



DOI: 10.20514/2226-6704-2024-14-4-284-297

УДК [613.291:644.41]-071.1(470.311)

EDN: XLDFAZ

**Д.О. Драгунов^{1,2}, А.В. Соколова^{1,2}, Г.П. Арутюнов¹**

¹ — Федеральное государственное автономное образовательное учреждение высшего образования «Российский национальный исследовательский медицинский университет имени Н.И. Пирогова» Министерства здравоохранения Российской Федерации, Москва, Россия

² — Научно-исследовательский институт организации здравоохранения и медицинского менеджмента Департамента здравоохранения города Москвы, Москва, Россия

РАЗРАБОТКА И ВАЛИДАЦИЯ ОПРОСНИКА ДЛЯ ОЦЕНКИ УРОВНЯ ПОТРЕБЛЕНИЯ СОЛИ У ВЗРОСЛОГО НАСЕЛЕНИЯ МОСКВЫ И МОСКОВСКОЙ ОБЛАСТИ ПРИ ПОМОЩИ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

D.O. Dragunov^{1,2}, A.V. Sokolova^{1,2}, G.P. Arutjunov¹

¹ — N.I. Pirogov Russian State Autonomous University of Medical Sciences, Ministry of Health of Russia, Moscow, Russia

² — SBU «NIIOZMM» DZM, Moscow, Russia

Development and Validation of a Questionnaire to Assess the Level of Salt Intake in the Adult Population of the Russian Federation Using Machine Learning Methods

Резюме

Цель: разработать и валидировать инструмент для оценки уровня потребления соли у взрослого населения РФ. **Материал и методы.** Респонденты заполняли пищевые дневники, где учитывался тип приема пищи, ее объем и факт дополнительного досаливания блюда. Для статистической обработки полученных данных использовали язык R, версия — 4.2.1, среда разработки RStudio (пакеты ggplot2, ggpubr, dplyr, tidyverse, gtsummary, rstatix). **Результаты.** Всего в исследование был включен 271 респондент, медианный возраст которых составил 52 [20; 70] года. Было установлено, что основными факторами высокого потребления натрия является досаливание, потребление соленых продуктов, меньший уровень потребления кондитерских изделий, для низкого уровня потребления соли характерно более высокое потребление молочных продуктов. Тест согласованности Козна составил $\kappa=0,48$ 95 % ДИ (0,08; 0,08), значения альфы Кронбаха $\alpha=0,8$. При пороговом значении ≥ 12 баллов по данным опросника, опросник имеет чувствительность 85 % по сравнению с медианной оценкой по данным 3-дневного пищевого дневника. При пороговом показателе < 12 баллов опросник имеет специфичность 74 % с медианной оценкой по данным 3-дневного пищевого дневника. **Заключение:** Опросник СОЛЬ может быть использован для определения оценки уровня потребления соли.

Ключевые слова: потребление соли, пищевой опросник, потребление натрия, машинное обучение

Конфликт интересов

Авторы заявляют, что данная работа, её тема, предмет и содержание не затрагивают конкурирующих интересов

Источники финансирования

Авторы заявляют об отсутствии финансирования при проведении исследования

Соответствие принципам этики

Исследование одобрено локальным этическим комитетом ГБУЗ «Городская клиническая больница № 4 ДЗМ» (Павловская больница, протокол № 44 от 20.09.2020г.). Все участники подписали Информированное согласие

Благодарности

Мы выражаем благодарность участникам исследования, без которых исследование было бы невозможным. А также студентам педиатрического факультета ФGAOY BO PИИМУ им. Н.И. Пирогова Минздрава России

Статья получена 26.02.2024 г.

Одобрена рецензентом 21.05.2024 г.

Принята к публикации 16.07.2024 г.

Для цитирования: Драгунов Д.О., Соколова А.В., Арутюнов Г.П. РАЗРАБОТКА И ВАЛИДАЦИЯ ОПРОСНИКА ДЛЯ ОЦЕНКИ УРОВНЯ ПОТРЕБЛЕНИЯ СОЛИ У ВЗРОСЛОГО НАСЕЛЕНИЯ МОСКВЫ И МОСКОВСКОЙ ОБЛАСТИ ПРИ ПОМОЩИ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ. Архивъ внутренней медицины. 2024; 14(4): 284-297. DOI: 10.20514/2226-6704-2024-14-4-284-297. EDN: XLDFAZ

Abstract

Purpose: To develop and validate a tool to assess salt intake in the adult population of the Russian Federation. **Material and Methods:** Respondents filled out food diaries, where the type of food intake, its volume, and the fact of additional salting of the dish were taken into account. R language, version 4.2.1, RStudio development environment (packages ggplot2, ggpubr, dplyr, tidyverse, gtsummary, rstatix) were used for statistical processing of the obtained data. **Results:** A total of 271 respondents were included in the study, with a median age of 52 [20; 70] years. It was found that the main factors for high sodium intake were pre-salting, consumption of salty foods, lower intake in confectionery, low salt intake was characterized by higher consumption of dairy products. Cohen's consistency test was $\kappa=0.48$ 95 % CI (0.08; 0.08), Cronbach's alpha values $\alpha=0.8$. At a threshold score of ≥ 12 points on the questionnaire, the questionnaire had a sensitivity of 85 % compared with the median score from the 3-day food diary data. At a threshold score < 12 points, the questionnaire has a specificity of 74 % compared with the median score from a 3-day food diary. **Conclusion:** The SOLE questionnaire can be used to determine the estimated level of salt intake by the population, but for wider application in the territory of the Russian Federation additional validation by regions is required.

Key words: salt consumption, food questionnaire, sodium consumption, machine learning

Conflict of interests

The authors declare no conflict of interests

Sources of funding

The authors declare no funding for this study

Conformity with the principles of ethics

The study was approved by the local ethics committee of the State Budgetary Healthcare Institution «City Clinical Hospital No. 4 DZM» (Pavlovsk Hospital, protocol No. 44 of September 20, 2020). All participants signed the Informed Consent

Acknowledgments

We express our gratitude to the study participants, without whom the study would not have been possible. And also, to students of the pediatric faculty of the Federal State Autonomous Educational Institution of Higher Education Russian National Research Medical University named after. N.I. Pirogov of the Russian Ministry of Health.

Article received on 26.02.2024

Reviewer approved 21.05.2024

Accepted for publication on 16.07.2024

For citation: Dragunov D.O., Sokolova A.V., Arutjunov G.P. Development and Validation of a Questionnaire to Assess the Level of Salt Intake in the Adult Population of the Russian Federation Using Machine Learning Methods. The Russian Archives of Internal Medicine. 2024; 14(4): 284-297. DOI: 10.20514/2226-6704-2024-14-4-284-297. EDN: XLDFAZ

24hDR — 24-hour diet recall, CCF — chronic heart failure, XGBoost — eXtreme Gradient Boosting, BMI — body mass index, NHANES — National Health and Nutrition Examination Survey, RSA — Republic of South Africa

Introduction

Since the 1960s, attempts have been made to study dietary behaviours using diet questionnaires [1]. First questionnaires to study sodium intake were developed in the 1980s [2]. In 1982 Pietinen et al. developed a questionnaire to classify salt intake levels. The questionnaire consisted of five questions about salting and self-assessed salt consumption, as well as the frequency of eating the seven products containing the highest amounts of salt. The questionnaire was completed by 1,471 people aged 14 to 65 years old, who also underwent daily urinalysis in order to assess the amount of sodium in urine. The resulting data were used to develop a questionnaire called the

Salt Index. The correlation between the 24-hour sodium extraction with urine and salt intake was weak: $r = 0.18$ in men and $r = 0.20$ in women ($p < 0.001$). This analysis used validation based on 24-hour urine collection, which for a long time was a golden standard for the study of sodium consumption. However, taking into account the recent data on the sodium exchange physiology [4], it is essential to understand that natriuresis is not constant and can change under the impact of a number of factors, for instance, differences in the respondent's diet during the week and on weekends [5]. Besides, sodium deposits can be released as a result of a long-lasting low-salt diet, as seen in study MARS520 [6].

Therefore, it is advisable to validate the questionnaire with the help of a diet diary. The study of sodium intake using the SALT questionnaire [8] is optimal, because it allows to take into account only one nutrient and variability in sodium consumption during several days.

Therefore, the purpose of our study was to develop and validate a tool for the assessment of salt intake by the adult population in Russia.

Materials and Methods

The first step was to develop a database of 7,641 products with the known salt content per 100 g of product. Data were sourced from guidelines edited by I. M. Skurikhin, Yu. B. Bulanov, E. V. Novikov [9–12], as well as the data on nutrients and sodium intake from

open official sources provided by the manufacturers of products and fast food certified in Russia. Products were grouped by types into categories, which were used to compile a diet diary. The study design is presented in Figure 1. The diary recorded the meal type, extent and fact of additional salting of food. Extent of additional salting was based on 0.1 g of salt per one additional salting [13]. Products containing less than 50 mg of sodium per serving were excluded from analysis. The response rate took into account the intake during the last week and excluded two weekdays and one weekend [14], since the mean salt intake on weekends can increase by 8–14 % vs. weekdays [15]. Meals during festive periods were disregarded. Respondents completed their diaries strictly on a same day basis. During the observation and diary completion, all respondents did not adhere to any special or low-salt diets. Servings were measured by weighing

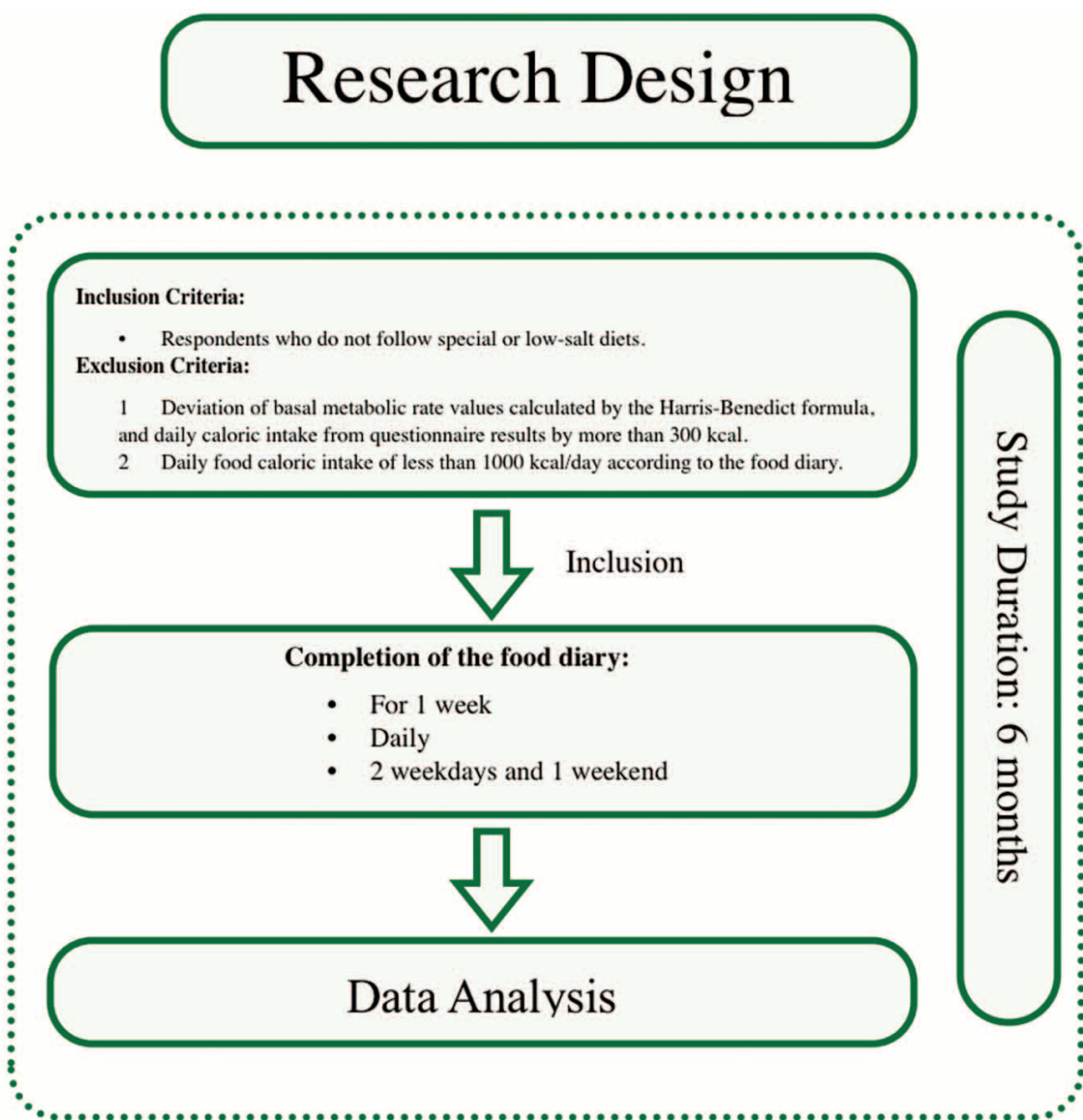


Figure 1. Study Design

the product or by package label (if any), or using an atlas of pictures of servings and food [14]. Subjects could were interviewed with the help of 24-hour diet recall, once per season. The data were collected using the weekday/week-end days ratio of 2 : 1 (patient follow-up/diary completion duration is not specified).

Development of the salt intake questionnaire

The purpose of the questionnaire was to have a tool to assess the median level of salt intake by respondents during the week. Therefore, immediately after questionnaire completion, the respondent was offered to complete a questionnaire of consumption of certain product groups or foods during the same period of time.

Products and foods were grouped by the content of sodium; the information on the content of salt was taken from the developed product and food database. All products were divided, taking into account categories and mean salt content in products and foods (e.g. Fish and seafood, etc.). All in all, there were 35 product categories.

Respondents were asked the following question: How often do you consume this type of products during the week? Depending on the frequency of consumption, responses were marked as “Never” — 0, “1–3 times a week” — 1, “4–6 times a week” — 2, “Once daily” — 3, “Twice daily” — 4, “More than three times daily” — 5. Additional salting was reported separately: an affirmative response added one more point.

Salt consumption rates used for data analysis

In order to validate the questionnaire result, it was compared to the median sodium intake calculated on the basis of the sum of three 24-hour diet diary results, which included information on salt consumption for each subject during two weekdays and one weekend day.

Salt intake by respondents was calculated using the formula:

$$z = \text{median} \sum_{d=1}^3$$

Where z is the median sum of salt consumption during three days, d is the day in the diet diary.

Analysis of the data obtained during various days of the week demonstrated a certain degree of variability in salt-containing product consumption; the median range for the sample was 4.2 [2.3; 7.3] g. Therefore, we used the median salt consumption to minimise individual variability of salt intake by each respondent. All patients who reported low salt consumption or special diets were excluded from the study and were disregarded in the analysis.

Machine learning was used to determine the product category and additional salting affecting salt consumption rates. We divided the sample into the teaching and test ones, then it was subjected to V-fold cross-check. The following algorithms were used: K mean method, random forest and eXtreme Gradient Boosting (XGBoost). Model optimality was evaluated using RMSE criteria and R^2 .

The sample size for the study was calculated with the formula based on the Bland–Altman limits of agreement [16]. According to the literature and pilot testing results, it was assumed that the expected mean difference between the questionnaire and 24-hour diet diary would be about 5 points per one gram of salt, which was partially based on the expected difference between the diet diary and questionnaire. Given some uncertainty in our estimates, possible mean difference and standard deviation, the coefficient of variation was 10 %, which ensured an acceptable degree of certainty that the estimated limits of agreement would be sufficiently reliable to confirm questionnaire results, with the 80 % power of the study. It corresponded to the minimal sampling size of 162 respondents for the evaluation of the lower Bland–Altman limit for the salt intake level. Taking into account possible exclusion of respondents due to partial compliance with the study protocol, the sampling size was to be 50 % larger than the calculated value and would be at least 243 people. This sampling size was adequate for an accurate evaluation of the Bland–Altman limits of agreement.

Statistics

Statistical processing of the data was performed using language R, v. 4.2.1, with the RStudio development framework. Normality of distribution was determined with the help of Shapiro-Wilk test and Kolmogorov-Smirnov test. Quantitative measures are presented as the mean value (M) $\pm$ standard deviation (S) or median, 25th and 75th percentile. When several groups were compared, Kruskal-Wallis test or analysis of variance were used. For comparison of the groups, the Student’s t-test was used for normal distribution, while the Wilcoxon’s test was used for non-normal distribution. Categorical variables were compared using Yates corrected χ^2 ; if the number of subjects in a group was below 5, then Fisher’s ratio test was used, followed by a post hoc analysis adjusted for multiple comparisons (Holm’s method). The relationships between variables were studied using Spearman’s rank correlation or Pearson correlation coefficient, depending on data distribution. Polynomial logistic regression was

used to test the association between categorial dependant variables with several categories. The Cronbach's alpha was calculated to determine the correspondence of data in the questionnaire. Questionnaire coherence was tested using the Cohen's kappa. The degree of difference between sodium intake estimates was visually assessed with the Bland–Altman plot. The zero hypothesis was discarded with the level of significance below 0.05.

Results

Clinical characteristics of the study group

Patient enrolment run from September 2021 till January 2023 in Moscow and Moscow region. The general cohort included 271 respondents: 220 (81 %) women and 51 (19 %) men. The mean age was 52 [20;70] years old. On average, women were younger than men (44.8 ± 26.2 years for women vs. 50.7 ± 25.1 years for men). Respondents below 30 years of age did not have any chronic conditions, whereas those over 30 years of age had various chronic diseases. The following diseases were diagnosed: chronic cardiac failure (CCF) was observed in 123 (45.4 %) subjects; 122 (45 %) patients were diagnosed with hypertension; 47 (17.3 %) had a history of myocardial infarction; atrial fibrillation and diabetes mellitus were diagnosed in 73 (26.9 %) and 39 (14.4 %) subjects, respectively. Lung diseases included bronchial asthma (13 (4.8 %) subjects) and chronic obstructive pulmonary disease (17 (6.3 %) subjects).

Quite a lot of patients over 30 years of age had chronic kidney disease (99 (36.5 %) respondents). Joint diseases were least numerous: 8 (3 %) patients (Table 1).

Salt intake variability in the groups

Based on the data distribution and study objectives, the following age groups were used: 18–40 years old, 41–70 years old, 71–80 years old and 80+ years old. It ensured a more detailed evaluation of various age categories and their impact on salt and calorie intake, as well as variability. Despite the fact that these age groups may not completely correspond to the generally recognised approaches used by the WHO, this distribution was chosen for a more accurate analysis of specific characteristics and consumer habits in each group in this study.

Table 2 shows salt and calorie intake and their variability in various age categories. There were no statistically significant differences in salt consumption in the age groups; the median average intake for three days was about 6 g. Also, calorie intake in the groups is not statistically different, unlike the expected basic metabolism, the mean value of which in all groups is below calorie intake. Body mass index is statistically lower in 18–40 years old group, then the median values are practically the same. Salt intake variability for three days recorded in the diet diary is not statistically different in the groups. The widest range in salt intake is observed in the groups of 18–40 and 41–70 years old, where variability varies from 20 to 50 %; the same trend is reported for

Table 1. Clinical characteristics of the group

Category	Value
Women	220 (81 %)
Men	51 (19 %)
Average age	52 [20;70] years
Average age of women	44.8±26.2 years
Average age of men	50.7±25.1 years
Chronic heart failure	123 (45.4 %)
Hypertensive disease	122 (45 %)
Myocardial infarction	47 (17.3 %)
Atrial fibrillation	73 (26.9 %)
Diabetes mellitus	39 (14.4 %)
Bronchial asthma	13 (4.8 %)
Chronic obstructive pulmonary disease	17 (6.3 %)
Chronic kidney disease	99 (36.5 %)
Joint diseases	8 (3 %)

Table 2. Salt intake, calorie consumption, and intake variability by age using a diary

Variables	All patients N=271 Me(IQR)	18 — 40 years n = 131 Me(IQR)	41 — 70 years n = 70 Me(IQR)	71 — 80 years n = 46 Me(IQR)	over 80 years n = 18 Me(IQR)	p-value
Average salt intake, g	6,2 (4,8; 8,5)	6,1 (4,7; 8,7)	6,4 (4,9; 8,4)	6,4 (4,7; 8,5)	6,0 (4,9; 8,1)	>0,9
Average calorie intake, calories	2119 (1765; 2604)	2095 (1725; 2592)	2178 (1819; 2650)	2120 (1846; 2526)	2080 (1703; 2623)	0,8
Salt intake to calorie ratio, mg/cal	2,93 (2,20; 3,76)	2,94 (2,19; 3,86)	2,89 (2,23; 3,60)	2,83 (2,15; 4,28)	3,04 (2,25; 3,29)	>0,9
Basal metabolism, calories	1434 (1378; 1501)	1434 (1380; 1540)*	1434 (1408; 1502)*	1434 (1400; 1434)**	1434 (1259; 1434)**	0,032
BMI, kg/m ²	23,2 (21,4; 25,1)	21,7 (19,7; 23,2)*	23,2 (23,2; 27,3)**	23,2 (23,2; 27,2)**	23,2 (23,2; 28,9)**	<0,001
Salt intake variability (%)	37 (27; 56)	37 (24; 58)	37 (29; 54)	39 (30; 57)	31 (16; 39)	0,2
Caloric intake variability (%)	24 (14; 34)	24 (16; 34)	22 (13; 34)	21 (16; 35)	20 (12; 34)	0,7
Gender						0,3
Female, n (%)	220 (81 %)	111 (85 %)	52 (74 %)	36 (78 %)	15 (83 %)	
Male, n (%)	51 (19 %)	20 (15 %)	18 (26 %)	10 (22 %)	3 (17 %)	

Notes: Data are presented as median and interquartile range (Me [IQR]) and n (%).
The p for trend is presented in the last column. Values for specific comparisons: *p <0.05 ** p <0.01 *** p <0.001

calorie intake. It appears that these variables are interrelated, and the higher the overall calorie intake with food, the higher the sodium intake; and this is especially true for younger subjects. Of note, despite its statistical insignificance, the ratio between salt and calorie intake is the highest in the age group of 80+ years old, it being a result of higher consumption of home-cooked preserves containing high amounts of salt and not so many calories.

Internal consistency check of the questionnaire

To test the consistency of the developed questionnaire, the internal consistency criterion was calculated; the Cronbach’s alpha was 0.8, i.e. evidencing a high degree of consistency in questions.

Questionnaire size reduction

Figure 2 shows five most commonly chosen products/product categories, distributed by the rate of consumption during the week. Figures indicate the median salt content in each product category, based on the diet diary per meal and taking into account the amount consumed. Products with extremely high median salt content were consumed not often; however, they could contain huge amounts of salt (e.g. Fish, the median value per meal was 5.3 g, i.e. the dietary reference intake per day). Also, it is important to take into account products containing moderate amount of sodium (about 1 g per meal, but taken quite often, i.e. at least once daily (e.g. bread,

0.6 g per meal, but three timely daily, i.e. 1.8 g per day from a single product)). Therefore, there is certain variability in consumption of products with high and moderate amounts of salt. That is why sodium intake should be considered on the weekly basis to reduce the degree of error.

When testing the degree of correlation between questions, several product categories were identified, which were highly correlated ($r > 0.3$). A number of categories with a high degree of correlation are mutually associated with products containing high amounts of sodium (e.g. fast food and pizza, etc.), therefore, some categories were combined in one category to reduce the number of questions. Of note, the pattern of high sodium intake is not observed in all categories. For instance, dairy products, soups (the median value is 0.4 and 0.7 g of salt per meal, respectively) are not associated with a high salt content. Thus, it is impossible to just combine these categories, because salt intake in this respondent category could differ. That is why we conducted an additional analysis using machine-aided cognition in order to identify product categories, which corresponded to the pattern of salt intake by respondents best of all.

Cluster analysis

To identify the main product categories in the questionnaire, which impacted the salt intake levels in patients, a cluster analysis with k-mean values was performed. Analysis of the scree plot showed that

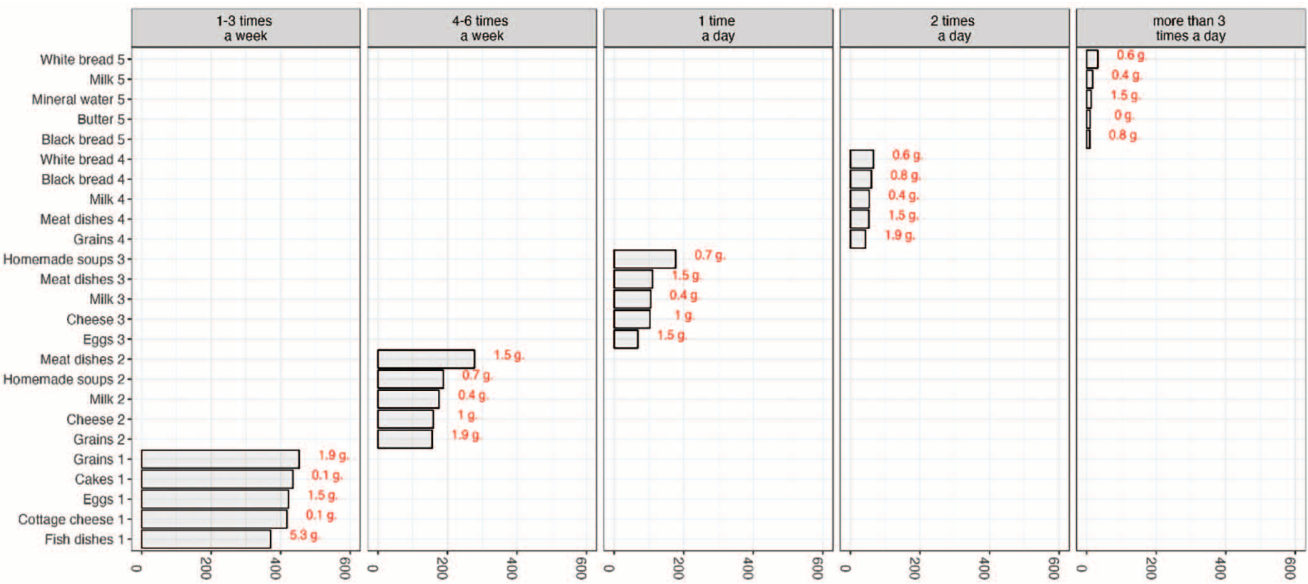


Figure 2. Frequency of food consumption (numbers indicate the frequency of food consumption by respondents: 1 — 1-3 times a week, 2 — 4-6 times a week, 3 — once a day, 4 — twice a day, 5 — more than 3 times a day)

the optimal number of clusters is three. The median intake in cluster 1 was 4.1 [2.9;5.1] g, in cluster 2 — 8.7 [7.6;10.6] g, and in cluster 3 — 17 [15.5;19.7] g ($p < 0.0001$). To identify the most important product categories in each cluster, a post hoc analysis was performed between each group, which identified products with statistically significant impact on salt intake values. Figure 3 shows the differences between clusters (A) and product groups (B), with statistically significant impact

on salt intake by respondents. Figures in Figure 3 show the mean frequency of intake (points). In other words, the main factors of high sodium intake are additional salting, consumption of salty products, lower consumption of confectionery (see “cakes” in the plot); lower salt consumption values are associated with higher dairy product consumption. Therefore, the ratio of the frequency of this or that product consumption can be used to predict median salt consumption.

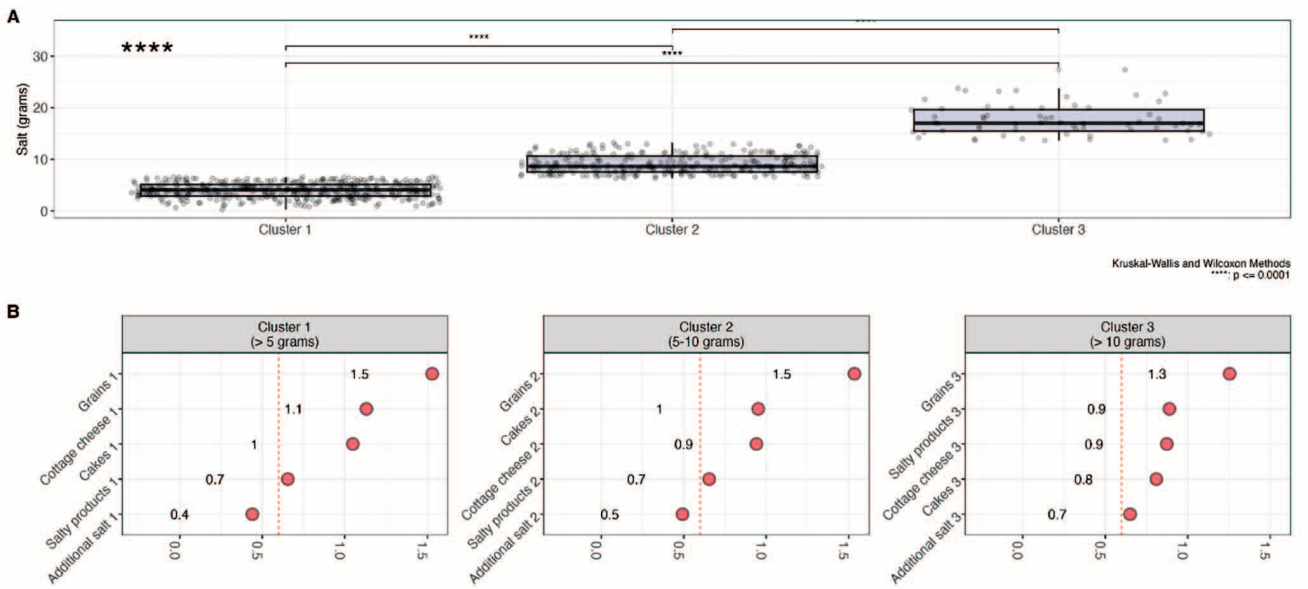


Figure 3. Comparative analysis of salt consumption in different clusters of respondents and the contribution of individual product categories

Algorithms of machine-aided learning

To select an optimal number of predictors, step-by-step linear regression plots were developed until an optimal model was identified. It resulted in a model with the following characteristics: $R^2 = 0.06$ (McFadden), $p < 0.0001$, and an optimal number of coefficients for forecasting. Table 3 shows the resulting optimal predictors. If compared with cluster analysis results, it is obvious that the majority of statistically significant predictors overlap. Also, the sign of the coefficient β can be

evaluated, which allows assessing the product category as increasing (positive coefficient β) or reducing salt intake (if coefficient β is negative).

Since it might be possible that the results did not have any linear dependence between variables, a method based on decision trees with a regression module was used. In Figure 4 (A), the resulting predictors are ranked depending on their degree of impact (incidence in clusters) on predicted salt intake. Figure 4 (B) shows first clusters of the resulting decision tree.

Table 3. The impact of various food categories on salt intake: results of regression analysis

Values	Coefficient (Beta)	Statistic	p-value
(Intercept)	6,83	13,41	p <= 0,0001
Adding salt	1,13	3,5	p <= 0,001
Cottage cheese	-0,6	-2,93	p <= 0,01
Fast food	0,89	3,32	p <= 0,001
Salty products	0,32	1,49	Not significant
Cakes	-0,42	-2,07	p <= 0,05
Lard	0,4	1,87	Not significant
Condensed milk	-0,6	-2,38	p <= 0,05
Groats	-0,44	-2,37	p <= 0,05
Canned foods	0,7	2,06	p <= 0,05
Meat dishes	0,27	1,54	Not significant
Powdered milk	1,36	1,5	Not significant

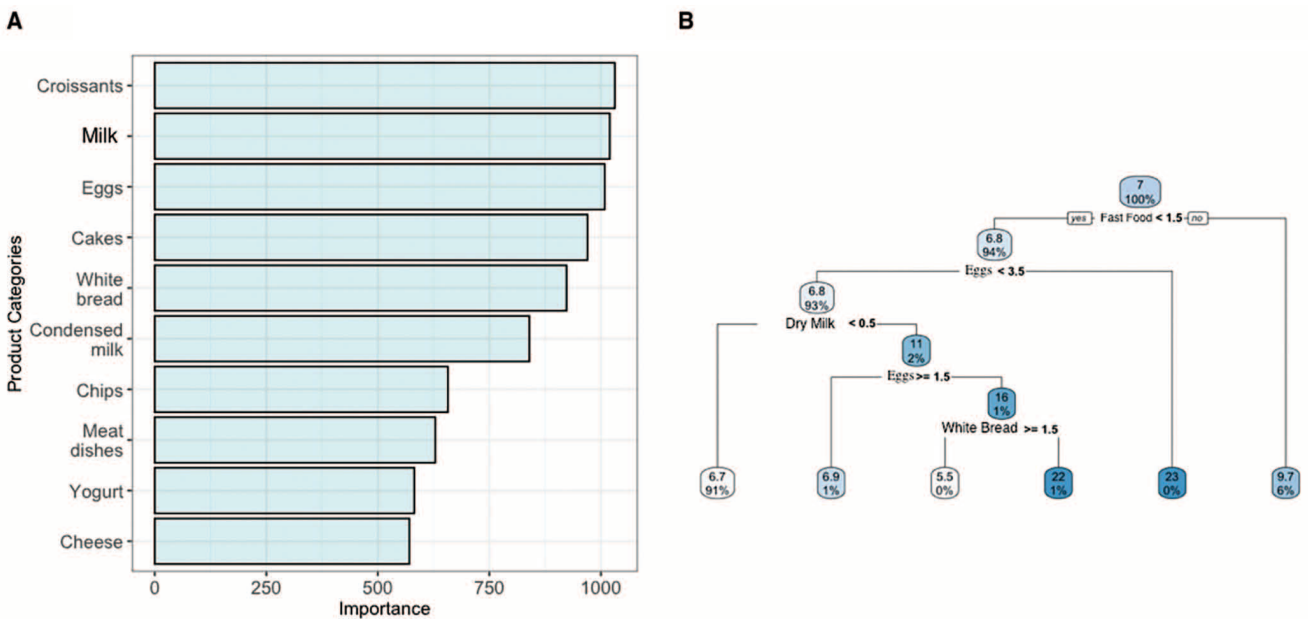


Figure 4. Decision trees. A — frequency of product categories in the node, B — pruned tree
Note. The decision tree divides the data starting with the condition “Fast food < 1.5” at the root node, where 97% of the samples fall into this branch. Then, the data is further split by the condition “Eggs < 3.5”, and subsequently by “Powdered milk < 0.5”. Each node shows the number of samples and the percentage of the total that meet the conditions. The terminal leaves, such as “6.7 (91%)”, represent data groups that cannot be further split. Nodes with zero values indicate the absence of samples meeting those conditions

Therefore, it is apparent that one of the main clusters is Fast Food consumption: if eaten more than once per week, salt intake is 9.7 g, and the share of such respondents is 6 %. At the same time, the level of correlation of the resulting model in a test sample was $r = 0.42$, $p = 0.0001$.

The most significant clusters in the two machine-aided learning models (random forest and XGBoost) were identified based on the results of test sample handling. The random forest model demonstrated a coefficient of correlation $r = 0.42$ with the confidence level of $p < 0.0001$, while XGBoost had a coefficient of correlation $r = 0.36$ and the confidence level of $p < 0.0001$. Although the forest model yielded better results in the test sample, a number of product categories overlap in these models.

We analysed product categories, which were most commonly used by various machine-aided learning algorithms to predict salt intake by respondents. Product categories used at least two times in predicting salt intake levels by various algorithms were divided into those with positive and negative impact on salt consumption. For example, the “Cakes” category reduced salt consumption, while “Fast food” category increased salt intake. The majority of algorithms used the “Cakes” category to predict salt intake by respondents. Only two algorithms used such categories as “Horse radish”, “Brown bread”, “Frankfurters”, “Lard”, “Preserves”, and “Yoghurt”. Therefore, the weight of these categories is much lower.

Increasing the questionnaire efficiency

We have optimised the questionnaire to make it more easy-to-use and efficient for routine use. First, we reduced the number of product categories to 10 to make the questionnaire more compact and easy-to-complete. We separated the following categories: “Dairy products”, “Confectionary”, “Eggs”, “Baked goods”, “Fast food”, “Home-cooked meals”, “Fish and seafood”, “Meat/poultry and meat/poultry products”, “Salty products/pickles/ mineral water”, “Additional salting”. We then optimised the frequency of responses for product consumption, by combining related categories. It allowed to specify what product categories impact salt intake and what product categories increase or reduce salt consumption. For instance, fast food products increase salt intake, while dairy products reduce it. Knowing these categories allowed us either to calculate coefficients for each category or apply inversion of points for a product category reducing the median salt intake levels during the day. Ignoring coefficients and point inversion, the correlation between points and salt intake was positive

and statistically significant; however, the correlation was weak ($r = 0.14$, $p = 0.0279$).

We chose to use the point inversion method as it offered the highest clarity for users. As a result, the correlation between the highest salt consumption by the respondent during the week and the total points was $r = 0.61$, $p < 0.0001$.

To test the association between the points resulting from questionnaire modifications, sodium intake levels were presented as “less than 5 g”, “5 to 10 g”, and “more than 10 g”, and a polynomial logistic regression model was developed. The resulting model has AIC = 191.9, OR = 2.38 [1.66;3.42] for intake of over 10 g; $p = 0.00001$ and OR = 1.6 [1.14;2.24] for intake of 5 to 10 g; $p = 0.0001$. Thus, the points in the questionnaire have statistically significant association with salt consumption by respondents and allow determining the points for various salt consumption levels: “less than 5 g” — 5.5 [4.25;6], “5 to 10 g” — 12.5 [10;14], “more than 10 g — 17 [15;18].

The resulting questionnaire titled **SALT** is presented in Appendix 1.

Testing validity and reliability of the SALT questionnaire

A Bland–Altman plot was generated to test the differences in the results obtained with the questionnaire and salt consumption values logged in respondents’ diet diaries (Fig. 5). To bring all changes to a single scale, a preliminary unifactor regression analysis was performed, with the dependent variable being salt and the predictor being points in the questionnaire ($R^2 = 0.12$, $p < 0.0001$). A relevant model was generated, which was then used to calculate predicted salt consumption in grams, depending on the points.

The bias is presented as the difference between sodium consumption in the SALT questionnaire and the median salt consumption over three days logged in diet diaries. The mean difference was 0 ± 4.36 at 95 % CI (-0.45; 0.45). Thus, satisfactory variability was observed between the optimised questionnaire results and actual salt intake, with a slight underestimate of sodium consumption. The plot shows that 95 % of the values were within the limits of agreement, although nine respondents were outside the limits of the two standard deviations in the mean difference, it being a result of extremely high salt intake in this group of respondents (21–27 g/day).

The Cohen’s kappa before an increase in the questionnaire efficiency was $\kappa = 0.004$ at 95 % CI (0.004; 0.004), after: $\kappa = 0.48$ at 95 % CI (0.08; 0.08), evidencing an average agreement and significant increase in agreement vs. the first variant. At the same time, the Cronbach’s alpha dropped to $\alpha = 0.38$ vs. the previous variant ($\alpha = 0.8$).

Weekly Salt Questionnaire (SALT Questionnaire)

Patient’s Full Name: _____

Date of Survey: “__” _____ 20__

	Never	Several times a week	Every day	More than once a day
Dairy Products ¹	0	2	1	1
Confectionery ²	0	2	1	1
Eggs and Egg Dishes	0	2	3	4
Bakery Products ³	0	2	3	4
Fast Food ⁴	0	2	3	4
Home-Cooked Food ⁵	0	2	1	1
Fish and Fish Dishes	0	2	3	4
Meat/Chicken and Meat/Chicken Dishes ⁶	0	2	3	4
Salty Products ⁷ / Pickles ⁷ / Mineral Water ⁸	0	2	3	4
Adding Salt	2			

- Note:**
- 1. Dairy Products:** cottage cheese, sour cream, milk of all kinds (excluding dry and condensed milk), including milk fruit juice, milkshakes, kefir, ryazhenka, cream, acidophilin, koumiss, yogurt (including yogurt drinks), butter, margarine.
 - 2. Confectionery:** cakes, muffins, pies, donuts, sweet buns, condensed milk, curd snacks, ice cream, wafers, croissants, bagels, cookies, lightly salted crackers, rusks, flatbreads, etc.
 - 3. Bakery Products:** white bread, black bread, gray bread, grain bread, pancakes.
 - 4. Fast Food:** chips, popcorn, regular crackers (salted), pizza, nuggets, burgers, French fries, rolls, sushi, etc.
 - 5. Home-Cooked Food:** food prepared at home, including homemade or restaurant soups, porridges, etc.
 - 6. Meat/Chicken and Meat/Chicken Dishes:** boiled sausage, smoked sausage, dry-cured sausage, bacon, salami, jerky, sausages, lard, pâté; meat/chicken dishes — steaks, minced meat, cutlets, meatballs, stews, lasagna, dumplings, etc.
 - 7. Salty Products:** all types of cheese, all types of semi-finished products, all types of canned goods, salted and pickled products (cucumbers, cabbage, including sauerkraut, olives, etc.), canned soups or soups made from dehydrated mixes, including kharcho soup, sauerkraut soup, salted nuts and seeds (including sunflower seeds), horseradish and salad dressings, including mayonnaise, soy sauce, ketchup, mustard, seasonings, tartar sauce, etc.
 - 8. Mineral Water:** Essentuki, Borjomi, Narzan, Dilijan, etc.

Result Interpretation

Salt Consumption Level	Number of Points
<5 grams/day	<7 points
6 — 10 grams/day	7 — 15 points
> 10 grams/day	> 16 points

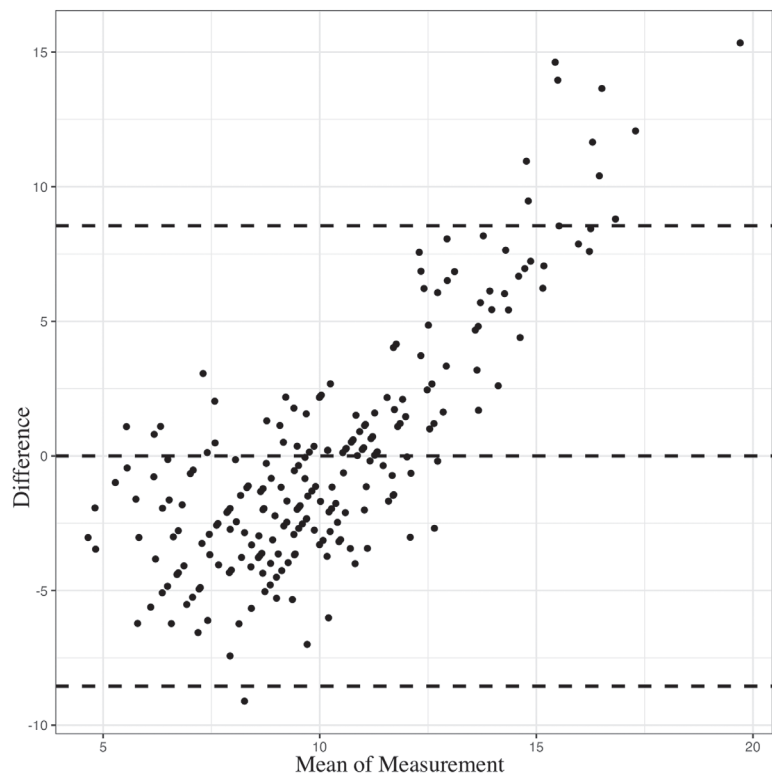


Figure 5. Bland-Altman plot of the systematic error (mean difference) between predicted sodium estimates from the SOL questionnaire and median dietary sodium intake from three days of questionnaire data

Note. The dotted line in the centre represents the mean difference between the two measurement methods. This value shows the systematic error, that is, the average amount by which one method differs from the other. The upper and lower dashed lines represent the limits of the 95% confidence interval for the mean difference. These lines show the range within which 95% of all differences between the two methods will lie

Sensitivity and specificity of the SALT questionnaire

At the threshold value of ≥ 12 points (corresponding to 7 g of salt per day) in the questionnaire, the sensitivity of the questionnaire is 85 % vs. the median estimate based on the three-day diet diary. At the threshold value of < 12 points, the specificity of the questionnaire is 74 % vs. the median estimate based on the three-day diet diary.

Discussion

The results of this study demonstrate that the *SALT* questionnaire can be a reliable tool for the assessment of the median sodium intake during the week. Some studies assessed eating habits in order to evaluate the daily salt intake. Mittelmark M. B. et al. [17] found out that 6 % of respondents did not use additional salt, although they stated otherwise. That is why salt consumption evaluation requires through statistical processing. In 1982, Pietinen et al. [2] used an approach based on statistical models to select products for the questionnaire. However, identification of the frequency of product consumption can be challenging. In this questionnaire, we used the questionnaire by Charlton K.E. et al.

[18], developed in the Republic of South Africa (RSA) and validated in three RSA populations. The assessment of salt intake was based on the mean daily consumption, taking into account coefficients used for the frequency of consumption. However, the authors noted that the coefficients limit the use of the questionnaire and simplified it before counting the points. The accuracy of the questionnaire was tested using the correlation with the sodium urine levels ($r = 0.171$) and diet diary ($r = 0.684$). The correlation with the questionnaire was higher, which can be attributed to the lack of sodium depot understanding in 2008. During data processing, we concluded that the number of gradations of possible answers on the frequency of sodium intake during the week was superfluous and could add to the errors in respondents' replies.

In this study, the salt intake variability was 4.2 [2.3; 7.3] g/day, whereas other authors mention fluctuations from 897 to 1,403 mg/day [19]. We used the median value instead of the mean salt consumption in order to reduce the impact from outliers and bring the results closer to the central trend for the respondent. In a study by Charlton K.E. et al. [18], only eight products with high sodium content (such as popcorn, meat, fish, etc.) significantly

correlated with the sodium urine levels, which can be associated with sodium excretion peaks. In a study by Sasaki S. et al. [20], the correlation between sodium excretion was low ($r = 0.14$ for men and $r = 0.23$ for women). Any attempts to increase the correlation between the questionnaire and sodium excretion did not yield any results. In a study by McLean R. M. [21], the correlation between the questionnaire and sodium excretion was low ($r = 0.257$ for the population, $r = 0.039$ for men and $r = 0.171$ for women). A systematic review [22] demonstrated that the satisfactory association between 24-hour sodium excretion and sodium intake levels was possible only with repeated daily urine tests. The most relevant (in terms of the rate of correlation between sodium intake and 24-hour sodium excretion) are the results of 7-day urine collection, as evidenced by the results of a study by Day N.E. et al. [23], where the correlation was $r = 0.47$, and it was lower than with nitrogen $r = 0.81$. This result is likely to be associated with sodium excretion peaks and sodium depots.

Studies, where patients are interviewed using the national product databases, show a higher linear correlation between questionnaire results and interview results. E.g. the Sodium Screener questionnaire containing 26 questions and validated in a study by Tangney C. [24] demonstrated the correlation $r = 0.83$ for men and $r = 0.85$ for women ($p < 0.001$) with the NHANES product database. However, of note, all validation methods have bias, and currently none of them is absolutely reliable [25].

According to the guidelines for the development of diet questionnaires for the population, all questionnaires must be verified in a sample of the population, in which they are planned to be used [13]. Our questionnaire was compared to the 24-hour diet diary, which is considered one of the most accurate methods. We decided to gather information for three days, because this period of time was used in other studies of sodium intake. For the assessment of sodium intake, we included common products and foods, as well as processed products containing higher amounts of sodium. However, the use of questionnaires validated in other countries is limited by the lack of national food, typically consumed by the study population. Therefore, our questionnaire is the only questionnaire in Russia to assess the sodium intake levels.

We used a machine-aided learning method to assess the accuracy of the SALT questionnaire; it allowed us to increase the sample size and check the results multiple times (10 random samples generated with the help of a V-fold cross-check, equivalent to 7,000 patients). This method helped in identifying the most important

product categories to assess sodium intake. We also used diet diaries to check the questionnaire results, because it allows to assess consumption during a longer period of time (days and weeks) and is independent of sodium depots [26] and sodium excretion peaks [27]. In this study, we used an eating behaviour pattern for the first time to develop a questionnaire, and this is an innovative approach.

Conclusions

This study demonstrated that the SALT questionnaire can be an acceptable tool for the assessment of sodium intake in Russia. We identified product categories, which correspond to the typical patterns of salt consumption in Russia, and the questionnaire demonstrates acceptable agreement with the data from the diet diary.

Study limitations

This questionnaire shows the categorial rate of salt intake in Moscow and Moscow Region and does not take into account diets in other regions of Russia. Additional validation is required in order to use this questionnaire in other regions of Russia.

Вклад авторов:

Все авторы внесли существенный вклад в подготовку работы, прочли и одобрили финальную версию статьи перед публикацией

Драгунов Д.О.: разработка концепции и дизайна исследования, определение его цели и задачи, а также методов сбора, анализа и интерпретации данных; статистическая обработка данных и машинное обучение; участие в сборе данных, их анализ и формирование выводов; написание рукописи, обоснование выводы и представление их в контексте предыдущих исследований

Соколова А.В.: разработка концепции и дизайна исследования; определение его цели и задачи, а также методов сбора, анализа и интерпретации данных; участие в сборе данных, их анализе и формировании выводов, а также в написании рукописи

Арутюнов Г.П.: создание концепции и дизайна исследования; определение его цели, задач и методов сбора; анализ и интерпретация данных

Author Contribution:

All the authors contributed significantly to the study and the article, read and approved the final version of the article before publication

Dragunov D.O.: The author makes a significant contribution to the concept and design of the study, defining its goals and objectives, as well as methods of data collection, analysis and interpretation. He actively participates in data collection, their analysis and formation of conclusions, as well as in writing the manuscript, justifying his conclusions and presenting them in the context of previous studies

Sokolova A.V.: The author plays a role in the conception and design of the study, defining its goals and objectives, as well as methods of data collection, analysis and interpretation. He is actively involved in the collection of data, their analysis and formation of conclusions, as well as in the writing of the manuscript

Arutjunov G.P.: The author plays a role in creating the concept and design of the study, defining its goals, objectives, and methods of data collection, analysis, and interpretation

Список литературы/ References:

- Hankin JH, Reynolds WE, Margen S. A Short Dietary Method for Epidemiologic Studies 123: II. Variability of Measured Nutrient Intakes. *The American Journal of Clinical Nutrition*. 1967; 20(9): 935–45. doi: 10.1093/ajcn/20.9.935
- Pietinen P, Tanskanen A, Tuomilehto J. Assessment of sodium intake by a short dietary questionnaire. *Scandinavian Journal of Social Medicine*. 1982; 10(3): 105–12. DOI: 10.1177/140349488201000307
- Titze J, Lang R, Ilies C, et al. Osmotically inactive skin Na⁺ storage in rats. *American Journal of Physiology-Renal Physiology*. 2003; 285(6): F1108–17. doi: 10.1152/ajprenal.00200.2003
- Kopp C, Linz P, Hammon M, et al. Seeing the sodium in a patient with hypernatremia. *Kidney International*. 2012; 82(12): 1343–4. doi: 10.1038/ki.2012.314
- Iaccarino Idelson P, D'Elia L, Cairella G, et al. Salt and health: Survey on knowledge and salt intake related behaviour in Italy. *Nutrients*. 2020; 12(2): 279. doi: 10.3390/nu12020279
- Rakova N, Jüttner K, Dahlmann A, et al. Long-term space flight simulation reveals infradian rhythmicity in human Na⁺ balance. *Cell Metabolism*. 2013; 17(1): 125–31. doi: 10.1016/j.cmet.2012.11.013
- Okada C, Takimoto H, Shimomura T. Development of a screening method for determining sodium intake based on the Dietary Reference Intakes for Japanese, 2020: A cross-sectional analysis of the National Health and Nutrition Survey, Japan. *PLOS ONE*. 2020; 15(9): e0235749. doi: 10.1371/journal.pone.0235749
- Cooper M, Simpson JR, Klutka R. Development and validation of a sodium Analysis tool (SALT). *Nutrition Journal*. 2020; 19(1). doi: 10.1186/s12937-020-00555-7
- Скурихин И.М., Тутельян В.А. Химический состав российских пищевых продуктов. Москва: ДеЛи принт. 2002; 235 с. Skurihin I.M., Tutelyana V.A. Chemical composition of Russian food products. Moscow: DeLi print. 2002; 235 p. [In Russian]
- Скурихин И.М., Волгарев М.Н. Химический состав пищевых продуктов: Справочные таблицы содержания основных пищевых веществ и энергетической ценности пищевых продуктов. 2-е изд., перераб. и доп. Москва: Агропромиздат. 1987; 224 с. Skurihin IM, Volgarev MN. Chemical composition of food products: Reference tables for the content of basic nutrients and energy value of food products. 2nd ed. Moscow: Agropromizdat. 1987; 224 p. [In Russian]
- Буланов Ю.Б. Химический состав продуктов: Пищевая ценность. Тверь: ГУПТО. 2003; 16 с. Bulanov Y.B. Chemical composition of products: Nutritional value. Tver: GUPTO. 2003; 16 p. [In Russian]
- Новикова Е.В., Черевко А.И. Энциклопедия питания. Нутриенты пищевых продуктов. т. 2 Москва: КноРус. 2023; 125 с. Novikova E.V., Cherevko AI. Encyclopedia of nutrition. Nutrients of food products. Vol. 2 Moscow: KnoRus. 2023; 125 p. [In Russian]
- Mason B, Ross L, Gill E, et al. Development and validation of a dietary screening tool for high sodium consumption in Australian renal patients. *J Ren Nutr*. 2014 Mar; 24(2): 123–34. e1–3. doi: 10.1053/j.jrn.2013.10.004. Epub 2014 Jan 3.
- Cade J, Thompson R, Burley V, et al. Development, validation and utilisation of food-frequency questionnaires: A review. *Public Health Nutrition*. 2002; 5(4): 567–87. doi: 10.1079/phn2001318
- Nowson C, Lim K, Land M-A, et al. Salt intake and dietary sources of salt on weekdays and weekend days in Australian adults. *Public Health Nutrition*. 2018; 21(12): 2174–82. doi: 10.1017/s1368980017004104
- Martin Bland J, Altman Douglas G. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*. 1986; 327(8476): 307–10. doi: 10.1016/s0140-6736(86)90837-8
- Mittelmark MB, Sternberg B. Assessment of salt use at the table: Comparison of observed and reported behavior. *Am J Public Health*. 1985 Oct; 75(10): 1215–6. doi: 10.2105/ajph.75.10.1215..
- Charlton KE, Steyn K, Levitt NS, et al. Development and validation of a short questionnaire to assess sodium intake. *Public Health Nutrition*. 2008; 11(1): 83–94. doi: 10.1017/S1368980007000146
- Luft FC, Aronoff GR, Sloan RS, et al. Intra- and Interindividual Variability in Sodium Intake in Normal Subjects and in Patients With Renal Insufficiency. *American Journal of Kidney Diseases*. 1986; 7(5): 375–80. doi: 10.1016/s0272-6386(86)80085-3
- Sasaki S, Yanagibori R, Amano K. Validity of a self-administered diet history questionnaire for assessment of sodium and potassium. *Japanese Circulation Journal*. 1998; 62(6): 431–5. doi: 10.1253/jcj.62.431
- McLean RM, Williams SM, Te Morenga LA, et al. Spot urine and 24-h diet recall estimates of dietary sodium intake from the 2008/09 New Zealand Adult Nutrition Survey: A comparison. *European Journal of Clinical Nutrition*. 2018; 72(8): 1120–7. doi: 10.1038/s41430-018-0176-0
- McLean RM, Farmer VL, Nettleton A, et al. Assessment of dietary sodium intake using a food frequency questionnaire and 24-hour urinary sodium excretion: A systematic literature review. *Journal of Clinical Hypertension (Greenwich, Conn)*. 2017; 19(12): 1214–30. doi: 10.1111/jch.13148
- Day N, McKeown N, Wong M, et al. Epidemiological assessment of diet: A comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *International Journal of Epidemiology*. 2001; 30(2): 309–17. doi: 10.1093/ije/30.2.309

24. Tangney CC, Rasmussen HE, Richards C, et al. Evaluation of a Brief Sodium Screener in Two Samples. *Nutrients*. 2019; 11(1): 166. doi: 10.3390/nu11010166
25. Souza DS, Santos BI, Costa BM, et al. Food frequency questionnaire for foods high in sodium: Validation with the triads method. *PLOS ONE*. 2023; 18(7): e0288123. doi: 10.1371/journal.pone.0288123
26. Artyukov I, Bukreeva I, Feshchenko R, et al. The first observation of osmotically neutral sodium accumulation in the myocardial interstitium. *Scientific Reports*. 2021; 11(1). doi: 10.1038/s41598-021-02956-y
27. Драгунов ДО, Арутюнов ГП, Соколова АВ. Современный взгляд на обмен натрия. *Клиническая Нефрология*. 2018; (1): 62–73. doi: 10.18565/nephrology.2018.1.62-73
Dragunov DO, Arutyunov GP, Sokolova AV. Modern view of sodium metabolism. *Klinicheskaya Nefrologiya*. 2018; (1): 62–73. [In Russian]

Информация об авторах

Драгунов Дмитрий Олегович  — к.м.н., доцент, доцент кафедры пропедевтики внутренних болезней ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России, заведующий ОМО по терапии ГБУ «НИИОЗММ» ДЗМ, Москва, e-mail: tamops2211@gmail.com, ORCID ID: <https://orcid.org/0000-0003-1059-8387>

Соколова Анна Викторовна — к.м.н., доцент, доцент кафедры пропедевтики внутренних болезней ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России, ведущий специалист ОМО по терапии ГБУ

«НИИОЗММ» ДЗМ, Москва, e-mail: sokolova2211@gmail.com, ORCID ID: <https://orcid.org/0000-0003-0823-9190>

Арутюнов Григорий Павлович — д.м.н., профессор, член-корр. РАН, зав. кафедры пропедевтики внутренних болезней ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России, Москва, e-mail: arut@ossn.ru, ORCID ID: <https://orcid.org/0000-0002-6645-2515>

Information about the authors

Dmitry O. Dragunov  — Candidate of Medical Sciences, Associate Professor, Associate Professor of the Department of Propaedeutics of Internal Diseases, N.I. Pirogov Russian State Autonomous University of Medical Sciences, Ministry of Health of Russia, Head of OMO for Therapeutics, SBU «NIIOZMM» DZM, Moscow, e-mail: tamops2211@gmail.com, ORCID ID: <https://orcid.org/0000-0003-1059-8387>

Anna V. Sokolova — Candidate of Medical Sciences, Associate Professor, Associate Professor of the Department of Internal Medicine Propaedeutics, N.I. Pirogov Russian National Research Institute of Medical Sciences, Russian Ministry of Health, leading specialist of OMO on therapeutics of SBU «NIIOZMM» DZM. Moscow, e-mail: sokolova2211@gmail.com, ORCID ID: <https://orcid.org/0000-0003-0823-9190>

Grigoriy P. Arutjunov — Doctor of Medical Sciences, Professor, Corresponding Member of the Russian Academy of Sciences, Head of the Department of Internal Medicine Propaedeutics, N.I. Pirogov Russian National Research Institute of Medical Sciences, Russian Ministry of Health. Moscow, e-mail: arut@ossn.ru, ORCID ID: <https://orcid.org/0000-0002-6645-2515>

 Автор, ответственный за переписку / Corresponding author