

УДК 519. 23

ПСЕВДОРАНДОМИЗАЦИЯ (PROPENSITY SCORE MATCHING) КАК СОВРЕМЕННЫЙ СТАТИСТИЧЕСКИЙ МЕТОД УСТРАНЕНИЯ СИСТЕМАТИЧЕСКИХ РАЗЛИЧИЙ СРАВНИВАЕМЫХ ГРУПП ПРИ АНАЛИЗЕ КОЛИЧЕСТВЕННЫХ ИСХОДОВ В ОБСЕРВАЦИОННЫХ ИССЛЕДОВАНИЯХ

© 2016 г. ¹⁻⁴А. М. Гржибовский, ⁵С. В. Иванов, ²М. А. Горбатова, ⁶А. А. Дюсупов¹Национальный институт общественного здравоохранения, г. Осло, Норвегия; ²Северный государственный медицинский университет, г. Архангельск; ³Северо-Восточный федеральный университет, г. Якутск;⁴Международный казахско-турецкий университет, г. Туркестан, Казахстан;⁵Северо-Западный государственный медицинский университет им. И. И. Мечникова, г. Санкт-Петербург;⁶Государственный медицинский университет г. Семей, Казахстан

В статье представлен метод псевдорандомизации (propensity score matching – PSM) – эффективный способ устранения конфаунд-эффекта различных факторов, искажающих результаты при сравнении наблюдаемых групп в обсервационных исследованиях. Метод PSM используется на этапе статистической обработки данных, сравним по эффективности с регрессионным анализом, но не требует при этом большого размера выборочной совокупности. В статье представлены основы данного метода и алгоритмы его применения с использованием статистического программного обеспечения STATA 13 для оценки различий между средними значениями количественной переменной исхода в изучаемых группах в обсервационном исследовании.

Ключевые слова: статистический анализ, обсервационные исследования, псевдорандомизация, propensity score, подбор пар, взвешивание, конфаундер, смещение результатов, ошибка отбора, количественная переменная

PROPENSITY SCORE MATCHING AS A MODERN STATISTICAL METHOD FOR BIAS CONTROL IN OBSERVATIONAL STUDIES WITH CONTINUOUS OUTCOME VARIABLE

¹⁻⁴A. M. Grjibovski, ⁵S. V. Ivanov, ²M. A. Gorbatoва, ⁶A. A. Dyussupov¹Norwegian Institute of Public Health, Oslo, Norway; ²Northern State Medical University, Arkhangelsk, Russia;³North-Eastern Federal University, Yakutsk, Russia; ⁴International Kazakh-Turkish University, Turkestan, Kazakhstan;⁵North-Western State Medical University, St. Petersburg, Russia; ⁶Semey State Medical University, Semey, Kazakhstan

The authors presents a propensity score matching (PSM) technique - an effective method to control the effect of confounding factors in observational studies. PSM has been shown to be as efficient as linear regression analysis, but can be performed using smaller samples. This article presents the basic principles of PSM and its practical application using STATA 13 software for the studies with continuous dependent variable.

Keywords: statistical analysis, observational studies, pseudo-randomization, propensity score, matching, weighting, confounder, bias, selection bias, continuous variable

Библиографическая ссылка:

Гржибовский А. М., Иванов С. В., Горбатова М. А., Дюсупов А. А. Псевдорандомизация (propensity score matching) как современный статистический метод устранения систематических различий сравниваемых групп при анализе количественных исходов в обсервационных исследованиях // Экология человека. 2016. № 7. С. 51–60.

Grjibovski A. M., Ivanov S. V., Gorbatoва M. A., Dyussupov A. A. Propensity Score Matching as a Modern Statistical Method for Bias Control in Observational Studies with Continuous Outcome Variable. *Ekologiya cheloveka* [Human Ecology]. 2016, 7, pp. 51-60.

Одной из целей исследований в медицине является достоверная оценка наличия, направления и силы связи между факторами воздействия и наблюдаемыми исходами. Примерами изучаемых факторов могут служить назначение определенных схем терапии, факторы окружающей среды, организационные мероприятия в области общественного здравоохранения и др. Примерами исходов могут служить количественные характеристики объектов исследования (уровень общего холестерина, длительность госпитализации, расходы на лечение и т. п.), а также категориальные показатели (наличие или отсутствие заболевания, летального исхода, осложнения и т. п.).

Следует учесть, что изучаемый исход подвержен влиянию как изучаемого в ходе исследования фактора, так и достаточного количества прочих факторов, которые могут быть связаны как с изучаемым воздействующим фактором, так и с исходом. Такие факторы называются конфаундерами [14]. Следствием действия конфаундеров является искажение результатов исследования, то есть возникновение различий между рассчитанным и фактическим значением меры и направления воздействия изучаемого фактора.

Данная проблема может решаться с помощью проведения рандомизированных контролируемых исследований (РКИ), которые при корректной рандоми-

зации обеспечивают сбалансированное распределение конфаундеров между основной группой и группой контроля и поэтому являются «золотым стандартом» оценки воздействия изучаемого фактора на исход. При этом влияние фактора на исход оценивается путем прямого сравнения исходов в основной и контрольной группах [5]. Но так как РКИ имеют существенные ограничения в использовании, к числу которых относится невозможность организовать процедуру рандомизации, а также незнание исследований, направленных на изучение лечебного воздействия, наиболее часто встречающимся инструментом оценки связи между фактором и исходом в медицине и особенно в общественном здравоохранении остаются обсервационные исследования [1, 2, 5].

Проведение обсервационных исследований требует корректной оценки мер эффекта действия фактора, учитывающей влияние конфаундеров. Например, в случае изучения эффективности какого-либо дополнительного метода лечения, используемого в составе комплексной терапии, прямое сравнение исхода в группе пациентов, получавших данный метод лечения в составе комплексной терапии, с исходом в группе пациентов, его не получавших, будет некорректным, так как группы сравнения, вероятнее всего, будут существенно различаться по базисным схемам терапии. В обсервационных исследованиях всегда есть риск включить в анализ группы сравнения, имеющие существенные систематические отличия друг от друга, что при прямом сравнении исходов в наблюдаемых группах не позволит экстраполировать результаты исследования на генеральную совокупность.

Таким образом, при анализе результатов обсервационных исследований необходимо использовать такой способ включения наблюдений в анализ, который мог бы обеспечить максимальную сопоставимость основной и референтной групп по имеющимся конфаундерам. Одним из таких способов является метод «Propensity score matching» (PSM), который был разработан P. R. Rosenbaum и D. B. Rubin еще в 1983 году [16]. Широкое внедрение метода PSM в исследовательский процесс в медицине, эконометрике, психологии и социологических исследованиях активно началось после 2000 г., и в настоящее время метод PSM использует все большее число исследователей по всему миру.

В обсервационных исследованиях, в отличие от экспериментальных, исследователь не может равномерно распределить распространенность конфаундеров между группами. Такое распределение происходит самопроизвольно, причем как случайным образом, так и вследствие набора неких характеристик объекта, связанных как с изучаемым фактором, так и с изучаемым исходом. Метод PSM позволяет рассчитать индекс соответствия (propensity score — PS) — вероятность попадания каждого объекта исследования в основную или контрольную группу наблюдения на основании набора его характеристик, то есть значение PS показывает вероятность для каждого объекта исследования попасть под действие изучаемого фактора

(значения PS, как и подобает вероятности, находятся в пределах от 0 до 1). Подробные математические выкладки для метода PSM описаны в специализированной литературе по статистике [9, 11, 13, 16].

Рассчитав значения условной вероятности попадания объекта исследования в основную группу, можно подобрать группу сравнения таким образом, чтобы данная вероятность была сбалансирована между группами сравнения. Таким образом будет скомпенсировано неравномерное распределение конфаундеров между основной группой и группой сравнения, то есть будет проведена процедура, сходная с процедурой рандомизации в РКИ, но только проводимая на этапе обработки, а не набора данных.

Особенность метода PSM заключается в том, что он позволяет свести широкий набор характеристик каждого наблюдения к единому вариационному ряду значений PS [8, 11, 12].

Алгоритм PSM включает в себя следующие основные этапы:

1. Отбор переменных-ковариат, кодирующих конфаундеры, для последующего включения в анализ.
2. Расчет значений PS для каждого наблюдения.
3. Проверка баланса распределения значений PS и переменных-ковариат между группами сравнения.
4. Расчет мер эффекта воздействия фактора на основании одного из имеющихся методов подбора пар или «взвешивания».
5. Оценка эффективности устранения дисбаланса ковариат после применения выбранного метода подбора пар или «взвешивания».
6. В случае сохранения дисбаланса повторное проведение анализа с использованием других методов подбора пар или «взвешивания».

Переменные, кодирующие фактор и исход, определяются целью исследования [1, 2, 5]. Переменная, кодирующая изучаемый фактор, должна быть дихотомической (есть фактор / нет фактора), переменная исхода должна быть количественной или качественной (в настоящей статье рассматриваются только случаи количественной переменной исхода). Следует отметить, что в зарубежных источниках переменная, кодирующая фактор, часто называется словом «treatment» — «лечение», независимо от того, к какой области науки относится исследование (медицина, общественное здравоохранение, эконометрика, психология или социология). Разумеется, спектр факторов, названных этим словом, гораздо шире, чем просто лечебное воздействие.

Согласно самой концепции PSM, отбор конфаундеров для включения в анализ должен быть основан на теоретических представлениях об их потенциальной связи как с изучаемым фактором, так и с исходом [7]. При этом в анализ не должны включаться переменные, которые являются звеном патогенетической цепи между изучаемым фактором и исходом. Следует отметить, что многие вопросы отбора конфаундеров в PSM не имеют однозначных ответов и остаются объектом обсуждения в литературе [7, 12].

Примерами конфаундеров могут служить социально-демографические характеристики участников исследования (возраст, пол, уровень доходов и т. п.), характеристика коморбидных состояний (сопутствующая патология, длительность анамнеза, частота госпитализаций и т. п.), локальные особенности (регион проживания, климатические особенности и т. п.), а также другие характеристики объектов исследования, удовлетворяющие вышеуказанным условиям.

Расчет значений PS производится с помощью логистической регрессии, в которой в качестве переменной отклика указывается факт принадлежности к основной группе (находящейся под действием «лечения»), а кодирующие конфаундеры ковариаты являются независимыми переменными. Заметим, что рассчитанные значения условной вероятности попадания наблюдения в основную группу могут распределиться таким образом, что определенная часть наблюдений будет иметь набор характеристик, создающий очень высокую вероятность подвергнуться действию фактора, а для другой части наблюдений, напротив, попадание под действие фактора будет очень маловероятным. Именно поэтому для оценки эффекта изучаемого фактора наиболее подходящими будут те наблюдения, которые попадают в зону перекреста областей значений PS («common support») [7, 15, 17].

Баланс распределения значений PS между основной и контрольной группами проводится путем разделения выборки наблюдений на страты таким образом, чтобы средние значения PS в пределах каждой страты в основной группе и группе сравнения статистически значимо не различались. После определения страт проводится проверка баланса всех ковариат в пределах каждого блока.

Далее используется один из следующих основных методов анализа [7, 9, 12]:

- Стратификация объектов исследования на основании значений PS.

- Подбор к каждому наблюдению основной группы одной или нескольких пар («ближайших соседей») — наблюдений с наиболее близким значением PS.

- Взвешивание Кернела, при котором каждому наблюдению основной группы присваивается весовой коэффициент, равный единице, а все наблюдения группы контроля ранжируются таким образом, чтобы более близкие к объектам основной группы наблюдения контрольной группы получали большие «веса» по сравнению с теми, которые являются более удаленными.

Вышеперечисленные методы необходимы для балансировки конфаундеров между основной и контрольной группами.

Основными мерами эффекта воздействия изучаемого фактора на исход являются следующие [17]:

1. Средний эффект воздействия фактора на объекты основной группы («average treatment effect for the treated» — АТТ).

2. Средний эффект воздействия фактора («average treatment effect» — АТЕ).

Следует отметить, что в случае анализа данных РКИ

эти два показателя будут практически равны между собой, так как благодаря процедуре рандомизации основная группа не должна иметь систематических отличий от группы контроля и генеральной совокупности.

Выбор между АТТ или АТЕ должен быть основан на цели и задачах исследования. Например, оценка эффективности продолжительной программы коррекции массы тела среди лиц с ожирением не может считаться целесообразной в отношении генеральной совокупности всех лиц с ожирением (АТЕ) вследствие различных трудностей, ограничивающих полноценное завершение подобных программ пациентами из целевой аудитории, и в данном случае следует рассчитывать показатель АТТ. Напротив, оценка эффективности информационной кампании по борьбе с курением (раздача соответствующих мотивирующих брошюр, социальная реклама и т. п.) может быть оценена с помощью показателя АТЕ, так как подобные мероприятия могут охватить всю генеральную совокупность практически без ограничений.

Метод PSM является методом коррекции влияния конфаундеров, используемым наряду с регрессионным анализом, поэтому сравнение эффективности данных двух методов представляет существенный интерес и обсуждается в литературе [7, 12]. Считается, что метод PSM более точен по сравнению с регрессионным анализом в том случае, когда в распоряжении исследователя имеется относительно небольшое число наблюдений на фоне большого числа конфаундеров, которые требуется включить в анализ. Выяснено также, что метод PSM обеспечивает менее смещенную оценку меры эффекта по сравнению с регрессионным анализом при наличии 7 и менее наблюдений на каждый конфаундер [10].

Как и в случае любого многоэтапного метода статистического анализа данных, для применения метода PSM необходимо специализированное статистическое программное обеспечение. В настоящей статье рассматривается профессиональный статистический программный пакет Stata 13.0 (Stata Corp, TX, USA) [4]. Демо-версию программы можно скачать с официального сайта разработчика — www.stata.com. Программа Stata позволяет задавать последовательность команд обработки данных и детализировать процесс анализа, который запускается путем ввода специализированных команд, имеющих определенный синтаксис [6]. Анализ также может проводиться с использованием выпадающего меню модуля «Statistics», но данный модуль имеет ограниченные функциональные возможности по сравнению с командным способом.

Далее будет рассмотрен синтаксис команд STATA и представлен алгоритм обработки данных с использованием различных вариантов PSM [12], а также интерпретация полученных результатов.

Для удобства восприятия в настоящей статье команды STATA будут записываться следующим образом:

- Жирный шрифт — команды обработки данных (например, **pscore**).

— Курсив — переменные, представленные в исходной базе данных (например, *gender*).

— Обычный шрифт — переменные, создаваемые в процессе анализа (например, *myscore*).

Запуск команд производится клавишей «Enter». Заметим, что в программе Stata в качестве десятичного разделителя используется не запятая, а точка, что важно при вводе значений в базу данных (иначе подобные переменные будут восприниматься программой как текстовые), а запятая используется для разделения разрядов. Программа имеет минимальную возможность воспринимать кириллические символы, поэтому для выбора названий переменных, открытия и сохранения файлов рекомендуется латиница и использование таких мест хранения на диске, которые в наименовании пути к файлу не имеют русскоязычных названий папок. В противном случае программа будет выдавать сообщение об ошибке.

Для начала работы с PSM необходимо дополнить базовую версию программы Stata специальными модулями, загрузив их из интернета. Для этого введем в поле «Command» программы STATA команду **findit pscore**.

После нажатия клавиши «Enter» программа откроет новое окно, в котором выберем модуль *st0026_2* и нажмем «click here to install» для установки модуля в имеющийся программный пакет.

Заметим, что в дальнейшем, если программа показывает, что введенная команда не распознана (сообщение «unrecognized command»), необходимо с помощью команды **findit** аналогичным образом найти соответствующий модуль и установить его (потребуется в отношении команды **psmatch2**).

Для наглядного примера выполнения PSM рассмотрим фрагмент базы данных, которые были собраны в ходе исследования, направленного на изучение метаболического синдрома и его детерминант в г. Туркестан (Казахстан) [3]. Одной из задач данного исследования была оценка связи между наличием у пациентов метаболического синдрома и уровнем гемоглобина.

Для загрузки базы данных в программу необходимо загрузить ее с сайта журнала (файл *STATams.dta*) и открыть через меню «File» — «Open».

В базе данных представлены следующие переменные:

— *bmi* (кодирует группу наблюдения): имеет значение «1» при отсутствии у пациента метаболического синдрома (основная группа) и значение «0» при его наличии (контрольная группа) кодирует «лечение» и имеет бинарный характер. Казалось бы, логичным было бы закодировать цифрой «1» наличие метаболического синдрома, но в случае PSM в качестве основной группы целесообразно использовать ту группу, в которой имеется меньшее число наблюдений, для расширения возможностей по подбору пар из группы сравнения к наблюдениям основной группы. В нашей базе данных 228 наблюдений, из которых 181 пациент имеет метаболический синдром, а оставшиеся 47 пациентов — нет;

— *hgb* (кодирует исход): уровень гемоглобина, количественная переменная;

— *gender, age, smoke, educ* — соответственно пол, возраст, табакокурение и уровень образования участника исследования. В нашем случае для уменьшения громоздкости примера были выбраны только эти ковариаты, из которых первая имеет дихотомический тип, вторая — количественный, третья и четвертая — порядковый.

Разумеется, в исследованиях, как правило, учитывается значительно более широкий набор конфаундеров, но в нашем примере будут использованы только четыре из них для уменьшения громоздкости изложения наглядного материала.

Далее рассмотрим пошаговый алгоритм анализа, конечной целью которого является расчет мер эффекта — значений АТТ или АТЕ.

1. Установка случайного порядка наблюдений

Случайный порядок наблюдений в базе данных важен для правильного расчета значений PS. Для установки случайного порядка наблюдений построчно введем в поле «Command» следующие команды, запуская каждую клавишей «Enter»:

```
set seed 123456
gen x=uniform()
sort x
```

2. Расчет значений PS

Введем в поле «Command» следующую команду:

```
pscore bmi gender age smoke educ,
pscore(myscore) blockid(myblock) logit
comsup
```

Заметим, что сразу после команды **pscore** в синтаксисе команды следует название переменной, кодирующей изучаемый фактор (в нашем случае — переменная, кодирующая группу наблюдения), после чего последовательно указываются названия ковариат. Команда создает новую переменную со значениями PS, имеющую название, которое мы указываем в скобках (в нашем случае выбрано название *myscore*). Команда **logit** указывает на то, что мы используем логистическую регрессию для расчета значений PS (если ее удалить, то будет использована пробит-модель, что также допустимо). Опция **comsup** фиксирует зону «перекреста» значений PS между основной группой и группой сравнения («common support»).

Если после опции **logit** печатать опцию **detail**, то программа детализирует все промежуточные расчеты, что может быть полезным для понимания процесса стратификации PS.

В результате выполнения команды **pscore** программа создаст в базе несколько новых переменных: значение PS для каждого наблюдения (переменная *myscore*), номер блока, в который объединены значения PS (переменная *myblock*) и факт соответствия каждого наблюдения области «перекреста» (переменная *comsup*). Созданные переменные можно увидеть, если войти в меню «Data» — «Data Editor» — «Data Editor (Edit)» (рис. 1).

| | bmi | hgb | gender | age | smoke | educ | x | mypscore | myblock | comsup |
|----|-----|-----|--------|-----|-------|------|----------|-----------|---------|--------|
| 1 | 0 | 159 | 1 | 57 | 0 | 1 | .0031677 | .19556506 | 1 | 1 |
| 2 | 0 | 153 | 1 | 40 | 0 | 1 | .0119713 | .46213224 | 3 | 1 |
| 3 | 0 | 129 | 0 | 66 | 0 | 2 | .015148 | .04789651 | 1 | 1 |
| 4 | 0 | 123 | 0 | 54 | 0 | 3 | .0227852 | .09433381 | 1 | 1 |
| 5 | 0 | 145 | 0 | 48 | 0 | 1 | .0248369 | .18399443 | 1 | 1 |
| 6 | 0 | 152 | 1 | 32 | 1 | 1 | .0252959 | .54403677 | 3 | 1 |
| 7 | 0 | 130 | 0 | 56 | 0 | 3 | .0266694 | .08238619 | 1 | 1 |
| 8 | 0 | 127 | 0 | 52 | 0 | 3 | .0301769 | .1078105 | 1 | 1 |
| 9 | 0 | 168 | 0 | 50 | 0 | 3 | .0369958 | .12295113 | 1 | 1 |
| 10 | 0 | 141 | 1 | 48 | 0 | 3 | .0395636 | .25490804 | 2 | 1 |
| 11 | 0 | 131 | 0 | 48 | 0 | 3 | .0438138 | .13988468 | 1 | 1 |
| 12 | 1 | 144 | 1 | 32 | 0 | 1 | .0463763 | .60881863 | 4 | 1 |

Рис. 1. Переменные базы данных STATAms.dta.

В процессе обработки данных программа последовательно представляет результаты расчетов, которые мы будем рассматривать, фиксируя внимание на ключевых моментах.

В первую очередь программа представляет описательную статистику вариационного ряда значений PS в виде процентилей (рис. 2), причем учитываются только значения PS, ограниченные областью перекреста: медиана значений PS равна 0,14 (процентиль «50%»), среднее значение — 0,20 («Mean»). Следует отметить, что в нашем примере в зону перекреста попадают 226 случаев из имеющихся в базе данных 228 наблюдений («Obs»).

| Percentiles | Smallest | | |
|-------------|----------|----------|-------------|
| 1% | .0439931 | .0373434 | |
| 5% | .0506804 | .0409733 | |
| 10% | .0625373 | .0439931 | Obs |
| 25% | .0943338 | .0439931 | Sum of Wgt. |
| | | | 226 |
| 50% | .1447108 | | Mean |
| | | | .2045035 |
| 75% | .2749465 | .6261202 | Std. Dev. |
| | | | .1499916 |
| 90% | .4434849 | .6768656 | Variance |
| | | | .0224975 |
| 95% | .5360098 | .7235613 | Skewness |
| | | | 1.292071 |
| 99% | .6768656 | .7352883 | Kurtosis |
| | | | 4.075862 |

Рис. 2. Описательная статистика значений PS для области «перекреста».

Для наглядного представления о распределении значений PS в основной группе и в группе сравнения сформируем гистограмму распределения (рис. 3) с помощью соответствующей команды:

psgraph, treated(bmi) pscore(mypscore)

Из представленной двойной гистограммы видно, что распределение значений PS в основной и референтной группах существенно различается, то есть имеются существенные различия между группами по распределению конфаундеров (пол, возраст, активность табакокурения и уровень образования).

3. Формирование страт (блоков) PS и проверка баланса (равновесия) значений PS и ковариат между основной группой и группой сравнения в пределах каждого блока PS

Далее команда **pscore** разделяет выборку на наблюдений на блоки и проверяет, насколько сбалансированными оказались блоки как по значениями PS, так и по каждой из ковариат. Балансировка PS

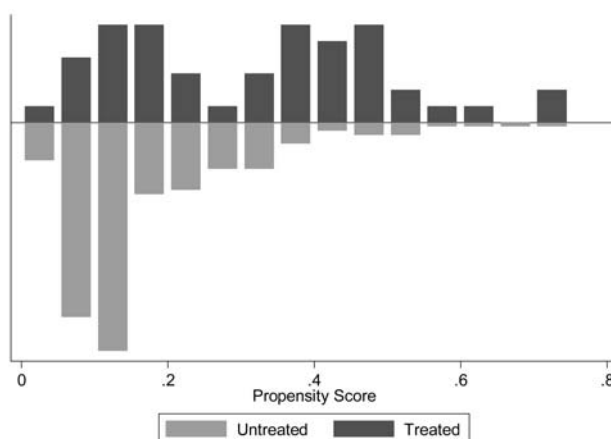


Рис. 3. Распределение значений PS в группе пациентов без метаболического синдрома («Treated») и группе пациентов с метаболическим синдромом («Untreated»).

проверяется путем сравнения средних значений PS в основной и контрольной группе в пределах каждого блока: если статистических различий нет, то такой блок считается сбалансированным. Если же значения PS в пределах блока между группами различаются, этот блок разделяется надвое и анализ повторяется, пока баланс PS не будет достигнут в каждом из блоков. Соответственно, если в начале анализа программа, как правило, формирует 4–5 блоков, то к концу процесса балансировки PS их количество может увеличиться.

В нашем случае количество блоков осталось равным 4. На рис. 4 представлено количество наблюдений в каждом блоке с разделением на группы сравнения (графа «bmi»).

| Inferior of block of pscore | bmi | | Total |
|-----------------------------|-----|----|-------|
| | 0 | 1 | |
| .0373434 | 125 | 17 | 142 |
| .2 | 43 | 13 | 56 |
| .4 | 9 | 14 | 23 |
| .6 | 2 | 3 | 5 |
| Total | 179 | 47 | 226 |

Note: the common support option has been selected

Рис. 4. Результаты деления выборки на блоки

Программа также проверяет все включенные в PSM ковариаты на предмет баланса между основной группой и группой сравнения в пределах каждого блока PS.

Сравнение осуществляется таким же образом, как на предыдущем этапе, но сравниваемыми вариационными рядами являются уже не значения переменной *mypscore*, а средние значения ковариат *gender*, *age*, *smoke* и *educ*.

Если в результате анализа выявлен дисбаланс какой-либо ковариаты в пределах одного или нескольких блоков, программа проинформирует об этом сообщением «The balancing property is not satisfied. Try a different specification of the propensity score». Также программа красным цветом выделит текст, в котором укажет, какие именно ковариаты и в каких блоках оказались несбалансированными.

В нашем случае баланс всех ковариат достигнут во всех блоках: «The balancing property is satisfied».

Что делать, если баланс всех ковариат во всех блоках не достигнут? Согласно опыту применения PSM, идеального баланса всех ковариат в каждом блоке, как правило, достигнуть не удастся. Рекомендуется взвесить потенциальное влияние несбалансированного конфаундера на результаты исследования: если на основании теоретических данных оно относительно невелико, то небольшой дисбаланс таких ковариат допустим. В ином случае существует ряд способов достижения баланса: использование не логит-, а пробит-модели расчета PS (для этого потребуется убрать опцию **logit** из команды **pscore**); удаление из анализа переменных, связанных с уже включенными в анализ ковариатами; включение в анализ одной или нескольких дополнительных ковариат; преобразование несбалансированных ковариат.

Далее рассмотрим методы, с помощью которых производится балансировка конфаундеров между основной и контрольной группами.

4. Подбор метода, позволяющего наиболее адекватно сбалансировать основную и контрольную группы по всем ковариатам

4.1. Метод стратификации (субклассификации) с использованием блоков PS

Метод использует стратификацию на основании сформированных ранее блоков PS. Данный способ позволяет рассчитать только значение АТТ:

```
atts hgb bmi gender age smoke,
pscore(mypscore) blockid(myblock) boot
reps(100)
```

Данная команда позволяет также уточнить значение стандартной ошибки АТТ методом бутстрэппинга (определение статистик вероятностных распределений, основанное на многократной генерации псевдобыборок на основе имеющейся выборки), причем количество повторов бутстрэппинга можно задавать вручную (в нашем примере заданы 100 повторов).

Результаты расчетов представлены на рис. 5. Программа представляет количество включенных в анализ наблюдений основной группы (47), группы

сравнения (179) и интересующую нас величину АТТ со стандартной ошибкой ($-7,4 \pm 4,3$).

Таким образом, в результате анализа установлено, что уровень гемоглобина у пациентов без метаболического синдрома (основная группа) в среднем на $(7,4 \pm 4,3)$ г/л ниже, чем у пациентов с метаболическим синдромом.

| n. treat. | n. contr. | ATT | Std. Err. | t |
|-----------|-----------|--------|-----------|--------|
| 47 | 179 | -7.364 | 4.283 | -1.719 |

Рис. 5. Результаты расчета АТТ с помощью метода стратификации (субклассификации)

4.2. Метод подбора пар 1:1 с помощью поиска «ближайшего соседа», соответствующего заданному отрезку PS

В данном случае каждому наблюдению основной группы подбирается «пара» из группы контроля, имеющая наиболее близкое значение PS, причем адекватный выбор значения отрезка PS позволит сделать так, чтобы наблюдения в парах были максимально похожими друг на друга по имеющемуся набору конфаундеров.

Значение данного отрезка определяется логитом PS. Для расчета логита PS потребуется последовательно ввести следующие две команды:

```
gen logitpscore = ln(mypscore /
(1-mypscore))
sum logitpscore
```

В результате программа выведет короткую таблицу (не приводится), в которой будет указано стандартное отклонение («Std. Dev.») логита PS (в нашем примере — 0,934). Установлено, что значение отрезка, равное 0,2 от стандартного отклонения логита PS, является вполне достаточным для обеспечения близкого соотношения наблюдений основной группы и группы контроля [8, 12].

Соответственно 0,2 стандартных отклонения логита PS будут равны $0,2 \times 0,934 = 0,19$.

Применяемая для подбора пар команда **psmatch2** использует рассчитанные ранее значения PS:

```
psmatch2 bmi, pscore(mypscore)
outcome(hgb) common caliper(0.19)
neighbor(1) ate
```

Результаты расчетов представлены на рис. 5. В верхней таблице на рис. 6 в графе «Difference» в соответствующих строках представлены значения мер эффекта: АТТ ($-5,98 \pm 6,31$) и АТЕ ($-3,66$). В нижней таблице представлено количество включенных в анализ наблюдений: из анализа были исключены 2 наблюдения группы сравнения, так как они не имели подходящей «пары». Заметим, что в настоящем примере базы данных нас интересует показатель АТТ, а не АТЕ, так как получение представлений о том, как бы изменился уровень гемоглобина во всей генеральной совокупности, если бы у всех пациентов в популяции

| Variable | Sample | Treated | Controls | Difference | S.E. | T-stat |
|----------|-----------|------------|------------|-------------|------------|--------|
| hgb | Unmatched | 123.506383 | 132.342541 | -8.83615846 | 3.40766609 | -2.59 |
| | ATT | 123.506383 | 129.489362 | -5.98297872 | 6.3112144 | -0.95 |
| | ATU | 132.234637 | 129.188827 | -3.04581006 | . | . |
| | ATE | | | -3.65663717 | . | . |

Note: S.E. does not take into account that the propensity score is estimated.

| psmatch2: Treatment assignment | psmatch2: Common support | | | Total |
|--------------------------------------|-----------------------------|------------|--|-------|
| | Off supp | On support | | |
| Untreated | 2 | 179 | | 181 |
| Treated | 0 | 47 | | 47 |
| Total | 2 | 226 | | 228 |

Рис. 6. Результаты расчета ATT и ATE при использовании метода подбора пар 1 : 1

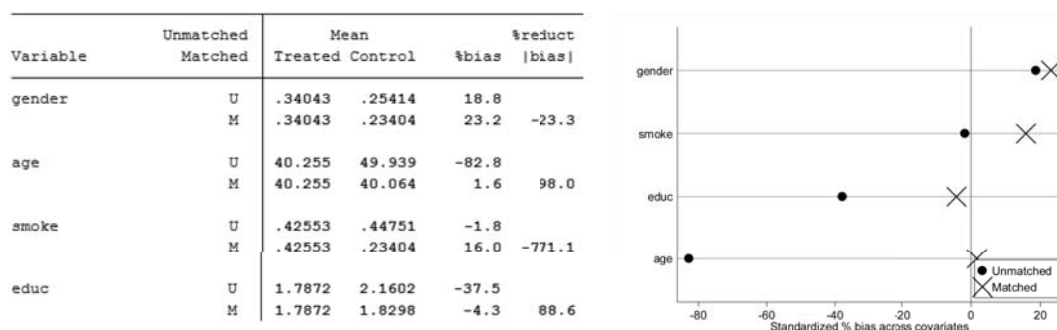


Рис. 7. Результаты проверки смещения конфаундеров в основной группе и группе контроля до (U) и после (M) подбора пар с использованием метода поиска «ближайшего соседа» 1 : 1

«исчез» метаболический синдром, в данном контексте не имеет практического смысла. Тем не менее в примерах мы будем использовать универсальный синтаксис команд, позволяющий рассчитать значения как ATT, так и ATE.

Но прежде чем принимать во внимание полученный результат, необходимо удостовериться в том, что ковариаты были адекватно сбалансированы между основной и контрольной группами. Для этого введем следующую команду:

```
psctest gender age smoke educ,
treated(bmi) both graph
```

На рис. 7 представлены результаты проверки балансировки ковариат после подбора пар. В таблице показано стандартизированное смещение средних значений всех 4 ковариат после проведения процедуры подбора пар (графа «%reduct |bias|», строки «M»): 23,2 % для ковариаты *gender*; 1,6 % для ковариаты *age*; 16,0 % для ковариаты *smoke* и -4,3 % для ковариаты *educ*. На представленном графике эта информация выглядит как приближение стандартизированной доли смещения для каждой ковариаты («Standardized % bias across covariates») к нулевому значению после процедуры подбора пар или «взвешивания». Таким образом, чем более сбалансированы ковариаты, тем ближе к нулевому значению будут располагаться соответствующие точки на графике (указаны знаком «x»). В данном случае для некото-

рых ковариат стандартизированное смещение средних значений оказалось большим, чем 10 %, что говорит о неудовлетворительном результате их балансировки.

Для достижения адекватной балансировки попробуем изменить «настройки» метода подбора пар. Например, можно уменьшить величину отрезка PS (сделать его значительно меньшим, чем 0,19) или увеличить количество наблюдений из группы сравнения, которые подбираются к каждому наблюдению основной группы (например, вместо подбора пар 1 : 1 использовать подбор пар 1 : 2). Используем второй вариант:

```
psmatch2 bmi, pscore(mypscore)
outcome(hgb) common caliper(0.19)
neighbor(2) ate
```

Результаты расчета ATE и ATT представлены на рис. 8. Результаты проверки балансировки ковариат представлены на рис. 9.

В данном случае балансировку ковариат можно считать удовлетворительной, так как после подбора пар 1 : 2 для каждой ковариаты доля стандартизированного смещения составила менее 10 %.

Таким образом, в результате анализа установлено, что уровень гемоглобина у пациентов без метаболического синдрома (основная группа) в среднем на $(7,2 \pm 5,2)$ г/л ниже, чем у пациентов с метаболическим синдромом. Заметим, что рассчитанное значение стандартной ошибки в данном случае не может

| Variable | Sample | Treated | Controls | Difference | S.E. | T-stat |
|----------|-----------|------------|------------|-------------|------------|--------|
| hgb | Unmatched | 123.506383 | 132.342541 | -8.83615846 | 3.40766609 | -2.59 |
| | ATT | 123.506383 | 130.680851 | -7.17446809 | 5.16863376 | -1.39 |
| | ATU | 132.234637 | 128.881564 | -3.35307263 | . | . |
| | ATE | | | -4.14778761 | . | . |

Note: S.E. does not take into account that the propensity score is estimated.

| | | | |
|----------------------|--------------------------|-----------|-------|
| psmatch2: | psmatch2: Common support | | |
| Treatment assignment | Off suppo | On suppor | Total |
| Untreated | 2 | 179 | 181 |
| Treated | 0 | 47 | 47 |
| Total | 2 | 226 | 228 |

Рис. 8. Результаты расчета ATT и ATE при использовании метода подбора пар 1 : 2

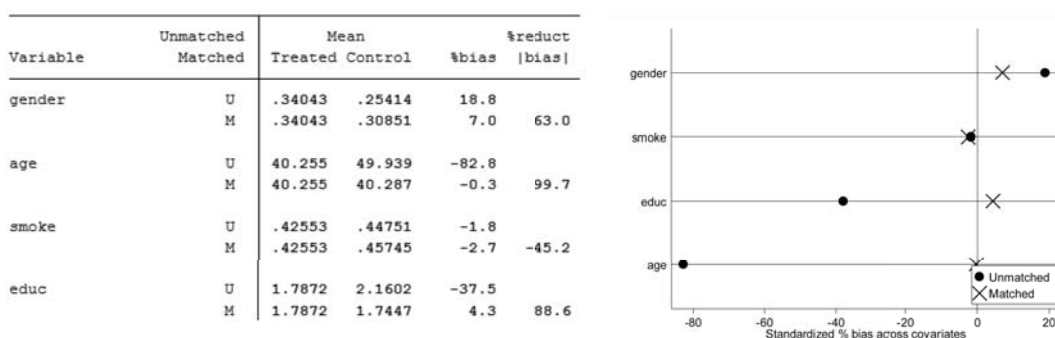


Рис. 9. Результаты проверки смещения конфаундеров в основной группе и группе контроля до (U) и после (M) подбора пар с использованием метода поиска «ближайшего соседа» 1 : 2

считаться очень точным, так как не учитывает факта предварительного расчета значений PS (сообщение программы «Note: S.E. does not take into account that the propensity score is estimated»).

4.3. Метод взвешивания Кернела

Несмотря на то, что адекватная балансировка ковариат была достигнута на предыдущем шаге анализа, выполним еще один способ балансировки, при котором каждому наблюдению контрольной группы придается определенный «вес» в зависимости от того,

насколько близко оно расположено к наблюдению основной группы по значению PS:

```
psmatch2 bmi, kernel outcome(hgb)
pscore(mypscore) ate
```

Результаты расчета ATT и ATE и результаты проверки баланса ковариат представлены на рис. 10 и рис. 11 соответственно.

В данном случае балансировка ковариат оказалась существенно более адекватной, чем при использовании метода подбора пар 1 : 2, а значит, и степень

| Variable | Sample | Treated | Controls | Difference | S.E. | T-stat |
|----------|-----------|------------|------------|-------------|------------|--------|
| hgb | Unmatched | 123.506383 | 132.342541 | -8.83615846 | 3.40766609 | -2.59 |
| | ATT | 123.506383 | 133.373298 | -9.86691547 | 4.26156106 | -2.32 |
| | ATU | 132.342541 | 129.042062 | -3.30047976 | . | . |
| | ATE | | | -4.65408712 | . | . |

Note: S.E. does not take into account that the propensity score is estimated.

| | | | |
|----------------------|--------------------------|--|-------|
| psmatch2: | psmatch2: Common support | | |
| Treatment assignment | On suppor | | Total |
| Untreated | 181 | | 181 |
| Treated | 47 | | 47 |
| Total | 228 | | 228 |

Рис. 10. Результаты расчета ATT и ATE при использовании метода взвешивания Кернела

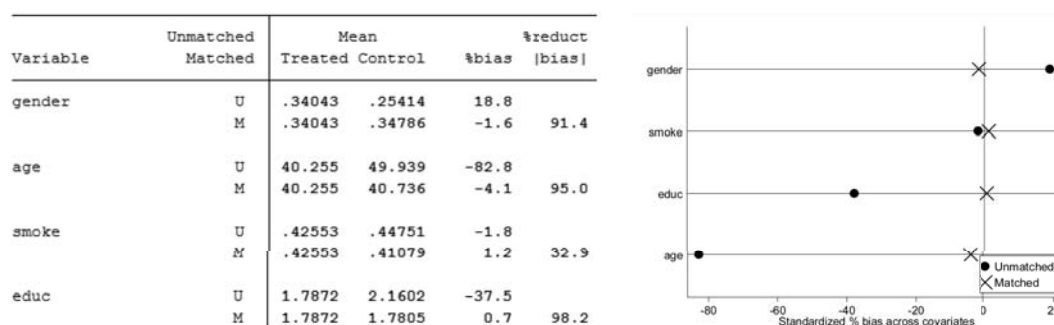


Рис. 11. Результаты проверки смещения конфаундеров в основной группе и группе контроля до (U) и после (M) подбора пар с использованием метода взвешивания Кернела

| hgb | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] | |
|--------|-----------|-----------|-------|-------|----------------------|-----------|
| bmi | -7.478707 | 3.132537 | -2.39 | 0.018 | -13.65202 | -1.305393 |
| gender | 23.01487 | 3.104728 | 7.41 | 0.000 | 16.89635 | 29.13338 |
| age | .363251 | .1124951 | 3.23 | 0.001 | .141566 | .5849559 |
| smoke | .0774382 | 1.199362 | 0.06 | 0.949 | -2.286152 | 2.441029 |
| educ | -.4725919 | 1.203901 | -0.39 | 0.695 | -2.845128 | 1.899945 |
| _cons | 109.3387 | 5.900077 | 18.53 | 0.000 | 97.71141 | 120.9661 |

Рис. 12. Результаты множественного линейного регрессионного анализа

искажения результатов стала меньше. Согласно результатам расчетов по методу «взвешивания» Кернела, уровень гемоглобина у пациентов без метаболического синдрома в среднем на $(9,9 \pm 4,3)$ г/л ниже, чем у пациентов с метаболическим синдромом.

Итак, применительно к имеющейся базе данных мы использовали несколько методов PSM для расчета отличия уровня гемоглобина у пациентов без метаболического синдрома по сравнению с имеющими метаболический синдром пациентами.

Сравним полученные результаты с результатами множественного линейного регрессионного анализа, обычно используемого для анализа подобных данных:

```
regress hgb bmi gender age smoke educ
```

На рис. 12 представлены результаты множественной линейной регрессии, из которых нас интересует коэффициент для переменной *bmi* (графа «Coef.»), равный $-7,5$, то есть, согласно данному методу, у пациентов без метаболического синдрома уровень гемоглобина в среднем на $7,5$ г/л ниже, чем у пациентов с метаболическим синдромом.

В таблице представлены результаты использования всех способов PSM и множественной линейной регрессии.

Различия между средними значениями гемоглобина в основной и контрольной группах с использованием различных методов анализа

| Метод | АТТ, г/л |
|--|----------|
| Множественная линейная регрессия | -7,5 |
| Метод стратификации (субклассификации) | -7,4 |
| Метод подбора пар 1 : 2 | -7,2 |
| Метод взвешивания Кернела | -9,9 |

В нашем случае возникает вопрос, какой из результатов использования методов PSM выбрать в качестве наиболее корректного? На этот вопрос нет однозначного ответа, так как каждый из методов PSM имеет свои особенности. Метод стратификации (субклассификации) является в историческом плане самой первой вариацией PSM и может считаться достаточно «грубым» способом балансировки ковариат. Напротив, методы, реализуемые с помощью команды **psmatch2**, были созданы позже и являются методологически более совершенными. В приведенном примере метод «взвешивания» Кернела позволил наиболее адекватно сбалансировать действие конфаундеров по сравнению с методом подбора пар и поэтому в случае нашего исследования его выбор можно считать наиболее целесообразным.

Список литературы

1. Гржибовский А. М., Иванов С. В. Когортные исследования в здравоохранении // Наука и Здравоохранение. 2015. № 3. С. 5–16.
2. Гржибовский А. М., Иванов С. В. Поперечные (одномоментные) исследования в здравоохранении // Наука и Здравоохранение. 2015. № 2. С. 5–18.
3. Садыкова К. Ж., Шалхарова Ж. С., Нускабаева Г. О., Садыкова А. Д., Жунисова М. Б., Маденбай К. М., Гржибовский А. М. Распространенность анемии, ее социально-демографические детерминанты и возможная связь с метаболическим синдромом в г. Туркестан, Южный Казахстан // Экология человека. 2015. № 8. С. 58–64.
4. Унгуяну Т. Н., Гржибовский А. М. Программное обеспечение для статистической обработки данных STATA: введение // Экология человека. 2014. № 1. С. 60–63.
5. Флетчер Р., Флетчер С., Вагнер Э. Клиническая эпидемиология. Основы доказательной медицины. М. : Медиа Сфера, 1998. 352 с.
6. Acock A. C. Gentle Introduction to Stata. USA, Texas : Stata Press, 2006. 289 p.

7. Austin P. C. An introduction to propensity score methods for reducing the effects of confounding in observational studies // *Multivariate Behavioral Research*. 2011. Vol. 46. P. 399–424.

8. Austin P. C. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies // *Pharmaceutical statistics*. 2011. Vol. 10 (2). P. 150–161.

9. Becker S. O., Ichino A. Estimation of average treatment effects based on propensity scores // *The Stata Journal*. 2002. Vol. 2 (4). P. 358–377.

10. Cepeda M. S., Boston R., Farrar J. T., Strom B. L. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders // *American Journal of Epidemiology*. 2003. Vol. 158 (3). P. 280–287.

11. D'Agostino R. B. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group // *Statistics in Medicine*. 1998. Vol. 17 (19). P. 2265–2281.

12. Garrido M. M., Kelley A. S., Paris J., Roza K., Meier D. E., Morrison R. S., Aldridge M. D. Methods for constructing and assessing propensity scores // *Health Services Research*. 2014. Vol. 49 (5). P. 1701–1720.

13. Guo Sh. Y., Mark W. Fraser. Propensity Score Analysis: Statistical Methods and Applications, 2nd ed. USA. SAGE Publications, 2015. 448 p.

14. Hernan M. A., Hernandez-Diaz S., Robins J. M. A structural approach to selection bias // *Epidemiology*. 2004. Vol. 15 (5). P. 615–625.

15. Patorno E., Grotta A., Bellocco R., Schneeweiss S. Propensity score methodology for confounding control in health care utilization databases // *Epidemiology Biostatistics and Public Health*. 2013. Vol. 10 (3). P. e8940.

16. Rosenbaum P. R., Rubin D. B. The central role of the propensity score in observational studies for causal effects // *Biometrika*. 1983. Vol. 70 (1). P. 41–55.

17. Sturmer T., Joshi M., Glynn R. J., Avorn J., Rothman K. J., Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods // *Journal of Clinical Epidemiology*. 2006. Vol. 59 (5). P. 437–447.

References

1. Grjibovski A. M., Ivanov S. V. Cohort studies in health sciences. *Nauka i Zdravooohranenie* [Science & Healthcare]. 2015, 3, pp. 5-16. [in Russian]

2. Grjibovski A. M., Ivanov S. V. Cross-sectional studies in health sciences. *Nauka i Zdravooohranenie* [Science & Healthcare]. 2015, 2, pp. 5-18. [in Russian]

3. Sadykova K. Zh., Shalkharova Zh. S., Shalkharova Zh. N., Nuskabaeva G. O., Sadykova A. D., Zhunisova M. B., Madenbay K. M., Grjibovski A. M. Prevalence of anemia, its socio-demographic determinants and potential association with metabolic syndrome in residents of Turkestan, Southern Kazakhstan. *Ekologiya cheloveka* [Human Ecology]. 2015, 8, pp. 58-64. [in Russian]

4. Unguryanu T. N., Grjibovski A. M. Introduction to STATA - software for statistical analysis. *Ekologiya cheloveka* [Human ecology]. 2014, 1, pp. 60-63. [in Russian]

5. Fletcher R., Fletcher C., Vagner E. Klinicheskaya epidemiologiya. *Osnovy dokazatel'noi meditsiny* [Clinical

epidemiology. Basics of the evidence-based medicine]. Moscow, Media Sphere Publ., 1998, 352 p.

6. Acok A. C. Gentle Introduction to Stata. USA, Texas, Stata Press, 2006, 289 p.

7. Austin P. C. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*. 2011, 46, pp. 399-424.

8. Austin P. C. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical statistics*. 2011, 10 (2), pp. 150-161.

9. Becker S. O., Ichino A. Estimation of average treatment effects based on propensity scores. *The Stata Journal*. 2002, 2 (4), pp. 358-377.

10. Cepeda M. S., Boston R., Farrar J. T., Strom B. L. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *American Journal of Epidemiology*. 2003, 158 (3), pp. 280-287.

11. D'Agostino R. B. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*. 1998, 17 (19), pp. 2265-2281.

12. Garrido M. M., Kelley A. S., Paris J., Roza K., Meier D. E., Morrison R. S., Aldridge M. D. Methods for constructing and assessing propensity scores. *Health Services Research*. 2014, 49 (5), pp. 1701-1720.

13. Guo Sh. Y., Mark W. Fraser. Propensity Score Analysis: Statistical Methods and Applications, 2nd ed. USA. SAGE Publications, 2015. 448 p.

14. Hernan M. A., Hernandez-Diaz S., Robins J. M. A structural approach to selection bias. *Epidemiology*. 2004, 15 (5), pp. 615-625.

15. Patorno E., Grotta A., Bellocco R., Schneeweiss S. Propensity score methodology for confounding control in health care utilization databases. *Epidemiology Biostatistics and Public Health*. 2013, 10 (3), p. e8940.

16. Rosenbaum P. R., Rubin D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983, 70 (1), pp. 41-55.

17. Sturmer T., Joshi M., Glynn R. J., Avorn J., Rothman K. J., Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *Journal of Clinical Epidemiology*. 2006, 59 (5), pp. 437-447.

Контактная информация:

Гржибовский Андрей Мечиславович — доктор медицины, старший советник Национального института общественного здравоохранения, г. Осло, Норвегия; руководитель отдела международных программ и инновационного развития ЦНИИ Северного государственного медицинского университета, г. Архангельск, Россия; профессор кафедры общественного здоровья и здравоохранения медицинского института Северо-Восточного федерального университета, г. Якутск, Россия; профессор, почетный доктор Международного казахско-турецкого университета, г. Туркестан, Казахстан; почетный профессор ГМУ г. Семей, Казахстан
Адрес: INFA, Nasjonalt folkehelseinstitutt, Postboks 4404 Nydalen, 0403 Oslo, Norway.

Е-mail: Andrej.Grjibovski@gmail.com

Тел.: +4745268913 (Норвегия), +79214717053 (Россия), +77471262965 (Казахстан)