

# Алгоритм фильтрации и поиска зависимостей при диагностике пациентов с доумеренным когнитивным снижением\*

Н. Н. Яхно<sup>1,II</sup>, В. Н. Гридин<sup>I</sup>, Н. Н. Коберская<sup>1,II</sup>, Д. С. Смирнов<sup>I</sup>, В. И. Солодовников<sup>I</sup>

<sup>I</sup>ФГБУН Центр информационных технологий в проектировании РАН, Московская обл., г. Одинцово, Россия

<sup>II</sup>ФГАОУ ВО Первый Московский государственный медицинский университет имени И.М. Сеченова Министерства здравоохранения Российской Федерации (Сеченовский университет), Москва, Россия

**Аннотация.** Рассмотрены алгоритмы и методы статистической обработки и преобразования медицинских данных пациентов. Предлагается методика выявления групп показателей, и отбора ключевых наблюдений, обладающих наибольшей согласованностью и разделительной способностью для априорных групп пациентов. Применяется комбинаторный канонический анализ для поиска наиболее важных факторов риска при когнитивном снижении.

**Ключевые слова:** сокращение размерности пространства, многомерные статистические методы, канонический анализ, когнитивные нарушения,

DOI 10.14357/20718632210204

## Введение

С увеличением продолжительности жизни когнитивные расстройства, достигающие степени деменции (самой частой причиной которых является болезнь Альцгеймера), стали серьезной медицинской и социально-экономической проблемой, что потребовало объединения усилий медицинских и исследовательских учреждений в определении адекватных и быстрых методов их диагностики. Поскольку эффективное лечение когнитивных нарушений на стадии деменции крайне затруднительно, более успешными будут лечебное воздействие и их профилактика до наступления слабоумия. Однако в выявлении этих состояний

и определении тактики ведения пациентов есть еще много нерешенных вопросов.

Деменции предшествует состояние «умеренное когнитивное расстройство» (УКР) (англ., *mild cognitive impairment*) [1], которое характеризуется объективным клинически значимым снижением когнитивных функций, при отсутствии выраженных нарушений повседневной деятельности, то есть деменции [2-6].

Состоянию УКР предшествуют «легкое когнитивное снижение» (ЛКС), когда клинико-психологическое обследование выявляет признаки снижения когнитивных функций, не достигающего степени УКР, и «субъективное когнитивное снижение» (СКС), при наличии субъективных жалоб, например, на снижение

\* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект № 19-29-01112 мк).

памяти при отсутствии объективных признаков такового. Эти состояния объединяются под названием «доумеренное когнитивное снижение» (ДУКС), проявляющиеся снижением когнитивных функций, которое находит свое отражение в жалобах пациента и выявляется клинически при выполнении сложных нейропсихологических тестов (показатели скрининговых шкал у этих пациентов остаются в нормальных границах) или диагностируются при наличии когнитивных жалоб и отсутствии объективных признаков когнитивного снижения [6-8].

Активно обсуждается роль сердечно-сосудистых факторов в развитии когнитивных нарушений. В наших предыдущих публикациях анализ возможного влияния сосудистых факторов риска при ЛКС показал их связь с ухудшением, главным образом, управляющих (регуляторных) функций [9, 10]. Это согласуется с известными данными о характере когнитивной дисфункции при цереброваскулярных расстройствах.

#### **Исходные данные**

В данной работе рассматривалась выборка из 447 пациентов (323 женщин и 124 мужчин) в возрасте от 50 до 75 лет. Из них 174 человека — лица с легким когнитивным снижением без сердечно-сосудистых факторов риска (ЛКС без ССФР), 87 человек — лица с субъективным когнитивным снижением без сердечно-сосудистых факторов риска (СКС без ССФР), 144 человека — лица с легким когнитивным снижением с сердечно-сосудистыми факторами риска (ЛКС с ССФР), 15 человек — лица с субъективным когнитивным снижением с сердечно-сосудистыми факторами риска (СКС с ССФР) и 27 здоровых людей.

По возрасту, гендерному соотношению и уровню образования группы пациентов и контрольная группа между собой достоверно не различались. Обследование включало общее клиническое, неврологическое и клинико-психологическое исследования с использованием количественных шкал и тестов - краткой шкалы оценки психического статуса (КШОПС), шкалы оценки лобной дисфункции (ШОЛД), исследование литеральных и категориальных ассоциаций, теста последовательного соединения цифр, цифр и букв (Trail making test, parts A,B), теста рисования часов, теста повторения

цифр в прямом и обратном порядке и оценки зрительной памяти по тесту Бентона.

Медицинские аспекты исследования данной выборки и первичные результаты статистической обработки данных изложены в нашей предыдущей работе [11]

Как было показано нами ранее [12], использование большого количества нейропсихологических тестов не позволяет получить согласованного статистического разбиения пациентов на группы по тяжести заболевания, а, следовательно, получить агрегированную количественную оценку степени когнитивного снижения. Такая оценка могла бы быть использована как результирующая переменная для предиктивных или регрессионных моделей, прогнозирующих степень когнитивного снижения, а также выделяющая главные соматические (сердечно-сосудистые и другие) факторы, влияющие на этот процесс. В том же исследовании был предложен метод обработки данных пациентов в целях обнаружения факторов, влияющих на когнитивные нарушения и свертки пространства нейропсихологических признаков. Однако в рамках текущей выборки с пятью априорными группами, предложенный ранее метод требует слишком большого объема вычислений, поэтому был предложен более простой метод свертки пространства.

Также, в связи с тем, что в исследуемых данных физиологические переменные представлены как качественные переменные, требуется использовать иные методы анализа эффективности разделения групп пациентов (в первом исследовании применялся метод MANOVA и тест Краскела-Уоллиса).

#### **Цель исследования**

Разработка алгоритма поиска статистически значимых зависимостей по результатам обследования пациентов при анализе «доумеренных» когнитивных расстройств при наличии зашумленных и слабо согласованных данных.

#### **Используемые понятия и допущения**

Априорная группа — это метка для пациента, которая была сделана в рамках медицинского обследования. Номера групп представлены в следующем порядке ЛКС без СФР, СКС без СФР, группа контроля, СКС с СФР, ЛКС с СФР, для максимального разделения групп

с наличием/отсутствием сердечно-сосудистых факторов рисков (см. параграф Апостериорный анализ), т.к. в целом группы пациентов с СФР/без СФР, разделяются слабее, чем группы пациентов с ЛКС/СКС.

**Применяемые теоретические и инструментальные методы и средства**

Были использованы язык программирования для статистической обработки данных и свободная программная среда вычислений с открытым исходным кодом R, версии 3.6.1.

**Результаты**

Решение задачи проводилось по алгоритму, изображенному на Рис. 1. Для нахождения подмножеств физиологических и нейропсихологических признаков, имеющих наибольшее влияние друг на друга, последовательно выполняется несколько анализов: комбинаторный анализ – для выделения лучших комбинаций нейропсихологических признаков, апостериорный анализ – для отбора «ядер» групп пациен-

тов, и канонический анализ – для определения статистических связей между отобранными нейропсихологическими признаками и группами соматических признаков в «ядерном» множестве пациентов.

**Комбинаторный анализ**

Для обобщенной оценки когнитивных функций мы ввели «интегральный показатель степени когнитивного снижения (ИПСКС)», который был получен при комбинаторном анализе по выявлению наиболее информативного набора нейропсихологических показателей пациентов. Для этого рассматривались все возможные неповторяющиеся комбинации нейропсихологических показателей, для которых рассчитывался набор характеристик, оценивающих эффективность разбиения на группы пациентов с ДУКС с ССФР, с ДУКС без ССФР и группу контроля.

Для каждой комбинации были вычислены следующие показатели: Среднее расстояние (*average distance*) [13], Компактность кластеров

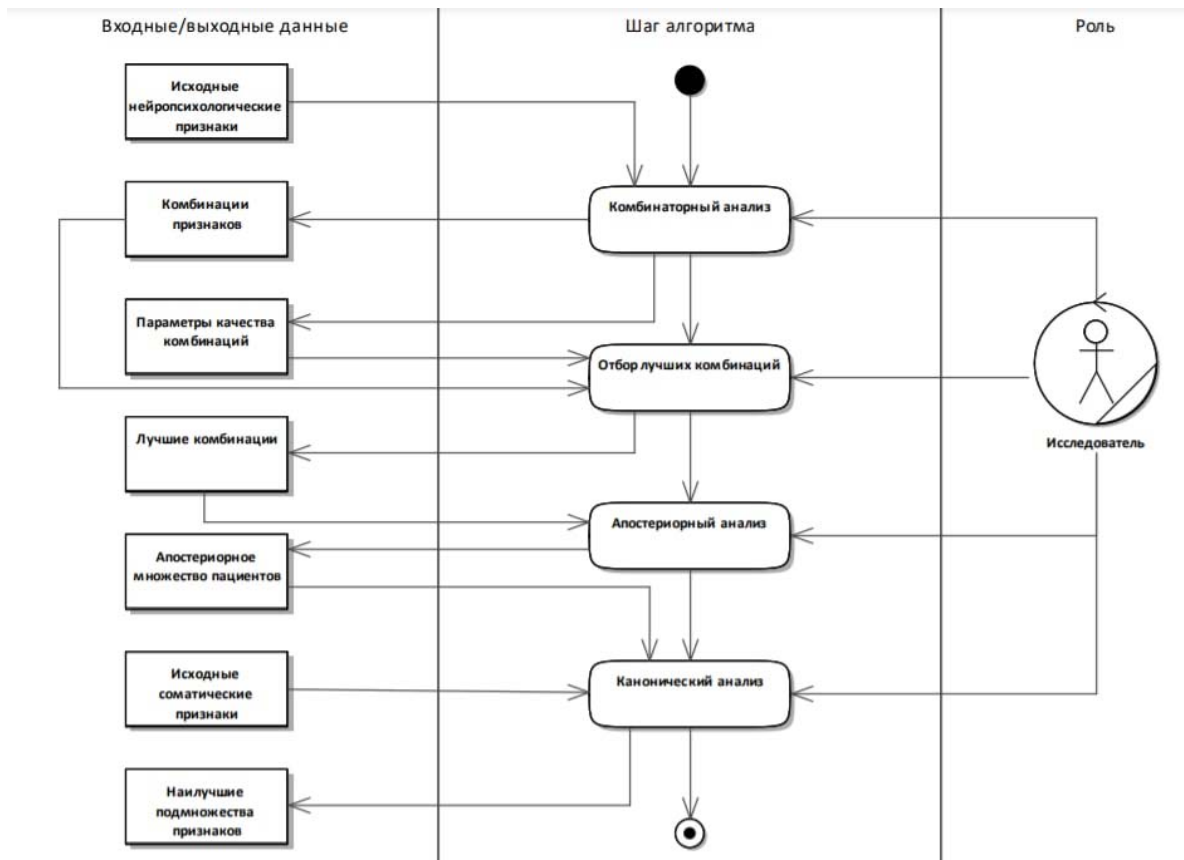


Рис. 1. Алгоритм приводимого анализа

(*Cluster Cohesion*) [13], Отделимость кластеров (*Cluster Separation*) [13], Средняя ширина силуэта (*Average Silhouette Width*) [13], Коэффициент конкордации Кендалла (коэффициент множественной ранговой корреляции) [14]

Приведенные статистические показатели позволяют оценить, насколько устойчиво и статистически значимо разделение пациентов на вышеупомянутые группы по данному набору нейропсихологических показателей, а также, насколько в целом нейропсихологические тесты, входящие в комбинацию, информативны и согласованы.

Стоит отметить, что средняя ширина силуэта обладает наименьшей вариацией значений и ее максимум невелик (значительно меньше оптимальных значений для данного показателя). Это объясняется в первую очередь тем, что по всем вариантам разделения сложно отделить группы с ДУКС при ССФР от групп с ДУКС без ССФР.

Далее, среди всех рассматриваемых комбинаций отбирались только те, у которых значения по вычисленным параметрам превосходят верхнюю квартиль (меньше нижней квартили, для компактности кластера и внутикластерного расстояния). Отобранные комбинации ранжировались в порядке возрастания (убывания, для компактности кластера и внутикластерного расстояния). В ранжированной таблице было оставлено 10% от общего числа комбинаций, среди которых отобрано 5 наиболее часто

встречающихся комбинаций — т.е. наилучших, с точки зрения многокритериального подхода.

Таким образом, отбиралась комбинация нейропсихологических тестов, которая обеспечивала бы наиболее четкое разделение групп, схожесть пациентов внутри одной группы, и согласованность между собой тестов, входящих в комбинацию. Среди всех показателей наиболее информативными оказалось сочетание тестов, представленных в Табл. 1.

Также как и в предыдущем нашем исследовании [12], первоначальная идея заключалась в формировании «интегрального показателя степени когнитивного снижения» (ИПСКС) для обобщенной оценки когнитивных функций, как результат применения метода главных компонент к отобранным комбинациям нейропсихологических тестов.

На Рис. 2 представлена визуализация пациентов в пространстве главных компонент одной из комбинаций. Маркерами разных типов обозначены различные априорные группы пациентов. Более крупными маркерами – центры соответствующих групп. Можно заметить, что центры различных групп логично распределены по главной компоненте (наибольшее значение – группа контроля, наименьшие значения – группы с легким когнитивным снижением), однако отдельные представители групп сильно перемешаны между собой, поэтому для дальнейшего анализа необходимо нивелировать этот феномен.

Табл. 1. Наиболее информативные комбинации нейропсихологических тестов

№ комбинации	Параметр 1	Параметр 2	Параметр 3	Параметр 4	Параметр 5
1	Асс 2	КШОПС общ.	ШОЛД общ.		
2	Асс 1	Асс 2	КШОПС 5	ШОЛД 1	ШОЛД 2
3	Асс 1	КШОПС 5	ШОЛД 1	ШОЛД 2	
4	Асс 1	Асс 2	ШОЛД 1	ШОЛД 2	
5	Асс 2	КШОПС 5	КШОПС 7	ШОЛД 3	

Примечание

Асс1-тест на называние литеральных ассоциаций

Асс2 – тест на называние категориальных ассоциаций

КШОПС общ. – общий балл по краткой шкале оценки психического статуса

КШОПС 5 – субтест «память» по краткой шкале оценки психического статуса

КШОПС 7 – субтест «повторение фразы» по краткой шкале оценки психического статуса

ШОЛД общ. – общий балл по шкале оценки лобной дисфункции

ШОЛД 1 – субтест «ассоциативное мышление» по шкале оценки лобной дисфункции

ШОЛД 2 – субтест «беглость речи» по шкале оценки лобной дисфункции

ШОЛД 3 – субтест «динамический праксис» по шкале оценки лобной дисфункции

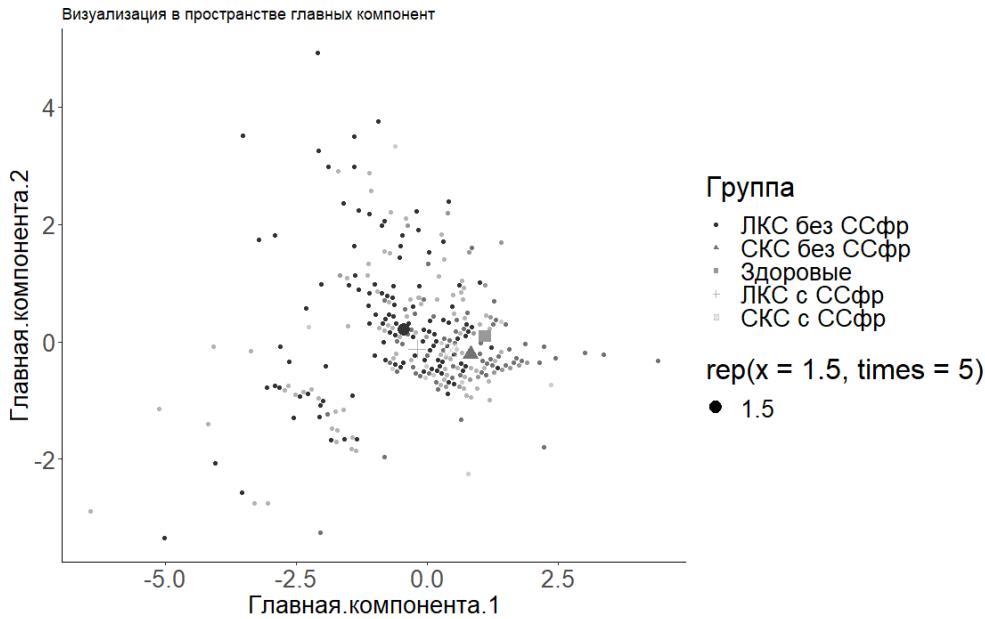


Рис. 2. Пример визуализации в пространстве главных компонент (Для комбинации №2)

**Апостериорный анализ**

Несмотря на отбор оптимальной комбинации тестов, группы пациентов остаются нечеткими, поэтому проведена коррекция групп: дальнейший алгоритм исключал пациентов, чьи показатели были не достаточно однозначны для включения в какую-либо группу. Было проведено формирование т.н. «ядер», которые наиболее отчетливо характеризуют основные закономерности группы. Это сокращение было реализовано по следующему алгоритму:

1. Для каждой комбинации отобранных признаков были выделены центры исходных групп.

2. Эти центры использовались для последующей K-means кластеризации, в качестве начальных центров. Причем кластеризация проводилась для 2-х вариантов исходной выборки — в одной содержались только стандартизированные значения нейропсихологических тестов отобранных в комбинации, в другой к ним добавлен номер исходной группы. Будем такие кластеры обозначать  $K^{(R)}$  и  $K^{(F)}$  соответственно.

3. Было последовательно проведено несколько итераций кластеризации. На каждом этапе среди множества всех пациентов, присутствующих в данной итерации  $P$ , определялись 4 множества:

a. Множество  $S$  — такие пациенты, у которых совпадает принадлежность априорной группы ( $G_z$ ) и ассоциированные с ней кластеры обоих типов ( $K_l^{(R)}$  и  $K_m^{(F)}$ ).

$$S = \{p \in P | (p \in G_z \wedge p \in K_l^{(R)} \wedge p \in K_m^{(F)}), (K_l^{(R)} \sim G_z, K_m^{(F)} \sim G_z), (G_z \subseteq P, K_l^{(R)} \subseteq P, K_m^{(F)} \subseteq P)\}$$

b. Множество  $W_1$  — такие пациенты, у которых совпадает принадлежность априорной группы и ассоциированного с ней кластера  $K^{(F)}$

$$W_1 = \{p \in P | (p \in G_z \wedge p \in K_m^{(F)}) (K_l^{(R)} \sim G_z, K_m^{(F)} \sim G_z), (G_z \subseteq P, K_m^{(F)} \subseteq P)\}$$

c. Множество  $W_2$  — такие пациенты, у которых совпадает принадлежности априорной группы и ассоциированной с ним кластера  $K^{(R)}$

$$W_2 = \{p \in P | (p \in G_z \wedge p \in K_l^{(R)}) (K_l^{(R)} \sim G_z, K_m^{(F)} \sim G_z), (G_z \subseteq P, K_l^{(R)} \subseteq P)\}$$

d. Множество  $W_3$  — такие пациенты, у которых совпадает принадлежности априорных групп и хотя бы одного из ассоциированного с ним кластера  $K^{(R)}$

$$W_3 = \left\{ p \in P \mid \left( p \in G_z \wedge \left( p \in K_l^{(R)} \vee p \in K_m^{(F)} \right) \right), \left( K_l^{(R)} \sim G_z, K_m^{(F)} \sim G_z \right), \left( G_z \subseteq P, K_l^{(R)} \subseteq P, K_m^{(F)} \subseteq P \right) \right\}$$

4. Также, для контроля и оценки качества отбора, рассчитывались 3 семейства параметров: Н-параметры, FF-параметры, FR-параметры. Н-параметры отражают долю каждого множества  $\{S, W1, W2, W3\}$ , долю «угаданных» кластеров при добавлении номера исходной группы и без добавления исходной группы. FF-параметры, FR-параметры — это параметры полноты (*recall*) [15, 16] для кластеров  $K^{(F)}$  и  $K^{(R)}$  соответственно. Значком # будем обозначать мощность множества, чтобы избежать путаницы со знаком модуля.

$$SH = \frac{\#S}{\#P}$$

$$WH1 = \frac{\#W1}{\#P}$$

$$WH2 = \frac{\#W2}{\#P}$$

$$WH3 = \frac{\#W3}{\#P}$$

$$FF_i = \frac{\#\{f \in P \mid (f \in G_i \wedge f \in K_i^{(F)}) \wedge (K_i^{(F)} \sim G_i), (G_i \subseteq P, K_i^{(F)} \subseteq P)\}}{\#\{f \in P \mid (f \in G_i \wedge f \in K_i^{(F)}) \wedge (K_i^{(F)} \sim G_i), (G_i \subseteq P, K_i^{(F)} \subseteq P)\} +$$

$$\#\{f \in P \mid (f \notin G_i \wedge f \in K_i^{(F)}) \wedge (K_i^{(F)} \sim G_i), (G_i \subseteq P, K_i^{(F)} \subseteq P)\}}$$

$$FR_i = \frac{\#\{f \in P \mid (f \in G_i \wedge f \in K_i^{(R)}) \wedge (K_i^{(R)} \sim G_i), (G_i \subseteq P, K_i^{(R)} \subseteq P)\}}{\#\{f \in P \mid (f \in G_i \wedge f \in K_i^{(R)}) \wedge (K_i^{(R)} \sim G_i), (G_i \subseteq P, K_i^{(R)} \subseteq P)\} +$$

$$\#\{f \in P \mid (f \notin G_i \wedge f \in K_i^{(R)}) \wedge (K_i^{(R)} \sim G_i), (G_i \subseteq P, K_i^{(R)} \subseteq P)\}}$$

5. На каждой итерации удаляются пациенты, не попавшие ни в одно из множеств
6. Критерием остановки алгоритма было «истощение» какой-либо из априорных групп

$$\#\{g \in P \mid g \in G_z, G_z \subseteq P\} = 0$$

При анализе этих показателей можно сделать следующие выводы:

- Во всех комбинациях алгоритм завершался на 3-й итерации. «Исчерпанными» (множествами с нулевой мощностью) оказываются группы контроля и пациенты с СКС. Т.е. Группы с ЛКС имеют наибольшие отличия от остальных групп.
- В большинстве комбинаций, лучше всего кластеризуются (как с добавлением номера группы, так и нет) группы пациентов с ЛКС (с ССФР и без ССФР). Худшие FF параметры меняются от комбинации к комбинации, а худшими FR-параметрами почти во всех комбинациях оказываются пациенты с СКС.
- Худшими значениями Н-параметров обладает комбинация 2. Однако ее параметры с точки зрения кластеризации групп похожи на остальные комбинации.

• 5-я комбинация обладает отличными от прочих комбинаций результатами при реализации алгоритма. Она позволяет наиболее полно кластеризовать пациентов при учете номера априорной группы, однако, без знаний о реальной принадлежности пациента единственной хорошо кластеризуемой группой оказывается группа контроля (что не наблюдается в остальных комбинациях).

#### Канонический корреляционный анализ

На заключительном этапе проведен канонический анализ. Канонический анализ (*Canonical Correlation Analysis, CCA*) — это статистический метод, поиска зависимостей, в ситуации, когда присутствует группа входных переменных и группа выходных переменных (в отличие от регрессионного анализа). Канонические коэффициенты определяются при решении задачи макси-

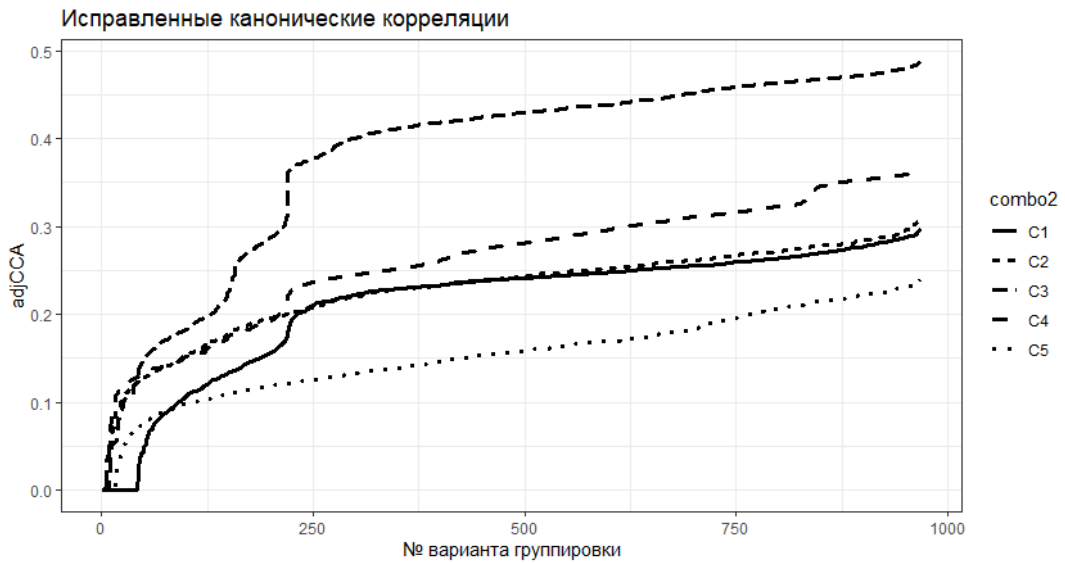


Рис. 3. Ранжированные канонические корреляции

мизации коэффициента корреляции с учетом всех возможных линейных комбинаций между входами и всех возможных линейных комбинаций между выходами [17]. Была использована компьютерная реализация алгоритма [18].

На Рис. 3 представлены расчеты канонических корреляций, для всех рассматриваемых комбинаций признаков. Исследовались следующие факторы риска: курение, прием алкоголя, язвенная болезнь желудка, инфаркт миокарда, сердечные нарушения (нарушения ритма и проводимости сердца), сахарный диабет, артериальная гипертензия, острое нарушение мозгового кровообращения, наличие онкологического заболевания. С учетом того, что количество рассматриваемых факторов риска в каждом варианте были различны, также как и различно было количество

«ядерных» пациентов, отобранных для каждой нейропсихологической комбинации, использовалась формула исправленной (adjusted) канонической корреляции [19].

На Рис. 3 видно, что в целом 3-я комбинация нейропсихологических тестов обладает большей канонической корреляцией. В группу соматических признаков, которые соответствуют наибольшей канонической корреляции, вошли: язвенная болезнь, перенесенный инфаркт миокарда, сахарный диабет, артериальная гипертензия, наличие онкологического заболевания.

В Табл. 2 приведены значения канонических коэффициентов для итоговой канонической переменной.

Как видно из таблицы, наибольшим влиянием на итоговую каноническую переменную со

Табл. 2. Результаты канонического анализа

Показатель X	Показатель Y	
Асс 1	0.997373	Язвенная болезнь -0.58317
КШОПС 5	0.088015	Инфаркт миокарда 1.949728
ШОЛД 1	0.36342	Сахарный диабет -0.46685
ШОЛД 2	-0.38722	Артериальная гипертензия 0.829842
		Онкология -0.46138

Примечание

Асс 1-тест на называние литеральных ассоциаций

КШОПС 5 – субтест «память» по Краткой шкале оценки психического статуса

ШОЛД 1 – субтест «ассоциативное мышление» по Шкале оценки лобной дисфункции

ШОЛД 2 – субтест «беглость речи» по Шкале оценки лобной дисфункции

стороны соматических факторов оказывают инфаркт миокарда в анамнезе и артериальная гипертензия.

## Заключение

Исследование показало, что требуются новые подходы к работе с медицинскими данными, связанные с изучением доумеренного когнитивного снижения, что вызвано большим разнообразием индивидуальных показателей на этой стадии когнитивного снижения. Данная проблема характерна не только для медицинской области, но и для широкого круга зашумленных и слабо согласованных данных, поэтому можно говорить о полезности полученных результатов для статистического анализа в поисковых научных исследованиях.

Предложен алгоритм для проведения статистической обработки результатов подобных исследований, связанный с интеллектуальными фильтрациями, как пространства признаков, так и самих наблюдений.

Подтверждены результаты предыдущих исследований о наиболее тесной связи когнитивного снижения с сердечно-сосудистыми факторами риска, такими как артериальная гипертензия или инфаркт миокарда в анамнезе.

Дальнейшие исследования по данной тематике могут быть связаны как с уточнением факторов риска (в настоящей работе не было изучено влияние эмоционально-аффективных факторов) и дискретностью оценки их тяжести, так и с развитием предиктивных моделей, на основе найденных зависимостей.

## Литература

1. Захаров В.В., Яхно Н.Н. Синдром умеренных когнитивных нарушений в пожилом возрасте – диагностика и лечение. *РМЖ*. 2004; 12(10): 573–6.
2. Amieva H., Le Goff M., Millet X., Orgogozo J.M., Pérès K., Barberger-Gateau P. et al. Prodromal Alzheimer's disease: Successive emergence of the clinical symptoms. *Ann. Neurol.* 2008; 64(5): 492–8
3. Mitchell A.J., Beaumont H., Ferguson D., Yadegarfar M., Stubbs B. Risk of dementia and mild cognitive impairment in older people with subjective memory complaints: meta-analysis. *Acta Psychiatr. Scand.* 2014; 130(6): 439–51.
4. Локшина А.Б., Захаров В.В. Лёгкие и умеренные когнитивные расстройства при дисциркуляторной энцефалопатии. *Неврологический журнал*. 2006; 11 (Прил. 1): 57–63.
5. Dubois B., Hampel H.J., Feldman H.H., Scheltens P., Aisen P., Andrien S. et al. Preclinical Alzheimer's disease: Definition, natural history and diagnostic criteria. *Alzheimers Dement.* 2016; 12(3): 292–323.
6. Seo E.H., Kim H., Lee K.H., Choo I.H. Altered executive function in pre-mild cognitive impairment. *J. Alzheimers Dis.* 2016; 54(3): 933–40
7. Jessen F., Amariglio R.E., van Boxtel M., Breteler M., Ceccaldi M., Chételat G. et al. Subjective Cognitive Decline Initiative (SCD-I) Working Group. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimers Dement.* 2014; 10(6): 844–52
8. Яхно Н.Н., Захаров В.В., Коберская Н.Н., Мхитарян Э.А., Гришина Д.А., Локшина А.Б., Савушкина И.Ю., Посохов С.И. "Предумеренные" (субъективные и лёгкие) когнитивные расстройства // *Неврологический журнал*. 2017. №4. URL: <https://cyberleninka.ru/article/n/predumerennye-subektivnye-i-lyogkie-kognitivnye-rasstroystva> (дата обращения: 03.03.2021).
9. Остроумова Т.М., Парфенов В.А., Остроумова О.Д., Перепелова Е. М., Перепелов В. А., Борисова Е.В. Возможности метода бесконтрастной магнитно-резонансной перфузии для выявления раннего поражения головного мозга при эссенциальной артериальной гипертензии // *Неврология, нейропсихиатрия, психосоматика*. - 2018. – Т. 10. - № 1. – С. 17–23.
10. Парфенов В.А., Остроумова Т. М., Перепелова Е. М., Перепелов В. А., Кочетков А. И., Остроумова О.Д. Перфузия головного мозга, когнитивные функции и сосудистый возраст у пациентов среднего возраста с эссенциальной артериальной гипертензией // *Кардиология*. – 2018. – Т. 58. - № 5. – С. 23–31.
11. Коберская Н.Н., Яхно Н.Н., Гридин В.Н., Смирнов Д.С. Влияние сердечно-сосудистых факторов риска на доумеренное когнитивное снижение в среднем и пожилом возрасте. *Неврология, нейропсихиатрия, психосоматика*. 2021;13(1):13-17. <https://doi.org/10.14412/2074-2711-2021-1-13-17>
12. Яхно Н.Н., Гридин В.Н., Смирнов Д.С., Панишев В.С., Парфенов В.А., Остроумова Т.М., Коберская Н.Н. Статистическая обработка и методика сокращения размерности пространства данных пациентов при анализе когнитивных нарушений // *Информационные технологии*. 2020. Т. 26. № 9. С. 515-522
13. Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). clValid: An R Package for Cluster Validation. *Journal of Statistical Software*, 25(4), 1 - 22. doi:<http://dx.doi.org/10.18637/jss.v025.i04>
14. Kendall, M.G. (1948) Rank correlation methods. London: Griffin
15. Labintsev, E. Метрики в задачах машинного обучения. 12 мая 2017. [Электронный ресурс], режим доступа: <https://habr.com/ru/company/ods/blog/328372/>
16. Marxer, R., & Purwins H. An F-Measure for Evaluation of Unsupervised Clustering with Non-Determined Number of Clusters. 2008. [Электронный ресурс] Режим доступа: <http://mtg.upf.edu/files/publications/unsuperf.pdf>



17. Hotelling H. (1936). Relations between two sets of variables. *Biometrika*, 28, 321–327. doi: 10.1093/biomet/28.3-4.321
18. González I., Déjean S. CCA: Canonical Correlation Analysis. [Электронный ресурс] Режим доступа: <https://cran.r-project.org/web/packages/CCA/index.html>
19. Yehia, Ahmed & Saleh, Mohamed & Ahmed, Abdul-Hadi. (2016). An Adjusted Correlation Coefficient for Canonical Correlation Analysis // *Journal of Egyptian Statistical Society*.

**Яхно Николай Николаевич.** ФГБУН Центра информационных технологий в проектировании РАН, Московская обл., г. Одинцово. Доктор математических наук, профессор, академик РАН. Количество печатных работ: более 649 (в том числе 20 монографий). Область научных интересов: неврология, нейропсихология. E-mail: [info@ditc.ras.ru](mailto:info@ditc.ras.ru)

**Гридин Владимир Николаевич.** ФГБУН Центра информационных технологий в проектировании РАН, Московская обл., г. Одинцово. Доктор технических наук, профессор. Количество печатных работ: более 410 (в том числе 9 монографий). Область научных интересов: информационные технологии, САД-системы, спектральные численно-аналитические методы. E-mail: [info@ditc.ras.ru](mailto:info@ditc.ras.ru)

**Коберская Надежда Николаевна.** ФГБУН Центра информационных технологий в проектировании РАН, Московская обл., г. Одинцово. Кандидат математических наук. Количество печатных работ: более 60. Область научных интересов: неврология, нейропсихиатрия, психосоматика. E-mail: [info@ditc.ras.ru](mailto:info@ditc.ras.ru)

**Смирнов Дмитрий Сергеевич.** ФГБУН Центра информационных технологий в проектировании РАН, Московская обл., г. Одинцово. Кандидат экономических наук. Количество печатных работ: 43 (в том числе 1 монография). Область научных интересов: математическая статистика, теория вероятности, имитационное моделирование. E-mail: [info@ditc.ras.ru](mailto:info@ditc.ras.ru)

**Солодовников Владимир Игоревич.** ФГБУН Центра информационных технологий в проектировании РАН, Московская обл., г. Одинцово. Кандидат технических наук. Количество печатных работ: более 85. Область научных интересов: информационные технологии, САД-системы, спектральные численно-аналитические методы. E-mail: [info@ditc.ras.ru](mailto:info@ditc.ras.ru)

## Algorithm of Data Filtration and Dependencies Searching in Analysis of Patients with Sub-Mild Cognitive Decline

N. N. Yakhno<sup>1,II</sup>, V. N. Gridin<sup>I</sup>, N. N. Koberskaya<sup>III</sup>, D. S. Smirnov<sup>I</sup>, V. I. Solodovnikov<sup>I</sup>

<sup>I</sup> Design Information Technologies Center Russian Academy of Sciences (DITC RAS)

<sup>II</sup> I. M. Sechenov First Moscow State Medical University (Sechenov University), V.N. Sklifosovsky Institute of Clinical Medicine, Department of Nervous Diseases and Neurosurgery

**Abstract.** Paper is aimed to develop algorithm for searching statically significant dependencies on medical data of patients with sub-mild cognitive decline. Algorithms and methods of statistical processing and transformation of medical data of patients are considered. A method for identifying groups of indicators that have the greatest consistency and separation capacity for a priori groups of patients and a method for selecting key observations are proposed. Combinatorial canonical analysis is applied to find the most important risk factors for cognitive decline. Confirmed the results of previous studies on the closest relationship between cognitive decline and cardiovascular risk factors such as arterial hypertension or myocardial infarction in anamnesis

**Keywords:** dimension reduction, multivariate statistical methods, canonical analysis, cognitive decline.

DOI 10.14357/20718632210204

## References

1. Zakharov V.V., Yakhno N.N. Sindrom umerennykh kognitivnykh narusheniy v pozhilom vozraste – diagnostika i lechenie [Mild cognitive impairment syndrome in old age - diagnosis and treatment]. *Rossiyskiy meditsinskiy zhurnal*. 2004; 12(10): 573–576.
2. Amieva H., Le Goff M., Millet X., Orgogozo J.M., Pérès K., Barberger-Gateau P. et al. Prodromal Alzheimer’s disease: Successive emergence of the clinical symptoms. *Ann. Neurol*. 2008; 64(5): 492–8
3. Mitchell A.J., Beaumont H., Ferguson D., Yadegarfar M., Stubbs B. Risk of dementia and mild cognitive impairment

- in older people with subjective memory complaints: meta-analysis. *Acta Psychiatr. Scand.* 2014; 130(6): 439–51.
4. Lokshina A.B., Zakharov V.V. Legkie i umerennye kognitivnye rasstroystva pri distsirkulyatornoy entsefalopatii [Mild to moderate cognitive impairment in discirculatory encephalopathy]. *Nevrologicheskiy zhurnal.* 2006; 11(Pril. 1): 57–63.
  5. Dubois B., Hampel H.J., Feldman H.H., Scheltens P., Aisen P., Andrien S. et al. Preclinical Alzheimer's disease: Definition, natural history and diagnostic criteria. *Alzheimers Dement.* 2016; 12(3): 292–323.
  6. Seo E.H., Kim H., Lee K.H., Choo I.H. Altered executive function in pre-mild cognitive impairment. *J. Alzheimers Dis.* 2016; 54(3): 933–40
  7. Jessen F., Amariglio R.E., van Boxtel M., Breteler M., Ceccaldi M., Chételat G. et al. Subjective Cognitive Decline Initiative (SCD-I) Working Group. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimers Dement.* 2014; 10(6): 844–52
  8. Yakhno N.N., Zakharov V.V., Koberskaya N.N., Mkhitarian E.A., Grishina D.A., Lokshina A.B., Savushkina I.Yu., Posokhov S.I. "Predumerennye" (sub)ektivnye i legkie kognitivnye rasstroystva ["Sub mild" (subjective and light) cognitive disorders] // *Nevrologicheskiy zhurnal.* 2017. №4. Available at: <https://cyberleninka.ru/article/n/predumerennye-subektivnye-i-lyogkie-kognitivnye-rasstroystva> ((accessed: March 03 2021).
  9. Ostroumova T.M., Parfenov V.A., Ostroumova O.D., Perepelova E. M., Perepelov V. A., Borisova E.V. Vozmozhnosti metoda beskontrastnoy magnitno-rezonansnoy perfuzii dlya vyyavleniya rannego porazheniya golovnoy mozga pri essential'noy arterial'noy gipertenzii [Possibilities of the non-contrast magnetic resonance perfusion method for detecting early brain damage in essential arterial hypertension] // *Nevrologiya, neyropsikhiatriya, psikhosomatika.* 2018; 10 (1): 17–23.
  10. Parfenov V.A., Ostroumova T. M., Perepelova E. M., Perepelov V. A., Kochetkov A. I., Ostroumova O.D. Perfuziya golovnoy mozga, kognitivnye funktsii i sosudisty vozrast u patsientov srednego vozrasta s essential'noy arterial'noy gipertenziey [Brain perfusion, cognitive function and vascular age in middle-aged patients with essential arterial hypertension] *Kardiologiya.* 2018; 58(5): 23–31.
  11. Koberskaya N.N., Yakhno N.N., Gridin V.N., Smirnov D.S. [The influence of cardiovascular risk factors on sub-moderate cognitive decline in middle age and old age.] *Nevrologiya, neyropsikhiatriya, psikhosomatika.* 2021;13(1):13-17. <https://doi.org/10.14412/2074-2711-2021-1-13-17>
  12. Yakhno N.N., Gridin V.N., Smirnov D.S., Panishchev V.S., Parfenov V.A., Ostroumova T.M., Koberskaya N.N. Ctatisticheskaya obrabotka i metodika sokrashcheniya razmernosti prostranstva dannykh patsientov pri analize kognitivnykh narusheniy [Statistical processing and methods of dimension reduction of the medical data of patients with subjective and mild cognitive decline]. *Informatsionnye tekhnologii.* 2020; 26(9): 515-522
  13. Brock, G., Pihur, V., Datta, S., & Datta, S. clValid: An R Package for Cluster Validation. *Journal of Statistical Software,* 2008;25(4): 1-22. doi:<http://dx.doi.org/10.18637/jss.v025.i04>
  14. Kendall, M.G. (1948) Rank correlation methods. London: Griffin.
  15. Labintcev, E. *Metriki v zadachakh mashinnogo obucheniya [Metrics in Machine Learning Problems].* May 12, 2017, Available at: <https://habr.com/ru/company/ods/blog/328372/>, (accessed March 5 2021)
  16. Marxer, R., & Purwins H. An F-Measure for Evaluation of Unsupervised Clustering with Non-Determined Number of Clusters. 2008. [Электронный ресурс] Режим доступа: <http://mtg.upf.edu/files/publications/unsuperf.pdf>
  17. Hotelling H. (1936). Relations between two sets of variables. *Biometrika,* 28, 321–327. doi: 10.1093/biomet/28.3-4.321.
  18. González I., Déjean S. CCA: Canonical Correlation Analysis. Available at: <https://cran.r-project.org/web/packages/CCA/index.html> (accessed March 5 2021)
  19. Yehia, Ahmed & Saleh, Mohamed & Ahmed, Abdul-Hadi. (2016). An Adjusted Correlation Coefficient for Canonical Correlation Analysis // *Journal of Egyptian Statistical Society.*

**Nikolay N. Y.** MD, professor, academician of the Russian Academy of Sciences, of the Design information technologies Center Russian Academy of Sciences, Moscow Region, Odintsovo, st. Marshal Biryuzova, 7a 143003, I.M. Sechenov First Moscow State Medical University, Trubetskaya St., 8, p. 2, Moscow. Number of publications: - more than 649 (including 20 monographs). Research interests: neurology, neuropsychology, e-mail: [info@ditc.ras.ru](mailto:info@ditc.ras.ru)

**Gridin V. N.** Doctor of Technical Sciences, Professor, Scientific director of the Design information technologies Center Russian Academy of Sciences, Research interests: CAD systems, information technology. The number of scientific publications is 410 (including 9 monographs). Moscow Region, Odintsovo, st. Marshal Biryuzova, 143003, e-mail: [info@ditc.ras.ru](mailto:info@ditc.ras.ru)

**Koberskaya N. N.** Candidate of Medical Sciences, of the Design information technologies Center Russian Academy of Sciences, Research interests: neuropsychiatry, psychosomatics. The number of scientific publications is 60. Moscow Region, Odintsovo, st. Marshal Biryuzova, 143003. I.M. Sechenov First Moscow State Medical University, Trubetskaya St., 8, p. 2, Moscow e-mail: [info@ditc.ras.ru](mailto:info@ditc.ras.ru)

**Smirnov D. S.** Candidate of Economic Sciences. Senior Researcher of the Design information technologies Center Russian Academy of Sciences, Research interests: mathematical statistics, probability theory, simulation. The number of scientific publications is 43. Moscow region, 7a, Marshal Biryuzov St, Moscow Region, Odintsovo, 143003, e-mail: [info@ditc.ras.ru](mailto:info@ditc.ras.ru)

**Solodovnikov V. I.** Candidate of Technical Sciences, director of the Design information technologies Center Russian Academy of Sciences, Research interests: CAD systems, information technology. The number of scientific publications is 85. Moscow Region, Odintsovo, st. Marshal Biryuzova, 143003, e-mail: [info@ditc.ras.ru](mailto:info@ditc.ras.ru)