, 50, 80, 120 единиц продукции с приемочным числом от нуля до четырех (приемочное число ограничивалось превышением рисков потребителя более 10% или поставщика — более 5%).

Анализируя данные табл. 2 и учитывая заданные риски потребителя и поставщика, а также долю заявленных дефектных изделий от поставщика (1%) и долю дефектных изделий, определенных потребителем (10%), установлено, что объем выборки, равный 30, 50, 80 и 120 единиц продукции, удовлетворяет требованиям со стороны как потребителя, так и поставщика при приемочном числе, равном нулю единиц продукции. С учетом доли дефектных единиц продукции при приемочных числах больше нуля наиболее подходящие объемы выборок равнялись 80 или 120 единицам.

Подход, основанный на применении ГОСТ Р ИСО 2859-1-2007 «Статистические методы. Процедуры выборочного контроля по альтернативному признаку», предполагает определение плана выборочного контроля последовательных партий на основе приемлемого уровня качества и может быть применен к различного рода данным или записям. Приемлемый уровень качества выражается в проценте несоответствующих единиц продукции или числе несоответствий на сто единиц продукции.

Нами было рассмотрено несколько вариантов формирования объемов выборок.

Сначала мы обратились к таблице «Коды объема выборки» [54]. В нашем случае общий уровень контроля равен II, специальный уровень контроля не используется. Так как объемы партий у нас находились в пределах от 1000 до 100 000, то для нас были интересны следующие коды: J, K, L, M. В данной работе план не имел многоступенчатости и не подразумевал переход на ослабленный или усиленный контроль. В связи с этим мы использовали данные из таблицы «Одноступенчатые планы при нормальном контроле (основная таблица)» [54]: для приемлемого уровня качества потребителя в 10% (для партий объемом от 501 до 10 000 исследований) объем выборки для контроля качества будет равен 125 единицам продукции с приемочным числом партии, равным нулю; для партий объемом от 10 001 до 150 000 объем выборки для контроля качества будет равен 500 единицам продукции с приемочным числом партии, равным единице.

Затем мы обратились к таблице «Риск изготовителя при нормальном контроле (процент непринятых партий для одноступенчатых планов)» [54] и получили риски поставщика 11,8% для выборки в 125 единиц продукции; 9,02% — для выборки в 500 единиц.

В табл. 3 приведена сводная информация по плюсам и минусам рассматриваемых подходов.

Рассмотренные нами подходы расчета объема выборки обладают рядом преимуществ перед широко распространенными подходами, основанными на фиксированных или табличных значениях. Так, например, подходы с применением готовых таблиц имеют конкретные статистические входные данные, что не позволяет считать их универсальными [46, 63, 64].

Таблица по методике В.И. Паниотто [47] содержит значения, которые рассчитаны для конкретных параметров: доля признака — 0,5; допустимая ошибка — 0,05; доверительная вероятность — 0,954 ($t=2$).

Таблица по методике N. Fox [65] также содержит конкретный параметр: доля признака — 0,35.

Значения из таблиц [47, 65] были рассчитаны по формулам точечной статистической оценки, описанной в данной статье выше. Если входные данные медико-биологического исследования не совпадают с параметрами таблиц, рекомендуем проводить расчеты, а не использовать указанные таблицы для определения объема выборки.

Таблица 3

Плюсы и минусы подходов формирования выборки

Особенности	Подход, основанный на точечной статистической оценке	Подход, основанный на проверке статистических гипотез (вариант 1)	Подход, основанный на проверке статистических гипотез (вариант 2)	Подход, основанный на применении ГОСТ Р ИСО 2859-1-2007
Плюсы	Позволяет определить объем выборки, используя статистические параметры Простые математические расчеты Возможность выбора доверительного интервала, достаточного для уверенного суждения о генеральных параметрах на основании известных выборочных показателей	Учитывает риски потребителя и поставщика Позволяет рассчитать риски в зависимости от объема выборки, долей дефектных единиц продукции и приемочного числа Позволяет визуально определить наиболее подходящие параметры для исследовательской работы	Учитывает риски потребителя и поставщика Позволяет рассчитать риски в зависимости от объема выборки, долей дефектных единиц продукции и приемочного числа Учитывает статистические параметры Позволяет рассчитать доверительные интервалы для полученных значений	Учитывает риски потребителя и поставщика Необходимы минимальные предварительные данные о генеральной совокупности Отсутствие необходимости в проведении математических расчетов Отсутствие необходимости использования дополнительного ПО для обработки данных Визуальное определение наиболее подходящих параметров для исследовательской работы Возможность применения различных планов контроля при низком или высоком уровне качества партии Зависимость объема выборки от высоких и низких объемов генеральной совокупности
Минусы	Не учитывается ошибка второго рода Статистические параметры и доверительный интервал в медико-биологических исследованиях имеют ограниченный предел, после которого результаты не имеют статистической значимости Требуется предварительных данных о генеральной совокупности	Требуется предварительных данных о генеральной совокупности Необходима предобработка данных с целью определения используемых формул Требуется фиксированных значений рисков и долей дефектных единиц продукции перед проведением исследовательской работы	Необходимы предварительные данные о генеральной совокупности Требуется фиксированных значений рисков перед проведением исследовательской работы Не учитывает случаи высокой стоимости единицы продукции, используемой для контроля качества Подразумевает использование платного ПО для статистической обработки	Не позволяет более тонко подобрать статистические параметры Не позволяет более точно рассчитать риски потребителя и поставщика Требуется фиксированных значений рисков и долей дефектных единиц продукции перед проведением исследовательской работы
Объем выборки	138	80	80, 120	125, 500

Заключение

Результаты данного исследования дают возможность подобрать наиболее подходящий подход для достижения целей медико-биологических исследований. Использование точечной статистической оценки и подхода, основанного на проверке статистических гипотез, позволяет наиболее гибко рассчитать объемы выборки в зависимости от входных параметров проводимого исследования. Применение ГОСТ Р ИСО 2859-1-2007 для формирования выборки является приоритетным, если эксперимент затрагивает взаимодействие исследователя и сторонней организации; позволяет учитывать риски и ошибки для обеих сторон, вовлеченных в процесс.

Оптимальное количество исследований при проведении контроля качества работы изучаемых нами СИИ для анализа медицинских изображений составляет 80 единиц продукции. Это удовлетворяет требованиям репрезентативности, баланса рисков

потребителя и поставщика услуг СИИ, а также оптимизации трудозатрат сотрудников, вовлеченных в процесс контроля качества результатов работы СИИ.

Финансирование. Данная статья подготовлена авторским коллективом в рамках научно-практического проекта в сфере медицины (№ЕГИСУ: 122112400040-1) «Эталонные наборы данных для устойчивого развития технологий искусственного интеллекта в медицинской диагностике с целью минимизации долгосрочных последствий пандемии коронавирусной инфекции для здоровья населения города Москвы».

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Литература/References

1. Васюта Е.А., Подольская Т.В. Проблемы и перспективы внедрения искусственного интеллекта в медицине.

Государственное и муниципальное управление. Ученые записки 2022; 1: 25–32, <https://doi.org/10.22394/2079-1690-2022-1-1-25-32>.

Vasyuta E.A., Podolskaya T.V. Challenges and prospects for the introduction of artificial intelligence in medicine. *Gosudarstvennoe i municipal'noe upravlenie. Uchenye zapiski* 2022; 1: 25–32, <https://doi.org/10.22394/2079-1690-2022-1-1-25-32>.

2. Ярмухаметов Р.Р. Обзор применений искусственного интеллекта в медицине. *Наукофера* 2020; 12–2: 172–178.

Yarmukhametov R.R. Overview of usages of artificial intelligence in medicine. *Naukosfera* 2020; 12–2: 172–178.

3. Алексеева М.Г., Zubov A.I., Novikov M.Yu. Искусственный интеллект в медицине. *Международный научно-исследовательский журнал* 2022; 7–2: 10–13, <https://doi.org/10.23670/irj.2022.121.7.038>.

Alekseeva M.G., Zubov A.I., Novikov M.Yu. Artificial intelligence in medicine. *Mezhdunarodnyj nauchno-issledovatel'skiy zhurnal* 2022; 7–2: 10–13. URL: <https://research-journal.org/archive/7-121-2022-july/artificial-intelligence-in-medicine>, <https://doi.org/10.23670/irj.2022.121.7.038>.

4. Карпов О.Э., Андриков Д.А., Максименко В.А., Храмов А.Е. Прозрачный искусственный интеллект для медицины. *Врач и информационные технологии* 2022; 2: 4–11, https://doi.org/10.25881/18110193_2022_2_4.

Karpov O.E., Andrikov D.A., Maksimenko V.A., Hramov A.E. Explainable artificial intelligence for medicine. *Vrac i informacionnye tehnologii* 2022; 2: 4–11, https://doi.org/10.25881/18110193_2022_2_4.

5. Афимина К.Г., Кушничук И.И. Применение методов искусственного интеллекта в медицине. *Известия Российской военно-медицинской академии* 2021; 40(S1–S3): 17–19.

Afimina K.G., Kushnirchuk I.I. Application of artificial intelligence methods in medicine. *Izvestia Rossijskoj voenno-meditsinskoj akademii* 2021; 40(S1–S3): 17–19.

6. Елизарова М.И., Уразова К.М., Ермашов С.Н., Пронькин Н.Н. Искусственный интеллект для медицины. *International Journal of Professional Science* 2021; 5: 81–85.

Elizarova M.I., Urazova K.M., Ermashov S.N., Pronkin N.N. Artificial intelligence in medicine. *International Journal of Professional Science* 2021; 5: 81–85.

7. Гусев А.В., Владимирский А.В., Шарова Д.Е., Арзамасов К.М., Храмов А.Е. Развитие исследований и разработок в сфере технологий искусственного интеллекта для здравоохранения в Российской Федерации: итоги 2021 года. *Digital Diagnostics* 2022; 3(3): 178–194, <https://doi.org/10.17816/dd107367>.

Gusev A.V., Vladzimirskiy A.V., Sharova D.E., Arzamasov K.M., Khramov A.E. Evolution of research and development in the field of artificial intelligence technologies for healthcare in the Russian Federation: results of 2021. *Digital Diagnostics* 2022; 3(3): 178–194, <https://doi.org/10.17816/dd107367>.

8. Морозов С.П., Владимирский А.В., Кляшторный В.Г., Андрейченко А.Е., Кульберг Н.С., Гомболевский В.А., Сергунова К.А. Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика). Серия «Лучшие практики лучевой и инструментальной диагностики». Вып. 57. М.; 2019; 51 с.

Morozov S.P., Vladzimirskiy A.V., Klyashtornyy V.G., Andreychenko A.E., Kul'berg N.S., Gomboleviskiy V.A.,

Sergunova K.A. *Klinicheskie ispytaniya programmogo obespecheniya na osnove intellektual'nykh tekhnologiy (luchevaya diagnostika). Seriya "Luchshie praktiki luchevoy i instrumental'noy diagnostiki"*. Вып. 57 [Clinical trials of software based on intelligent technologies (radiology). Series "Best practices of radiological and instrumental diagnostics". Issue 57]. Moscow; 2019; 51 p.

9. Приказ Министерства здравоохранения РФ от 15.09.2020 №980н «Об утверждении Порядка осуществления мониторинга безопасности медицинских изделий».

Приказ Министерства здравоохранения РФ от 15.09.2020 №980н «Об утверждении Порядка осуществления мониторинга безопасности медицинских изделий» [Order of the Ministry of Health of the Russian Federation of September 15, 2020 No.980n "On approval of the Procedure for monitoring the safety of medical devices"].

10. Решение Коллегии Евразийской экономической комиссии от 22.12.2015 №174 «Об утверждении Правил проведения мониторинга безопасности, качества и эффективности медицинских изделий».

Reshenie Kollegii Evraziyskoy ekonomicheskoy komissii ot 22.12.2015 No.174 "Ob utverzhdenii Pravil provedeniya monitoringa bezopasnosti, kachestva i effektivnosti meditsinskikh izdeliy" [Decision of the Board of the Eurasian Economic Commission of December 22, 2015 No.174 "On approval of the Rules for monitoring the safety, quality and efficiency of medical devices"].

11. World Health Organization. *Guidance for post-market surveillance and market surveillance of medical devices, including in-vitro-diagnostics*. WHO; 2020. URL: https://www.who.int/docs/default-source/essential-medicines/in-vitro-diagnostics/draft-public-pmsdevices.pdf?sfvrsn=f803f68a_2.

12. European Commission. *Guidance on Clinical Evaluation (MDR)/Performance Evaluation (IVDR) of Medical Device Software*. Luxembourg; 2020. URL: https://ec.europa.eu/health/system/files/2020-09/md_mdcg_2020_1_guidance_clinic_eva_md_software_en_0.pdf.

13. U.S. Food and Drug Administration. *Postmarket Surveillance Under Section 522 of the Federal Food, Drug, and Cosmetic Act. Guidance for Industry and Food and Drug Administration Staff*. 2021. URL: <https://www.fda.gov/media/81015/download>.

14. Федеральный закон от 21.11.2011 №323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации» (в ред. от 01.01.2022).

Federal'nyy zakon ot 21.11.2011 No.323-FZ "Ob osnovakh okhrany zdorov'ya grazhdan v Rossiyskoj Federatsii" (v red. ot 01.01.2022) [Federal Law of November 21, 2011 No.323-FZ "On the basics of protecting the health of citizens in the Russian Federation" (as amended on January 1, 2022)].

15. Benjamins S., Dhunoo P., Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med* 2020; 3: 118, <https://doi.org/10.1038/s41746-020-00324-0>.

16. Kelly C.J., Karthikesalingam A., Suleyman M., Corrado G., King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019; 17(1): 195, <https://doi.org/10.1186/s12916-019-1426-2>.

17. U.S. Food and Drug Administration. *Proposed Regulatory Framework for Modifications to Artificial Intelligence Machine Learning (AI ML)-Based Software as a Medical Device (SaMD). Discussion Paper and Request for Feedback*. 2019. URL: <https://www.fda.gov/media/122535/download>.

18. IMDRF Software as a Medical Device (SaMD) Working Group. "Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations. 2014. URL: <https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf>.
19. Park Y., Jackson G.P., Foreman M.A., Gruen D., Hu J., Das A.K. Evaluating artificial intelligence in medicine: phases of clinical research. *JAMIA Open* 2020; 3(3): 326–331, <https://doi.org/10.1093/jamiaopen/ooaa033>.
20. Article 78 — post-market surveillance system of the manufacturer. URL: https://lexpacency.org/eu/32017R0746/ART_78.
21. Florey C.D. Sample size for beginners. *BMJ* 1993; 306(6886): 1181–1184, <https://doi.org/10.1136/bmj.306.6886.1181>.
22. Адлер Ю.П. Выборка: «все или ничего». *Контроль качества продукции* 2015; 8: 26–32.
- Adler Yu.P. Sample: "all or nothing". *Kontrol' kachestva produktsii* 2015; 8: 26–32.
23. Адлер Ю.П. Репрезентативна ли ваша выборка? *Контроль качества продукции* 2016; 5: 39–43.
- Adler Yu.P. Is your sample representative? *Kontrol' kachestva produktsii* 2016; 5: 39–43.
24. Бондаренко Н.Н., Писарик В.М., Атрашкевич Т.И., Новик И.И. Формирование репрезентативной выборки для STEPS-исследования в республике Беларусь. *Вопросы организации и информатизации здравоохранения* 2018; 2: 30–38.
- Bandarenko N.N., Pisaryk V.M., Atrashkevich T.I., Novik I.I. Forming of the representative sample for steps-survey in the republic of Belarus. *Voprosy organizatsii i informatizatsii zdavoohranenia* 2018; 2: 30–38.
25. Burmeister E., Aitken L. Sample size: how many is enough? *Aust Crit Care* 2012; 25(4): 271–274, <https://doi.org/10.1016/j.aucc.2012.07.002>.
26. Naing L., Winn T., Rusli B.N. Practical issues in calculating the sample size for prevalence studies. *Arch Orofac Sci* 2006; 1: 9–14.
27. Braganza O. *Economically rational sample-size choice and irreproducibility*. arXiv; 2019; URL: <https://arxiv.org/pdf/1908.08702v2.pdf>.
28. Singh A.S., Masuku M.B. Sampling techniques & determination of sample size in applied statistics research: an overview. *Int J Economics Commerce Manag* 2014; 2(11): 1–22.
29. Lwanga S.K., Lemeshow S. *Sample size determination in health studies*. World Health Organization; 1991; 80 p.
30. Kim J., Seo B.S. How to calculate sample size and why. *Clin Orthop Surg* 2013; 5(3): 235–242, <https://doi.org/10.4055/cios.2013.5.3.235>.
31. Шарафутдинова Н.Х., Киреева Э.Ф., Николаева И.Е., Павлова М.Ю., Халфин Р.М., Шарафутдинов М.А., Борисова М.В., Латыпов А.Б., Галикеева А.Ш. *Статистические методы в медицине и здравоохранении*. Уфа: ФГБОУ ВО БГМУ Минздрава России; 2018; 131 с.
- Sharafutdinova N.Kh., Kireeva E.F., Nikolaeva I.E., Pavlova M.Yu., Khalfin R.M., Sharafutdinov M.A., Borisova M.V., Latypov A.B., Galikееva A.Sh. *Statisticheskie metody v meditsine i zdavoookhraneni* [Statistical methods in medicine and public health]. Ufa: FGBOU VO BGMU Minzdrava Rossii; 2018; 131 p.
32. Skaff P.A., Sloan J. Design and analysis of equivalence clinical trials via the SAS system. *Proc SUGI* 1998; 23: 1166–1171.
33. Койчубеков Б.К., Сорокина М.А., Мхитарян К.Э. Определение размера выборки при планировании научного исследования. *Международный журнал прикладных и фундаментальных исследований* 2014; 4: 71–74.
- Koichubekov B.K., Sorokina M.A., Mkhitaryan X.E. Sample size determination in planning of scientific research. *Mezhdunarodnyj zurnal prikladnyh i fundamental'nyh issledovanij* 2014; 4: 71–74.
34. Noordzij M., Tripepi G., Dekker F.W., Zoccali C., Tanck M.W., Jager K.J. Sample size calculations: basic principles and common pitfalls. *Nephrol Dial Transplant* 2010; 25(5): 1388–1393, <https://doi.org/10.1093/ndt/gfp732>.
35. Cody J. *Sample size calculation using SAS®, R, and nQuery software*. SAS Global Forum; 2020. URL: <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2020/4675-2020.pdf>.
36. Тараканова В.В., Наумкин Б.И. Формирование выборочной совокупности. *Эксперимент и инновации в школе* 2009; 3: 46–49.
- Tarakanova V.V., Naumkin B.I. Formation of the sample population. *Ekspieriment i innovatsii v skole* 2009; 3: 46–49.
37. Israel G.D. *Determining sample size*. Florida: University of Florida, IFAS extension; 2012.
38. Kadam P., Bhalerao S. Sample size calculation. *Int J Ayurveda Res* 2010; 1(1): 55–57, <https://doi.org/10.4103/0974-7788.59946>.
39. Dell R.B., Holleran S., Ramakrishnan R. Sample size determination. *ILAR J* 2002; 43(4): 207–213, <https://doi.org/10.1093/ilar.43.4.207>.
40. Lakens D. Sample size justification. *Collabra Psychol* 2022; 8(1): 1–32.
41. Kirby A., GebSKI V., Keech A.C. Determining the sample size in a clinical trial. *Med J Aust* 2002; 177(5): 256–257, <https://doi.org/10.5694/j.1326-5377.2002.tb04759.x>.
42. Jones S.R., Carley S., Harrison M. An introduction to power and sample size estimation. *Emerg Med J* 2003; 20(5): 453–458, <https://doi.org/10.1136/emj.20.5.453>.
43. Rebrova O.Yu., Gusev A.V. Sample size calculation for clinical trials of medical decision support systems with binary outcome. *Sovremennye tehnologii v medicine* 2022; 14(3): 6, <https://doi.org/10.17691/stm2022.14.3.01>.
44. Schilling E.G., Neubauer D.V. *Acceptance sampling in quality control*. Taylor & Francis Group, LLC; 2008; 709 p.
45. Полунина Н.В. *Общественное здоровье и здравоохранение*. М: Медицинское информационное агентство; 2010; 544 с.
- Polunina N.V. *Obshchestvennoe zdorov'e i zdavoookhranenie* [Public health and healthcare]. Moscow: Meditsinskoe informatsionnoe agentstvo; 2010; 544 p.
46. Наркевич А.Н., Виноградов К.А. Методы определения минимально необходимого объема выборки в медицинских исследованиях. *Социальные аспекты здоровья населения* 2019; 65(6): 10.
- Narkevich A.N., Vinogradov K.A. Methods for determining the minimum required sample size in medical research. *Social'nye aspekty zdorov'a naselenia* 2019; 65(6): 10.
47. Паниотто В.И., Максименко В.С. *Количественные методы в социологических исследованиях*. Киев; 2003. URL: https://www.kiis.com.ua/materials/books/376072_C6170_paniotto_v_i_maksimenko_v_s_kolichestvennye_metody_v_sociolo.pdf.

- Paniotto V.I., Maksimenko V.S. *Kolichestvennye metody v sotsiologicheskikh issledovaniyakh* [Quantitative methods in sociological research]. Kiev; 2003. URL: https://www.kiis.com.ua/materials/books/376072_C6170_paniotto_v_i_maksimenko_v_s_kolichestvennye_metody_v_sociolo.pdf.
48. Сырцова Л.Е., Косаговская И.И., Авксентьева М.В. *Основы эпидемиологии и статистического анализа в общественном здоровье и управлении здравоохранением*. М; 2003; 91 с.
- Syrtsova L.E., Kosagovskaya I.I., Avksent'eva M.V. *Osnovy epidemiologii i statisticheskogo analiza v obshchestvennom zdorov'e i upravlenii zdravookhraneniem* [Fundamentals of epidemiology and statistical analysis in public health and health management]. Moscow; 2003; 91 p.
49. Агишева Д.К., Зотова С.А., Матвеева Т.А., Светличная В.Б. *Математическая статистика*. Волгоград: ВПИ (филиал) ВолгГТУ; 2010; 159 с.
- Agisheva D.K., Zotova S.A., Matveeva T.A., Svetlichnaya V.B. *Matematicheskaya statistika* [Mathematical statistics]. Volgograd: VPI (filial) VolG GTU; 2010; 159 p.
50. Румянцев П.О., Саенко В.А., Румянцева У.В., Чекин С.Ю. *Статистические методы анализа в клинической практике*. 2009. URL: <https://medstatistic.ru/articles/StatMethodsInClinics.pdf>.
- Rumyantsev P.O., Saenko V.A., Rumyantseva U.V., Chekin S.Yu. *Statisticheskie metody analiza v klinicheskoy praktike* [Statistical methods of analysis in clinical practice]. 2009. URL: <https://medstatistic.ru/articles/StatMethodsInClinics.pdf>.
51. Taherdoost H. Determining sample size; how to calculate survey sample size. *Int J Econ Manag Syst* 2017; 2: 237–239.
52. Blackwelder W.C. Equivalence trials. In: *Encyclopedia of biostatistics*. Volume 2. New York: John Wiley and Sons; 1998; p. 1367–1372.
53. Chow S.C., Shao J., Wang H. *Sample size calculations in clinical research*. 2nd Edition. Florida: Chapman & Hall/CRC Biostatistics Series; 2008.
54. ГОСТ Р ИСО 2859-1-2007. *Статистические методы. Процедуры выборочного контроля по альтернативному признаку. Часть 1. Планы выборочного контроля последовательных партий на основе приемлемого уровня качества*. М: Стандартинформ; 2007; 101 с.
- GOST R ISO 2859-1-2007. *Statisticheskie metody. Protседury vyborochnogo kontrolya po al'ternativnomu priznaku. Chast' 1. Plany vyborochnogo kontrolya posledovatel'nykh partiy na osnove priemlegomogo urovnya kachestva* [Statistical methods. Sampling procedures for inspection by attributes. Part 1: sampling schemes indexed by acceptance quality limit for lot-by-lot inspection]. Moscow: Standartinform; 2007; 101 p.
55. Шарашкина Т.П. *Статистические методы в управлении качеством*. Саранск: Мордовский государственный университет; 2013; 91 с.
- Sharashkina T.P. *Statisticheskie metody v upravlenii kachestvom* [Statistical methods in quality management]. Saransk: Mordovskiy gosudarstvennyy universitet; 2013; 91 p.
56. Бородачев С.М. *Статистические методы в управлении качеством*. Екатеринбург: Издательство Уральского университета; 2016; 87 с.
- Borodachev S.M. *Statisticheskie metody v upravlenii kachestvom* [Statistical methods in quality management]. Ekaterinburg: Izdatel'stvo Ural'skogo universiteta; 2016; 87 p.
57. Клячкин В.Н. *Статистические методы в управлении качеством*. Ульяновск: УлГТУ; 2013; 156 с.
- Klyachkin V.N. *Statisticheskie metody v upravlenii kachestvom* [Statistical methods in quality management]. Ul'yanovsk: UIGTU; 2013; 156 p.
58. Ефимов В.В. *Основы бережливого производства*. Ульяновск: УлГТУ; 2011; 160 с.
- Efimov V.V. *Osnovy berezhlivogo proizvodstva* [Fundamentals of lean manufacturing]. Ul'yanovsk: UIGTU; 2011; 160 p.
59. Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department. *Experiment on the use of innovative computer vision technologies for analysis of medical images in the Moscow healthcare system*. URL: <https://www.clinicaltrials.gov/ct2/show/NCT04489992>.
60. Andreychenko A.E., Logunova T.A., Gombolevskiy V.A., Nikolaev A.E., Vladzmyrskyy A.V., Sinityn V.E., Morozov S.P. *A methodology for selection and quality control of the radiological computer vision deployment at the megalopolis scale*. medRxiv; 2010, <https://doi.org/10.1101/2022.02.12.22270663>.
61. *Международная классификация болезней 10-го пересмотра (МКБ-10)* 2021. URL: <https://mkb-10.com>.
- Mezhdunarodnaya klassifikatsiya bolezney 10-go peresmotra (МКБ-10)*. [International Classification of Diseases of the 10th Revision (ICD-10)]. 2021. URL: <https://mkb-10.com>.
62. *Приказ Департамента здравоохранения города Москвы от 24.02.2022 №160 «Об утверждении Порядка и условий проведения эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы»*.
- Prikaz Departamenta zdravookhraneniya goroda Moskvy ot 24.02.2022 No.160 "Ob utverzhdenii Poryadka i usloviy provedeniya eksperimenta po ispol'zovaniyu innovatsionnykh tekhnologiy v oblasti komp'yuternogo zreniya dlya analiza meditsinskikh izobrazheniy i dal'neyshego primeneniya v sisteme zdravookhraneniya goroda Moskvy"*. [Order of the Department of Health of the city of Moscow dated February 24, 2022 No.160 "On approval of the Procedure and conditions for conducting an experiment on the use of innovative technologies in the field of computer vision for the analysis of medical images and further application in the health care system of the city of Moscow"].
63. Баврина А.П. Основные понятия статистики. *Медицинский альманах* 2020; 3: 101–111.
- Bavrina A.P. Basic concepts of statistics. *Medicinskij al'manah* 2020; 3: 101–111.
64. Кошевой О.С., Карпова М.К. Определение объема выборочной совокупности при проведении региональных социологических исследований. *Известия высших учебных заведений. Поволжский регион* 2011; 2: 98–104.
- Koshevoy O.S., Karpova M.K. Sample size determination in the course of regional sociological research. *Izvestia vyssih ucebnyh zavedenij. Povolzskij region* 2011; 2: 98–104.
65. Fox N., Hunn A., Mathers N. *Sampling and sample size calculation*. Sheffield: Trent RDSU; 2007; 41 p.