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 Technologies[1];
K.M. Arzamasov, MD, PhD, Head of the Department of Medical Informatics,
Radiomics and Radiogenomics[1];
A.E. Andreichenko, PhD, Leading Researcher, Department of Medical Informatics,
Radiomics and Radiogenomics[1]; Head of the Group of Artificial Intelligence[2]; Leading Researcher[3];
V.P. Novik, Researcher, Sector of the Development of Systems for the Implementation
of Intelligent Medical Technologies[1];
T.M. Bobrovskaya, Junior Researcher, Sector of the Development of Systems for the Implementation
of Intelligent Medical Technologies[1];
A.V. Vladzimirsky, MD, DSc, Deputy Director for Research[1]; Professor, Department of Information
and Internet Technologies[4]

[1]Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies
of the Moscow Health Care Department, 24/1 Petrovka St., Moscow, 127051, Russia;
[2]K-SkAI LLC, 17 Naberezhnaya Varkausa, Petrozavodsk, the Republic of Karelia, 185031, Russia;
[3]ITMO National Research University, 49 Kronverksky Pr., Saint Petersburg, 197101, Russia;
[4]First 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.

Introduction

Artificial intelligence (AI) in medicine is a tool that automates routine processes such as filling out electronic records, analyzing and diagnosing medical images, doing analytics, and patient treatment planning. AI helps to reduce labor costs in the process of medical and biological activities, as well as improve the accuracy of

---

Corresponding author: Kirill M. Arzamasov, e-mail: ArzamasovKM@zdrav.mos.ru

recommendations, diagnostics, prescriptions, etc. [1–7].

When introducing AI systems into clinical practice, an important part is the quality control of AI results [8–10] in order to confirm their safety and effectiveness [11–20]. For an autonomous and adaptive AI system, the operational periodic quality control ensuring adjustment with minimal risks and a short response time is of particular value.

All types of defects in the AI operation can be conditionally divided into two groups according to the method of detection: those detected during automated and manual checks. An automated check provides quality control of the entire population over a certain period. A manual check is carried out on a limited sample and requires significant resource costs, as it involves opening and viewing images under study; clinical analysis of the original image; assessment of the AI results; management of the quality control records, etc. In this regard, the issue of sampling for quality control of the AI results is relevant and should solve at least two important tasks:

1) representativeness and correctness of the proportions of distribution (content) of the studied trait [21–24];

2) accounting for the labor costs of employees involved in the manual quality control of the AI results.

Many approaches have been proposed for calculating sample sizes in various fields of science [25–43], however, it is not possible to make a reasoned choice of one or another approach when planning a biomedical research or when introducing AI for practical use. Thus, in paper [37], it is reported that, depending on the approach used, different sample sizes are obtained.

The aim of the study is to evaluate the efficacy of approaches to sampling when conducting periodic quality control of the AI results in biomedical practice.

Materials and Methods

Sampling approaches based on point statistical estimation [31, 44–47], statistical hypothesis testing — variant 1 [48–51] and variant 2 [52, 53], the application of GOST R ISO 2859-1-2007 "Statistical methods. Sampling procedures for inspection by attributes" [54–58] were considered.

The analysis of approaches to sampling was carried out as part of an experiment on the use of innovative technologies in the field of computer vision for the analysis of medical images and their further application in the healthcare system of Moscow (Russia) [59, 60]. Previously, various types of defects were found in the results of processing medical images by AI [59], which reduce the clinical and diagnostic value of the systems we studied.

In this article, a set of medical AI-processed studies over a certain period of time will be called a batch. Each batch consists of product items of the same type, class, size, and composition, processed under almost identical conditions over the same period of time. A product item is an AI-processed study (connected to the radiological information system "Unified Radiological Information Service of the City of Moscow" (URIS)) for a particular reporting period.

Each study contains the following data:

population parameters (gender and age indicators, ethnic composition, regions of residence, etc.); depersonalization data; information on medical facilities which are the source for the formation of a data set;

study characteristics (anatomical area(s); modality; projections; types of medical products — diagnostic devices; types and characteristics of research protocols);

target pathology in accordance with the International classification of diseases [61];

cases of presence/absence of pathological findings.

The quality control of the AI results is carried out repeatedly. When controlling the quality of the results of the AI systems which we study, a manual review of the studies is carried out by experts. Due to a large number (more than 1000) of studies in the batch, there is no possibility of quality control in full due to time constraints, as well as due to the small number of experts. As part of the experiment [59], it has been found that no more than 10% of defective product items are contained in the general population. This means that the entire batch for the reporting period contains no more than 10% of product items with technological defects [62]. Thus, the article describes approaches that correspond to a qualitative feature, where the proportion of cases in which the studied trait occurs is known. The volume of the general population ranges from 1000 to 100,000 product items.

The solution to the problem of the quality control of a batch with AI results was provided on the basis of selective observation, based on the concepts of general and sample population.

This article describes serial repetition-free sampling, where a selected product item was drawn from the entire volume of the general population and not returned back; thus, the probability of getting the remaining product items increased.

Calculations were performed using the PASS 15 Update — NCSS (www.ncss.com) and LibreOffice Calc (www.libreoffice.org) software.

Results and Discussion

The approach based on point statistical estimation considers the deviation of the results of a sample study from the general values. With this approach, a sample size is calculated by the formula:

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

where n is the sample size; N is the size of the general population; t is a coefficient showing with what probability (reliability) it is possible to guarantee the reliability of the result obtained or the critical value of the Student's criterion at the appropriate significance level (for a significance

level of 0.05, the coefficient t=1.96); Δ is the limiting error of the indicator; w is the proportion of the investigated trait; q=($1-w$) is the proportion where the investigated trait is absent.

Thus, when the proportion of the investigated trait (w) is 0.9, the level of statistical significance is 0.95, and the maximum permissible error (Δ) is 0.05, we get a sample size (n) equal to 138 product items.

The statistical hypothesis testing approach (variant 1) involves testing the statistical hypothesis H_0 (the quality of the batch meets the requirements) in the presence of an alternative hypothesis H_1 (the quality of the batch does not meet the requirements). If among n-product items the number of defective ones (m) does not exceed the acceptance number c ($m \leq c$) (the maximum permissible number of rejected sample items, which allows making a decision on the acceptance of a product batch in terms of quality), then the product batch is accepted; otherwise, it is rejected.

To select a control plan (determination of a sample), the following formula is used:

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

where m is the number of defective items in a sample of n; $p_n(m)$ is the probability of occurrence of m defective product items in the n sample; c is the acceptance number.

Since in the scope of the experiment [59], the size of the entire batch exceeded the sample size by more than 10% [56], the operational characteristics were determined by the formula:

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

where C_n^m is the number of combinations of the appearance of m defective product items in the sample of n:

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

In the experimental example [59], the acceptance number equal to two product items was used, calculations were made and curves were plotted for samples of 30, 50, 80, 138 product items (see the Figure).

The figure shows the consumer's and the supplier's risks. The supplier's risk is the probability of rejecting a good quality batch (i.e., in the general population, the proportion of defective items of products is less than 10%). Taking into account the proportion of declared defective products from the supplier, the risk is assumed to be 1%. The consumer's risk is the probability of accepting
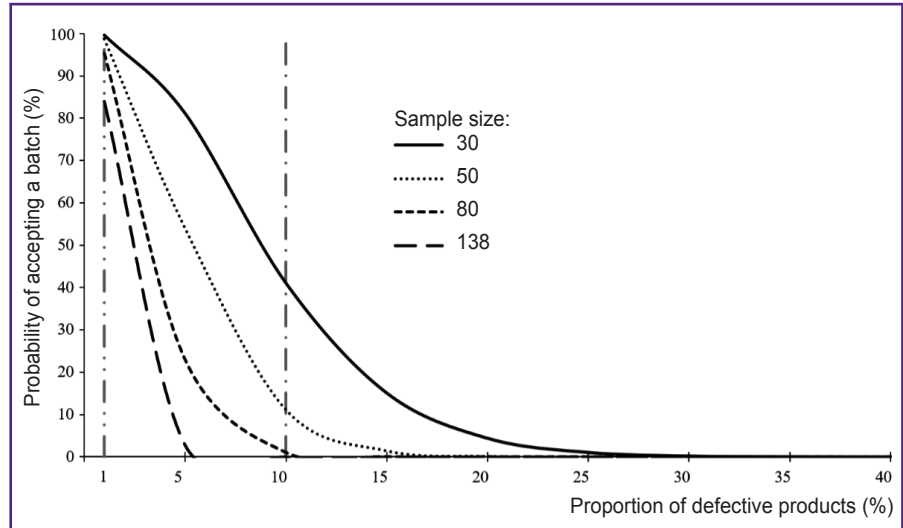
Operational characteristics for different sample sizes:
vertical dash-dotted line with two dots — supplier's risk; vertical dash-dotted line with one dot — consumer's risk

T a b l e 1

Dependence of risks on the sample size

Sample size	Consumer's risk (%)	Supplier's risk (%)
30	41	0
50	11	1
80	1	5
138	0	16

a low-quality batch. Accounting for the proportion of defective products identified by the consumer, the risk is assumed to be 10%.

Analyzing the data from Table 1 and taking into account the specified risks to the consumer and the supplier at the level of no more than 10% and no more than 5%, respectively, we found that the sample size, equal to 80 product items, meets the requirements of both the consumer and the supplier.

The approach based on statistical hypothesis testing (variant 2) is based on the principles of the probability of rejecting the null hypothesis; it takes into account both the supplier's and the consumer's risks. The null hypothesis H_0 assumes that if the general population contains more than 10% of defective product items, then the entire batch for the reporting period contains more than 10% of product items with technological defects. Accordingly, under the alternative hypothesis H_1, less than 10% of product items have technological defects. The probability of rejecting the null hypothesis is at least 80%.

Calculations were made (Table 2) for samples of 30, 50, 80, 120 product items with the acceptance number from zero to four (the acceptance number was limited by exceeding the risks to the consumer of over 10% or the supplier — of over 5%).

Table 2

Dependence of risks on the sample size and acceptance number

Sample size	Acceptance number/proportion of defects (%)	Probability of deviation from the null hypothesis (%)	Supplier's risk (%)	Consumer's risk (%)
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

Analyzing the data from Table 2 and taking into consideration the specified risks to the consumer and the supplier, as well as the proportion of declared defective items from the supplier (1%) and the proportion of defective items identified by the consumer (10%), it has been found that the sample size equal to 30, 50, 80, and 120 product items, meets the requirements of both the consumer and the supplier with the acceptance number equal to zero product items. Taking into account the proportion of defective items with acceptance numbers greater than zero, the most appropriate sample sizes were 80 or 120 items.

The approach based on the application of GOST R ISO 2859-1-2007 "Statistical methods. Sampling procedures for inspection by attributes" involves the determination of a sampling scheme for successive batches based on an acceptable level of quality and can be applied to various types of data or records. The acceptable level of quality is expressed as the percentage of nonconforming product items or the number of nonconformities per hundred product items.

We have considered several variants of sample size determination.

First, we addressed the "Sample size codes" table [54]. In our case, the general control level is II, special inspection levels are not used. Since our batch sizes ranged from 1000 to 100,000, the J, K, L, M codes were of interest for us. In this work, we did not plan multi-stage sampling, neither did we imply a transition to a weakened or enhanced control. In this regard, we used the data from the table "Single-stage plans under normal control (main

Table 3

Pros and cons of the approaches to sampling

Features	Approach based on point statistical estimation	Approach based on statistical hypothesis testing (variant 1)	Approach based on statistical hypothesis testing (variant 2)	Approach based on the application of GOST R ISO 2859-1-2007
Pros	Ability to determine the sample size using statistical parameters. Simple math calculations. Ability to choose a confidence interval sufficient for a confident judgment about general parameters based on known sample indicators	The consumer's and supplier's risks are taken into account. Ability to calculate risks depending on the sample size, the proportion of defective product items and acceptance number. Ability to visually determine the most appropriate parameters for research	The consumer's and supplier's risks are taken into account. Ability to calculate risks depending on the sample size, the proportion of defective product items and acceptance number. Statistical parameters are taken into account. Ability to calculate confidence intervals for the obtained values	The consumer's and supplier's risks are taken into account. Minimum preliminary population data are required. No need for mathematical calculations. No need for additional data processing software. Visual determination of the most suitable parameters for research. Possibility to apply different control plans at low or high batch quality levels. Dependence of a sample size on high and low population sizes
Cons	Type II error is not taken into account. Statistical parameters and confidence interval in biomedical research have a limited scope, beyond which the results are not statistically significant. Preliminary data on general population are required	Preliminary population data are required. Data preprocessing is required to determine the formulas used. Fixed values of risks and proportions of defective product items are required before conducting research	Preliminary population data are required. Fixed risk values are required before conducting research. Cases of high item cost used for quality control are not taken into consideration. The use of paid software for statistical processing is implied	Finer choice of statistical parameters is not allowed. More accurate calculation of the consumer's and supplier's risks is not allowed. Fixed values of risks and proportions of defective product items before conducting research are required
Sample size	138	80	80, 120	125, 500

   S.F. Chetverikov, K.M. Arzamasov, A.E. Andreichenko, V.P. Novik, T.M. Bobrovskaya, A.V. Vladzimirsky

table)" [54]: for an acceptable consumer quality level of 10% (for batches of 501 to 10,000 studies), the sample size for quality control will be equal to 125 product items with the batch acceptance number equal to zero; for batches of 10,001 to 150,000, the quality control sample size will be 500 items with the batch acceptance number equal to one.

We then addressed the table "Producer's risk under normal control (percentage of rejected batches for single-stage plans)" [54] and obtained the supplier's risk of 11.8% for a sample of 125 product items; 9.02% for a sample of 500 product items.

Table 3 summarizes the pros and cons of the considered approaches.

The approaches to sample size calculation that we have considered have a number of advantages over the widely used approaches based on fixed or tabular values. Thus, for example, the approaches with application of ready-made tables have specific statistical input data, which does not make it possible to consider them universal [46, 63, 64].

The table based on V.I. Paniotto's methodology [47] contains values that are calculated for specific parameters: the proportion of the trait is 0.5; allowable error — 0.05; confidence level — 0.954 ($t=2$).

The table based on N. Fox' methodology [65] also contains a specific parameter: the proportion of the trait is 0.35.

The values from the tables [47, 65] were calculated using the formulas of the point statistical estimation described in this article above. If the input data of a biomedical study do not match the parameters of the tables, we recommend performing calculations rather than using the given tables to determine the sample size.

Conclusion

The results of this study provide a choice of the most appropriate approaches to achieving the goals of biomedical research. The use of point statistical estimation and the approach based on statistical hypothesis testing allows for the most flexible calculation of sample sizes depending on the input parameters of the study. The use of GOST R ISO 2859-1-2007 for sampling is a priority if the experiment involves the interaction of the researcher and a third-party organization. This allows taking account of the risks and errors to both parties involved in the process.

The optimal number of studies during quality control of the AI systems that we have studied for the analysis of medical images is 80 product items. This meets the requirements of representativeness, balance of risks to the consumer and the providers of AI services, as well as optimization of labor costs of employees involved in the quality control of artificial intelligence.

References

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

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

3. Alekseeva M.G., Zubov A.I., Novikov M.Yu. Artificial intelligence in medicine. *Mezdunarodnyj naucno-issledovatel'skij zurnal* 2022; 7–2: 10–13, https://doi.org/10.23670/irj.2022.121.7.038.

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

6. 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. Gusev A.V., Vladzymyrskyy 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. Morozov S.P., Vladzimirskiy A.V., Klyashtornyy V.G., Andreychenko A.E., Kul'berg N.S., Gombolevskiy V.A., Sergunova K.A. *Klinicheskie ispytaniya programmnogo obespecheniya na osnove intellektual'nykh tekhnologiy (luchevaya diagnostika). Seriya "Luchshie praktiki luchevoy i instrumental'noy diagnostiki"*. Vyp. 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. *Prikaz Ministerstva zdravookhraneniya RF ot 15.09.2020 No.980n "Ob utverzhdenii Poryadka osushchestvleniya monitoringa bezopasnosti meditsinskikh izdeliy"* [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. *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. *Federal'nyy zakon ot 21.11.2011 No.323-FZ "Ob osnovakh okhrany zdorov'ya grazhdan v Rossiyskoy 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. Benjamens S., Dhunnoo 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://lexparency.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. Adler Yu.P. Sample: "all or nothing". *Kontrol' kacestva produkcii* 2015; 8: 26–32.

23. Adler Yu.P. Is your sample representative? *Kontrol' kacestva produkcii* 2016; 5: 39–43.

24. 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 organizacii i informatizacii zdravoohranenia* 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. Sharafutdinova N.Kh., Kireeva E.F., Nikolaeva I.E., Pavlova M.Yu., Khalfin R.M., Sharafutdinov M.A., Borisova M.V., Latypov A.B., Galikeeva A.Sh. *Statisticheskie metody v meditsine i zdravookhranenii* [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. Koichubekov B.K., Sorokina M.A., Mkhitaryan X.E. Sample size determination in planning of scientific research. *Mezdunarodnyj 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. Tarakanova V.V., Naumkin B.I. Formation of the sample population. *Eksperiment i innovacii 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. Polunina N.V. *Obshchestvennoe zdorov'e i zdravookhranenie* [Public health and healthcare]. Moscow: Meditsinskoe informatsionnoe agentstvo; 2010; 544 p.

46. 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. 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. 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. Agisheva D.K., Zotova S.A., Matveeva T.A., Svetlichnaya V.B. *Matematicheskaya statistica* [Mathematical statistics]. Volgograd: VPI (filial) VolgGTU; 2010; 159 p.

50. 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. *GOST R ISO 2859-1-2007. Statisticheskie metody. Protsedury vyborochnogo kontrolya po al'ternativnomu priznaku. Chast' 1. Plany vyborochnogo kontrolya posledovatel'nykh partiy na osnove priemlemogo 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. Sharashkina T.P. *Statisticheskie metody v upravlenii kachestvom* [Statistical methods in quality management]. Saransk: Mordovskiy gosudarstvennyy universitet; 2013; 91 p.

56. Borodachev S.M. *Statisticheskie metody v upravlenii kachestvom* [Statistical methods in quality management]. Ekaterinburg: Izdatel'stvo Ural'skogo universiteta; 2016; 87 p.

57. Klyachkin V.N. *Statisticheskie metody v upravlenii kachestvom* [Statistical methods in quality management]. Ul'yanovsk: UlGTU; 2013; 156 p.

58. Efimov V.V. *Osnovy berezhlivogo proizvodstva* [Fundamentals of lean manufacturing]. Ul'yanovsk: UlGTU; 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., Vladzymyrskyy A.V., Sinitsyn 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. *Mezhdunarodnaya klassifikatsiya bolezney 10-go peresmotra (MKB-10)* [International Classification of Diseases of the 10th Revision (ICD-10)]. 2021. URL: https://mkb-10.com.

62. *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. Bavrina A.P. Basic concepts of statistics. *Medicinskij al'manah* 2020; 3: 101–111.

64. 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.