

NON-COMMERCIAL JOINT STOCK COMPANY
«KARAGANDA MEDICAL UNIVERSITY»

ANNOTATION

of the dissertation for the degree of Doctor of Philosophy (PhD)

«Prognostic assessment of the influence of clinical and genetic factors on the risk of development and clinical course of spontaneous intracerebral hemorrhage in the main ethnic population of Kazakhstan»

Educational program: 8D10100 «Medicine»

Author: Muratbekova Shynar Sabitovna

Scientific supervisors:

Grigolashvili M.A., Candidate of Medical Sciences, Associate Professor, Head of the Department of Neurology, Psychiatry and Rehabilitation, NJSC “KMU”

Kadyrova I.A., PhD, Professor, Director of the Institute of Life Sciences, NJSC “KMU”

Foreign scientific consultant:

Vaiva Hendrixson, PhD, MD, Professor, Institute of Biomedical Sciences, Faculty of Medicine, Vilnius University (Vilnius, Lithuania)

Karaganda, 2026

Relevance of the research

Spontaneous intracerebral hemorrhage (ICH) ranks second among all types of stroke and is characterized by the highest rates of mortality and disability. According to the Global Burden of Disease (GBD) 2021 study, 93.8 million cases of stroke of all types were registered worldwide, with ICH accounting for 45.6% of all stroke deaths and 49.5% of the total stroke burden measured in DALYs. Thirty-day mortality in ICH reaches 30–40% in developed countries and 40–55% in low-income countries, and a good functional outcome is observed in only 31.2% of surviving patients.

In Kazakhstan, according to Zhakhina et al. (2022), ICH accounts for 19% of all hospitalized cases of acute cerebrovascular accident; the mean age of patients with ICH (59.2 years) is significantly lower than in ischemic stroke (64.8 years), and 58.3% of patients are of working age. The genetic determinants of ICH in Central Asian, including Kazakh, populations remain practically unstudied.

The *COL4A2* gene, encoding the $\alpha 2$ chain of type IV collagen - a critical component of the basement membranes of cerebral vessels - is considered a promising candidate in the study of genetic predisposition to ICH. In European populations, the minor C allele of the rs9515199 polymorphism of the *COL4A2* gene is associated with an increased risk of deep ICH: OR = 1.24–1.28, whereas in Asian populations the data are contradictory: in the Japanese population the association was not confirmed, in the Chinese Han population other polymorphisms of the same gene are associated with a similar risk. For the Kazakh population, which possesses a unique mixed East- and West-Eurasian gene pool, such studies have not been conducted previously.

Existing prognostic scales for ICH (ICH Score, FUNC Score, MAX ICH Score, ICH-FOS, MICH Score) are post-event, require performed neuroimaging, do not contain genetic markers, and are validated predominantly on European and North American cohorts, which limits their applicability for the tasks of preclinical risk stratification in the Kazakh population.

Thus, the high incidence of ICH in Kazakhstan, the pronounced population plasticity of the *COL4A2* gene, and the absence of population-specific prognostic algorithms determine the relevance of this dissertation research.

Aim of the research: to assess clinical and genetic risk factors for spontaneous intracerebral hemorrhage in the Kazakh population, taking into account epidemiological data, to study their relationship with clinical and neuroimaging characteristics of the disease in order to develop a prognostic algorithm for risk stratification.

Research objectives:

1. To conduct an epidemiological analysis of spontaneous ICH in the Karaganda region with calculation of age-standardized incidence rates, hospital mortality, and population mortality to identify ethnic characteristics.

2. To assess the relationship of major non-modifiable (sex, age) and modifiable (arterial hypertension, diabetes mellitus, alcohol consumption, smoking, obesity) factors with the presence of spontaneous ICH in a case-control study.
3. To determine the allele and genotype frequencies of the rs9521732, rs9521733, and rs9515199 polymorphisms of the *COL4A2* gene, to assess their association with the risk of ICH, and to compare the direction of association with data from European populations.
4. To conduct an exploratory analysis of the relationship of the rs9515199 polymorphism of the *COL4A2* gene with the localization and clinical-neuroimaging characteristics of ICH.
5. To develop a clinical-genetic classification model (risk score scale) for separating individuals with ICH and the control group, to assess its internal discriminatory ability and the incremental value of the genetic marker.

Object of the research - patients with verified spontaneous intracerebral hemorrhage (ethnic Kazakhs) and individuals of the control group examined in medical organizations of the Karaganda region in the period 2020–2021 (n=327: 162 patients, 165 controls); at the first stage, a retrospective epidemiological cohort of patients with ICH of the Karaganda region for 2018–2020 was analyzed (n=1657).

Scientific novelty of the study:

1. For the first time, ethnic differences in the incidence of ICH (intracerebral hemorrhage) have been shown using data from the Karaganda region with the use of age standardization (a 1.47-fold increase), in the absence of such differences in unstandardized rates (1.04), which indicates the existence of population-level risk factors.
2. For the first time, using data from the Karaganda region, a comprehensive study of clinical and genetic factors associated with spontaneous intracerebral hemorrhage was conducted in a sample of individuals of Kazakh nationality.
3. For the first time in the studied sample, the polymorphisms rs9521732, rs9521733, and rs9515199 of the *COL4A2* gene were investigated. A tendency toward an association of the C allele of rs9515199 with lower odds of the disease was identified (unadjusted OR = 0.23–0.50; log-additive model adjusted for diabetes mellitus, alcohol consumption, age, and sex: OR = 0.55; 95% CI: 0.39–0.77). A difference in the frequency distribution was established (0.47 in the Kazakh population versus 0.40 in European populations) and an opposite direction of effect, which casts doubt on the possibility of direct extrapolation of European data to Central Asian populations.
4. An exploratory analysis was conducted of the possible association of the rs9515199 polymorphism with the localization and clinical-neuroimaging characteristics of intracerebral hemorrhage.
5. A prognostic algorithm was proposed (author's certificate of the Republic of Kazakhstan No. 71562), integrating clinical and genetic predictors based on a

previously unused combination of weight coefficients. For the first time, a quantitative assessment of the incremental value of the genetic marker rs9515199 was performed; however, its addition did not lead to a significant improvement in the discriminative ability of the model ($p = 0.137$), which clarifies the role of common genetic variants in ICH risk stratification in the Kazakh population. Its use in clinical practice must be preceded by external prospective validation in an independent sample.

Theoretical significance

1. The obtained data expand understanding of population differences in the epidemiology of spontaneous intracerebral hemorrhage and demonstrate that comparison of crude incidence rates between ethnic groups leads to false-negative conclusions if the difference in age composition of the compared groups is not taken into account.
2. The established opposite direction of the association of the C allele of the rs9515199 polymorphism of the *COL4A2* gene with the risk of intracerebral hemorrhage compared with European populations indicates the impossibility of direct extrapolation of the results of genetic association studies performed in European populations to populations of Central Asia and substantiates the need for population-specific studies.
3. The obtained data on allele and genotype frequencies of three polymorphisms of the *COL4A2* gene for the first time supplement the genetic characterization of the population of the Republic of Kazakhstan and can be used as reference values in planning subsequent genetic-epidemiological studies in the region.
4. The results of the work demonstrate a methodological limitation of the case-control design in studying diseases with a highly prevalent risk factor: under conditions of complete separation of the data, the odds ratio for such a factor has no finite estimate, which must be taken into account when planning and interpreting studies of cerebrovascular diseases.
5. For the first time in the Kazakh population, a comparison of clinical and genetic factors was performed within a single prognostic model. It was shown that the contribution of a single common genetic variant (rs9515199) to the discriminative ability of the model is limited against the background of the dominant clinical factor (arterial hypertension), which clarifies ideas about the real potential of using individual SNPs in risk stratification of intracerebral hemorrhage. Its use in clinical practice must be preceded by external prospective validation in an independent sample.

Practical significance

1. In the studied sample, the significance of such modifiable risk factors for intracerebral hemorrhage as arterial hypertension, diabetes mellitus, and moderate/heavy alcohol consumption was confirmed, which can be taken into account when stratifying the population by degree of risk at the primary health care level.

2. The developed prognostic algorithm (author's certificate of the Republic of Kazakhstan No. 71562) can be considered as a preliminary tool for ICH risk stratification in the Kazakh population. Its use in clinical practice must be preceded by external prospective validation in an independent sample.
3. The applied methodology of age standardization of rates by ethnic affiliation can be used in the analysis of other regions of the republic.

Statements submitted for defense:

1. Differences in the epidemiological characteristics of spontaneous intracerebral hemorrhage between ethnic groups of the population of the Karaganda region are detected only after accounting for the age composition of the groups: incidence in Kazakhs is higher, while in-hospital mortality is lower, as a result of which population mortality does not differ between the groups.
2. The modifiable risk factors for spontaneous intracerebral hemorrhage in the studied sample are arterial hypertension, diabetes mellitus, and moderate or heavy alcohol consumption; smoking and obesity showed no association with the disease.
3. Carriage of the C allele of the rs9515199 polymorphism of the *COL4A2* gene is associated with lower odds of spontaneous intracerebral hemorrhage in the Kazakh population; the direction of the association is opposite to that established in European populations.
4. Carriage of the C allele of the rs9515199 polymorphism of the *COL4A2* gene is associated with lower odds for deep localization of spontaneous intracerebral hemorrhage in the Kazakh population; the direction of the association is opposite to that established in European populations.
5. The developed clinical-genetic model has high internal discriminative ability. The addition of rs9515199 provides only a slight increase in prognostic ability compared with the clinical model.

The study was approved by the Local Ethics Committee of NPJSC "Karaganda Medical University" (protocol No. 57 dated June 19, 2020; protocol No. 11 dated June 3, 2026). Written informed consent was obtained from each participant (or their legal representative). The study was conducted in accordance with the principles of the Declaration of Helsinki of the World Medical Association.

Connection of the dissertation with other research works:

The dissertation was carried out within the framework of the grant funding project of the Ministry of Education and Science of the Republic of Kazakhstan No. 0120PK00254 "Prognostic assessment of the influence of clinical and genetic factors on the risk of development and the nature of the course of the disease in spontaneous intracerebral hemorrhage" Program code AP08957527.

Approbation of the dissertation research results. The main provisions of the dissertation were presented, reported, and discussed at:

1. IX Neuroforum "Week of Neurosciences in Kazan - 2023" / IX All-Russian Scientific and Practical Conference "Bekhterev Readings-2023" (Kazan, Russia, February 14–17, 2023), 1st degree diploma;
2. International Scientific and Practical Conference "Current Problems of Clinical, Experimental Neurology, Neurosurgery and Neurophysiology" (Almaty, Kazakhstan, May 26–27, 2023), 1st degree diploma;
3. Conference "Young Scientist" (Karaganda, Kazakhstan, May 27, 2023);
4. International Scientific and Practical Conference "Genetics and Physiology: Past, Present, Future" (Almaty, Kazakhstan, May 27–30, 2025).

Personal contribution of the author. The author independently conducted an analytical review of the literature, performed the collection and processing of primary clinical material, and the formation of an electronic database. Biological material was collected and molecular-genetic research was conducted (DNA extraction, genotyping of three polymorphisms of the *COL4A2* gene). Statistical analysis was personally performed and, based on the obtained data, a prognostic algorithm was developed (author's certificate No. 71562), scientific articles were prepared and published.

Implementation of the research results into practice. The results of the dissertation research have been implemented in the educational process of NPJSC "Karaganda Medical University", as well as in clinical practice (LLP "City Center of Primary Health Care"). The fact of application is confirmed by implementation acts.

Publications on the topic of the dissertation. On the topic of the dissertation, 4 works have been published, including 2 articles in journals indexed in the Scopus database (28th and 26th percentiles), 1 article in a journal recommended by the Committee for Quality Assurance in the Sphere of Science and Higher Education of the Ministry of Science and Higher Education of the Republic of Kazakhstan (List 2); 1 author's certificate of the Republic of Kazakhstan has been obtained.

Structure and volume of the dissertation. The dissertation is presented on 111 pages of typewritten text and consists of an introduction, three chapters (literature review; materials and methods; results and discussion), a conclusion, and a list of used sources, including 108 titles. The work is illustrated with 41 tables and 10 figures. The appendices contain 1 author's certificate and 3 implementation acts.

Design, materials and methods of the research

The study was conducted in two stages. The first stage was a retrospective epidemiological study of an extended cohort of patients with ICH of the Karaganda region for 2018–2020 (n=1657) with calculation of age-standardized incidence rates, hospital mortality, and population mortality depending on ethnicity. The second stage was a clinical observational case-control study (n=327; main group - 162 patients, control group - 165 people), conducted in 2020–2021 on the basis of medical organizations of the Karaganda region.

The main group included patients with spontaneous ICH verified by neuroimaging (CT/MRI), ethnic Kazakhs (at least two generations of ancestors of

Kazakh nationality), over 18 years of age, who signed informed consent. Exclusion criteria: traumatic genesis of ICH; ICH due to rupture of an aneurysm or arteriovenous malformation; ICH against the background of a brain tumor; secondary ICH against the background of anticoagulant therapy; severe autoimmune diseases. The control group consisted of individuals comparable in sex, age, and ethnicity without stroke, brain tumors, and autoimmune diseases in the anamnesis.

All participants had demographic data and modifiable risk factors (arterial hypertension, diabetes mellitus, alcohol consumption, smoking, obesity) recorded; for the main group, hematoma characteristics according to CT/MRI data were additionally recorded. The molecular-genetic study included extraction of genomic DNA (GeneJET Genomic DNA Purification Kit; quality control on NanoDrop 2000) and genotyping of three polymorphisms of the *COL4A2* gene (rs9521732, rs9521733, rs9515199) by real-time PCR (TaqMan SNP Genotyping Assay, QuantStudio 5, Applied Biosystems).

Statistical processing was performed in Statistica 13.3 (StatSoft) and R 4.2.0 (packages pROC, rms, caret, dcurves). Quantitative variables are presented as $M \pm SD$, categorical - as n (%); comparison was performed using Pearson's χ^2 test / Fisher's exact test, Mann–Whitney test. The association of polymorphisms with the risk of ICH was assessed by logistic regression in five inheritance models with correction for multiple comparisons (Benjamini–Hochberg method, FDR). Independent predictors were determined by multivariate logistic regression analysis with stepwise selection (stepwise, AIC); due to quasi-complete separation of data on arterial hypertension, Firth's penalized regression was used as the main multivariate model. Based on β -coefficients, a risk score scale was constructed. Discriminatory ability was assessed by ROC analysis (AUC, DeLong method, Youden index, cut-off threshold), calibration - by the calibration curve; clinical usefulness was assessed by Decision Curve Analysis (DCA) as a descriptive characteristic of the sample; improvement of reclassification - by NRI and IDI indicators. Internal validation was performed by bootstrap (1000 samples). The significance level was taken as $p < 0.05$.

Results

The epidemiological analysis of the extended cohort ($n=1657$) showed that after age standardization, the incidence of spontaneous ICH among ethnic Kazakhs was 72.8 versus 49.4 per 100,000 person-years among individuals of other ethnic groups (ratio 1.47; 95% CI 1.34–1.63); when comparing non-standardized indicators, no difference was detected (ratio 1.04). Hospital mortality among Kazakhs was lower - 24.0% versus 38.2% (odds ratio adjusted for age and sex 0.51; 95% CI 0.41–0.63), as a result of which standardized population mortality between the groups did not differ significantly (ratio 0.91; 95% CI 0.76–1.09).

The main and control groups of the case-control study were comparable in sex and age ($p > 0.05$). Arterial hypertension was detected in 100% of patients of the main group and in 24.0% of genotyped individuals of the control group, as a result of which its odds ratio in this design cannot be finally point-estimated (quasi-

complete separation of data). In the unadjusted analysis, diabetes mellitus was associated with an increase in the odds of ICH (OR = 3.11; 95% CI 1.19–8.09), moderate/heavy alcohol consumption - to a greater extent (OR = 4.43; 95% CI 2.12–9.27); in the multivariate model (Firth's penalized regression, without arterial hypertension), both factors retained significance: alcohol - OR = 4.49 (95% CI 2.09–10.30), diabetes mellitus - OR = 3.83 (95% CI 1.45–11.67). Smoking and obesity did not show a statistically significant association with risk ($p > 0.05$).

Among the three studied polymorphisms of the *COL4A2* gene, a statistically significant association with the risk of ICH was established only for rs9515199. The frequency of the C allele in the control group was 0.47 (which corresponds to East Asian indicators and exceeds European ones - 0.40) versus 0.30 in patients with ICH ($p < 0.001$). Carriage of the C allele is associated with a decrease in the odds of the disease: codominant model (C/C versus T/T) - OR = 0.23 (95% CI 0.11–0.47), dominant - OR = 0.45 (0.28–0.73), recessive - OR = 0.30 (0.16–0.58), log-additive - OR = 0.50 (0.36–0.70); the overdominant model did not reveal an association. After adjustment for diabetes mellitus, alcohol consumption, age, and sex, the association persisted almost unchanged (C/C versus T/T - OR = 0.29; 0.14–0.59; log-additive - OR = 0.55; 0.39–0.77) and remained significant after correction for multiple comparisons (Benjamini–Hochberg method), with the exception of the overdominant model. The direction of the association is directly opposite to European data, where the C allele is a risk factor (OR = 1.14–1.44).

Subgroup analysis by hematoma localization (deep ICH - 128 people, 79.0%; lobar - 18 people, 11.1%) did not reveal statistically significant differences in clinical and genetic parameters in a direct comparison of deep and lobar hemorrhages ($p > 0.05$ for all comparisons, including the distribution of rs9515199 genotypes - $p = 0.783$). When compared with the control group in the deep ICH subgroup, strong and statistically significant associations persisted with the rs9515199 genotype (T/T - 51.7% versus 31.2% in the control; C/C - 9.5% versus 26.0%; $p < 0.001$), alcohol consumption (23.4% versus 6.1%; $p < 0.001$), and diabetes mellitus (11.7% versus 3.6%; $p = 0.011$); in the multivariate model for this subgroup (without arterial hypertension), carriage of the C allele retained a protective association (OR = 0.454; 95% CI 0.266–0.772). In the lobar ICH subgroup ($n=18$), no statistically significant associations with the control group were detected for any of the studied factors ($p > 0.05$ for all comparisons, including genotype - $p = 0.451$), which is partly due to the small size of the subgroup.

Based on the identified predictors, a prognostic algorithm was developed (score scale and logistic regression model) integrating clinical (arterial hypertension, diabetes mellitus, alcohol) and genetic (rs9515199) factors. The full logistic model demonstrated AUC = 0.920 (95% CI 0.889–0.951) versus AUC = 0.908 (0.878–0.938) for the model without the genetic marker; the increase $\Delta\text{AUC} = 0.012$ did not reach statistical significance (DeLong test, $p = 0.137$). The score scale showed AUC = 0.895; at a cut-off threshold of ≥ 2 points, sensitivity was 100%, specificity - 68.8%; at a threshold of ≥ 3 points - 67.8% and 86.4%, respectively. The addition of the genetic marker rs9515199 did not lead to a statistically significant improvement in either discrimination or reclassification (continuous NRI = -0.234; IDI = 0.0025).

Bootstrap validation (1000 samples) confirmed the stability of the model: the adjusted AUC values (0.906 and 0.910) practically did not differ from the original ones. The calibration curve demonstrated good agreement between predicted probabilities and the observed frequency of ICH; at the same time, since in the case-control design the ratio of groups does not reflect the true prevalence of the disease in the population, predictive values, absolute risk calibration, and Decision Curve Analysis (DCA) are not population-valid without external prospective verification and are presented as descriptive characteristics of the sample (DCA showed a higher net benefit of the model compared to the strategies "treat all"/"treat none" in the range of threshold probabilities of 8–45%), and not as a justification for clinical implementation.

Conclusions

1. Differences in the incidence of spontaneous intracerebral hemorrhage between ethnic groups of the population of the Karaganda region are detected only after accounting for the age composition of the groups. The incidence for 2018–2020 was 72.8 per 100,000 person-years in Kazakhs and 49.4 in representatives of other ethnic groups (ratio 1.47; 95% CI: 1.34–1.63); when comparing crude rates, no difference is detected (ratio 1.04). In-hospital mortality in Kazakhs is lower — 24.0% vs. 38.2% (odds ratio adjusted for age and sex 0.51; 0.41–0.63), as a result of which population mortality does not differ between the groups (ratio 0.91; 0.76–1.09).
2. Modifiable risk factors for spontaneous intracerebral hemorrhage in the studied sample are arterial hypertension, diabetes mellitus, and moderate or heavy alcohol consumption. Hypertension was detected in all patients of the main group and in 25.45% of the control group, therefore its odds ratio is not assessed in this design. Diabetes mellitus increases the odds of the disease (OR = 3.11; 95% CI: 1.19–8.09), alcohol consumption to a greater extent (OR = 4.43; 2.12–9.27); when all factors are considered simultaneously, both remain significant (3.83 and 4.49, respectively). Smoking and obesity showed no association with the disease. The groups are comparable in sex and age.
3. The rs9515199 polymorphism of the *COL4A2* gene is associated with the risk of spontaneous intracerebral hemorrhage: carriage of the C allele is associated with lower odds of the disease (OR = 0.23 for the C/C genotype in the codominant model: 95% CI: 0.11–0.47; log-additive model OR = 0.50; 0.36–0.70). The association persists after adjustment for diabetes mellitus, alcohol consumption, age, and sex (0.29; 0.14–0.59) and remains significant after correction for multiple comparisons. The contribution of the polymorphism in addition to arterial hypertension cannot be assessed in this design. The frequency of the C allele in the control group was 0.47, which corresponds to East Asian populations and exceeds European indicators (0.40); the direction of the association is opposite to European data, where the same allele is considered a risk factor. The rs9521733 polymorphism showed no statistically significant association with the disease; the results for the rs9521732

polymorphism are preliminary and are not included in the evidence part of the work.

4. The association of the rs9515199 polymorphism of the *COL4A2* gene with the risk of spontaneous intracerebral hemorrhage was confirmed for deep localization (basal ganglia, thalamus): carriage of the C allele is associated with a reduction in the odds of the disease ($p = 1.96 \times 10^{-4}$). For lobar hemorrhages, no statistically significant association was obtained ($p = 0.451$), however, due to the small size of the subgroup ($n=18$), this result does not allow a final conclusion about differences in genetic effects between subtypes.
5. A classification model was developed that includes clinical (arterial hypertension, diabetes mellitus, alcohol consumption) and genetic (rs9515199) factors, and its internal discriminative ability was evaluated. The model distinguishes patients from control subjects with AUC = 0.895 for the point scale and 0.920 (95% CI: 0.889–0.951) for the logistic model; at a threshold of ≥ 2 points, sensitivity is 100%, specificity 68.8%. The addition of the genetic component did not significantly improve discrimination. External validation of the model was not performed.

LIST OF USED SOURCES

- 1 GBD 2021 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021 // *The Lancet Neurology*. - 2024. - Vol. 23, No. 10. - P. 973-1003.
- 2 An S.J., Kim T.J., Yoon B.W. Epidemiology, risk factors, and clinical features of intracerebral hemorrhage: an update // *Journal of Stroke*. - 2017. - Vol. 19, No. 1. - P. 3-10.
- 3 Zhakhina G., Zhalmagambetov B., Gusmanov A., Sakko Y., Yerdessov S., Matmusaeva E., et al. Incidence and mortality rates of strokes in Kazakhstan in 2014–2019 // *Scientific Reports*. - 2022. - Vol. 12, No. 1. - P. 16041.
- 4 Rannikmae K., Sivakumaran V., Millar H., Malik R., Anderson C.D., Chong M., et al. *COL4A2* is associated with lacunar ischemic stroke and deep ICH: meta-analyses among 21,500 cases and 40,600 controls // *Neurology*. - 2017. - Vol. 89, No. 18. - P. 1829-1839.
- 5 Wang H., Zhu Y., Zheng L., Chen M., Hao Z., et al. Association of the *COL4A2* gene polymorphisms with primary intracerebral hemorrhage risk and outcome in Chinese Han population // *Molecular Neurobiology*. - 2024. - Vol. 61, No. 11. - P. 8787-8796.
- 6 Hemphill J.C., Bonovich D.C., Besmertis L., Manley G.T., Johnston S.C. The ICH score: a simple, reliable grading scale for intracerebral hemorrhage // *Stroke*. - 2001. - Vol. 32, No. 4. - P. 891–897.