

УДК 303.725.33; 330.43

DOI: 10.17323/1813-8691-2025-29-2-337-374

## Эффект воздействия высшего образования на здоровье в моделях возрастных когорт<sup>1</sup>

Косорукова М.А., Коссова Е.В., Потанин Б.С., Шило А.В.

Работа посвящена оцениванию среднего эффекта воздействия высшего образования на здоровье в предположении существования структурных различий в зависимостях между исследуемыми показателями по полу и когорте рождения индивидов. Оценивание эффектов проводится отдельно для каждой когорты мужчин и женщин с применением параметрических (многомерная рекурсивная пробит-модель) и непараметрических (мэтчинг) методов. Использование принципиально разных подходов позволяет проверить устойчивость результатов к нарушению предпосылок применяемых методов. В ходе исследования для женщин выявлены статистические свидетельства в пользу существования значимых отрицательных средних эффектов воздействия высшего образования на вероятность наличия ожирения и вероятность гипертонии и значимых положительных эффектов на вероятность высокой самооценки здоровья. Для мужчин найдены статистические свидетельства в пользу существования значимых положительных средних эффектов воздействия на вероятность наличия заболеваний сердца, вероятность гипертонии, аллергии и высокой самооценки здоровья. Обнаружена значительная дифференциация средних эффектов воздействия в зависимости от возрастной когорты. Для определенных видов заболеваний подтверждается гипотеза о когортном снижении отдачи от образования. Делается вывод о необходимости изучения влияния образования на здоровье индивидов в отдельности

<sup>1</sup> Работа выполнена при поддержке Российского научного фонда (РНФ), проект 21-18-00427.

**Косорукова Мария Александровна** — аспирант департамента прикладной экономики, факультет экономических наук Национального исследовательского университета «Высшая школа экономики». E-mail: kosorukova.mariya@gmail.com

**Коссова Елена Владимировна** — к.ф.-м.н., доцент департамента прикладной экономики, факультет экономических наук Национального исследовательского университета «Высшая школа экономики». E-mail: ekossova@hse.ru

**Потанин Богдан Станиславович** — к.э.н., доцент департамента прикладной экономики, факультет экономических наук Национального исследовательского университета «Высшая школа экономики». E-mail: bogdanpotanin@gmail.com

**Шило Александр Васильевич** — студент магистратуры, факультет экономических наук Национального исследовательского университета «Высшая школа экономики». E-mail: alexandr.vas.shilo@gmail.com

Статья поступила: 23.01.2025/Статья принята: 23.04.2025.

для мужчин и женщин, родившихся в разные периоды времени, а также о нецелесообразности использования агрегированных показателей здоровья при анализе воздействия образования. Результаты исследования позволяют определить характер когортных изменений во влиянии образования на состояние здоровья населения России и могут учитываться при формировании политики в сфере образования как источника положительных внешних эффектов на здоровье.

**Ключевые слова:** образование; здоровье; эффект воздействия; возрастные когорты; мэтчинг; многомерная пробит-модель.

**Для цитирования:** Косорукова М.А., Коссова Е.В., Потанин Б.С., Шило А.В. Эффект воздействия высшего образования на здоровье в моделях возрастных когорт. *Экономический журнал ВШЭ*. 2025; 29(2): 337-374.

**For citation:** Kosorukova M.A., Kossova E.V., Potanin B.S., Shylo A.V. Treatment Effect of Higher Education on Health in Age Cohort Models. *HSE Economic Journal*. 2025; 29(2): 337-374. (In Russ.)

## Введение

Тема взаимосвязи образования и здоровья нередко обсуждается в научной литературе. Став предметом дискуссий исследователей еще в конце прошлого столетия [Berger, Leigh, 1989], она не перестает быть актуальной и до сих пор, что подтверждается различными обзорными работами, описывающими многочисленные исследования последних лет, посвященные образованию и здоровью [Eide, Showalter, 2011; Hamad et al., 2018].

При исследовании связи образования и здоровья возникают три методологических вопроса. Первый из них относится к факту существования и характеру связи между уровнем образования и здоровьем индивидов. В исследованиях часто можно встретить подтверждение наличия корреляционной связи между ними [Albarrán et al., 2020; Silles, 2009]. Однако до сих пор в литературе отсутствует однозначный ответ на вопрос, является ли эта связь причинно-следственной. Одним из главных источников неопределенности, возникающей при ответе на данный вопрос, является сложность учета ненаблюдаемых факторов при проведении исследований на нерандомизированных данных [Mehta, 2001]. Главная проблема таких данных заключается в отсутствии у исследователя информации о механизме разделения объектов на группы. В частности, индивиды самостоятельно принимают решение о получении высшего образования под воздействием различных факторов, среди которых могут быть такие, которые одновременно влияют как на вероятность получения образования, так и на здоровье индивида. Наличие таких факторов среди ненаблюдаемых приводит к эндогенности переменной образования и возникновению смещения самоотбора [Schlotter et al., 2011]. По этой причине при оценивании средних эффектов воздействия возникает необходимость в применении особых методов, позволяющих учесть указанную проблему.

Существует два основных подхода к оцениванию эффектов воздействия – параметрический и непараметрический, каждый из которых имеет свои преимущества и недостатки. Для оценивания воздействия образования на здоровье чаще используются пара-

метрические методы оценивания, такие как метод инструментальных переменных, дизайн разрывной регрессии, построение системы бинарных уравнений. Однако главный недостаток параметрического подхода заключается в необходимости введения предположений о распределении случайных ошибок и функциональной форме связи между зависимой переменной и объясняющими факторами. В этом смысле непараметрические методы оценивания, такие как, например, мэтчинг, оказываются более предпочтительными, поскольку не требуют выдвижения подобных предположений. Однако и у них есть свои недостатки, заключающиеся, в том числе, в сложности соблюдения требований к отбираемым для оценивания переменным. Сравнение разных подходов, проводимое в настоящем исследовании, позволит выявить устойчивые к нарушению предпосылок результаты.

Второй вопрос касается переменных, которые исследователи используют как прокси для образования индивида и уровня его здоровья. В качестве индикатора уровня образования зачастую выступает число лет обучения [Delaruelle et al., 2015]. При этом у такой аппроксимации есть недостаток – количественные показатели уровня образования, как правило, довольно слабо отражают его качество. Причем большинство работ, в которых применяется данный показатель, посвящено воздействию конкретно школьного образования на здоровье индивидов, а именно изучается влияние одного дополнительного года обучения в школе [Clark, Royer, 2013; Fonseca et al., 2020; Silles, 2009]. Однако подобный подход к оцениванию эффектов связан со спецификой законодательства относительно школьного образования в том или ином государстве и приемлем не для любой страны. Кроме того, остается открытым вопрос о влиянии высшего образования на здоровье индивидов. Учитывая тот факт, что Россия принадлежит к числу стран-лидеров по доле населения с высшим образованием, изучение вопроса воздействия именно высшего образования на здоровье российского населения является актуальной задачей.

Рассматривая прокси-переменные здоровья индивидов, исследователи обычно используют самооценку здоровья или агрегированные показатели здоровья, такие как число имеющихся у индивида хронических заболеваний [Черкасов и др., 2019] или индикатор наличия любого хронического заболевания. В настоящем исследовании выдвигается гипотеза о том, что в России влияние высшего образования на наличие хронических заболеваний может быть разнонаправленным в зависимости от вида заболевания. Эта гипотеза подкрепляется результатами зарубежных исследований, в которых удалось выявить статистические свидетельства в пользу повышения вероятности наличия аллергии и онкологических заболеваний [Dalstra et al., 2005; Dilmaghani, 2020; Fonseca, Zheng, 2011], но при этом снижения вероятности наличия ожирения (когда индекс массы тела (ИМТ) превышает 30 кг/м<sup>2</sup>), сахарного диабета, гипертонии, повышенного уровня холестерина в крови и низкой самооценки здоровья [Davies et al., 2018; Fonseca, Zheng, 2011; Grabner, 2009; Viego, Temporelli, 2017] у образованных индивидов. Если схожая дифференциация справедлива и для России, то при использовании суммарных показателей здоровья влияние образования может быть не выявлено в силу взаимного компенсирования эффектов с различными знаками. Поэтому в данной работе в качестве показателей здоровья помимо субъективной оценки будут выступать переменные, отражающие наличие конкретных хронических заболеваний.

Наконец, третий методологический вопрос, возникающий при изучении взаимосвязи образования и здоровья и занимающий центральную позицию в данном исследовании, – вопрос об однородности наблюдений. Ключевой особенностью настоящей работы

является оценивание взаимосвязи образования и здоровья в разрезе возрастных когорт. Большинство предыдущих исследований проводилось без учета различий индивидов по периоду их рождения и, соответственно, получения образования. Однако в последнее время все большее распространение среди исследователей приобретает мнение, что образование, полученное в разные исторические периоды, может оказывать различное влияние на здоровье индивидов. Иными словами, существуют различия в эффектах образования в зависимости от возрастных когорт.

В ряде исследований [Delaruelle et al., 2015; Lynch, 2003] обсуждались возможные причины дифференциации отдачи от образования между представителями различных возрастных когорт. Были выдвинуты две конкурирующие гипотезы о характере и причинах соответствующих различий. Согласно гипотезе об увеличении отдачи (the cohort accretion hypothesis) предполагается, что чем моложе когорта, тем сильнее влияние образования на здоровье. Это может быть связано с тем, что более старшие когорты страдают от естественных возрастных проблем, поэтому образование может уже не оказывать значимого влияния на здоровье пожилых людей. При этом гипотеза о снижении отдачи (the cohort diminishing hypothesis), напротив, говорит об уменьшении силы воздействия образования для более молодых когорт: если эффект образования на здоровье является положительным и кумулятивным, то он окажется выше для представителей старших когорт в результате своего накопления в течение жизни индивидов.

В работе [Delaruelle et al., 2015] на данных по населению европейских стран было обнаружено, что чем моложе когорта, тем ниже отдача от образования, что подтверждает гипотезу о снижении отдачи. Полученный результат вызывал у авторов опасения по поводу возможного появления потенциальных барьеров на пути развития образования в тех или иных странах и, как следствие, развития общества в целом. Гипотеза о снижении отдачи также подтвердилась на данных по населению Соединенных Штатов Америки [Lynch, 2003]. Однако на российских данных соответствующие гипотезы протестированы не были, что подчеркивает актуальность изучения особенностей влияния образования на здоровье российского населения в разрезе возрастных когорт.

Таким образом, цель настоящего исследования заключается в выявлении когортных различий в эффектах образования на здоровье населения России. Выдвигаются гипотеза о наличии структурной неоднородности связи между образованием и переменными здоровья по полу и когорте рождения индивидов и гипотеза о когортном снижении отдачи от образования на здоровье. Для проверки данных гипотез эффект образования на различные аспекты здоровья оценивается отдельно для каждой когорты мужчин и женщин. При этом используются как параметрический, так и непараметрический подходы к оцениванию, что дает возможность проверить устойчивость оценок. Результаты исследования позволят определить характер изменений во влиянии образования на здоровье населения России и учитывать его при формировании политики в сфере образования.

## 1. Методология и данные

### 1.1. Понятие среднего эффекта воздействия

При изучении воздействия бинарной переменной на целевой показатель (outcome, исход объекта) удобно использовать концепцию эффекта воздействия [Ениколопов, 2009;

Rubin, 1974; Rubin, 1977]. Предположим, что на основе значения бинарной переменной воздействия (treatment)  $D_i$ ,  $i = 1, \dots, N$ , где  $N$  – размер выборки, все объекты выборки делятся на две группы – группу тех, кто подвергся изучаемому воздействию (т.е. группу воздействия,  $D_i = 1$ ), и группу тех, на кого воздействие не оказывалось (контрольная группа,  $D_i = 0$ ). В таком случае для каждого объекта существует пара потенциальных исходов:  $Y_i(1)$  – исход объекта с номером  $i$ , если бы он оказался в группе воздействия,  $Y_i(0)$  – исход этого же объекта с номером  $i$ , если бы он оказался в группе контроля. По определению истинный средний эффект воздействия (Average Treatment Effect, ATE) – это математическое ожидание разности потенциальных исходов:  $ATE = E(Y_i(1) - Y_i(0))$ , а его оценка, соответственно, это оценка данного математического ожидания:

$$\widehat{ATE} = \hat{E}(Y_i(1) - Y_i(0) | X_i),$$

где  $X_i$  – набор контрольных переменных.

Если в контрольные переменные включены все факторы, влияющие и на отбор объекта в группу, и на его исход, то оценку эффекта воздействия можно интерпретировать с точки зрения существования причинно-следственной связи между воздействием (в данной работе – высшим образованием) и исходом (переменной, которая выступает индикатором здоровья). Таким образом, величина  $\widehat{ATE}$  при выполнении предпосылок методов оценивания будет отражать, насколько в среднем выше или ниже вероятность определенного состояния здоровья для высокообразованных индивидов по сравнению с теми, у кого нет высшего образования.

Проблема с использованием в качестве оценки среднего эффекта воздействия простой разности средних потенциальных исходов объектов состоит в том, что в реальности объект может попасть только в одну группу, поэтому для него наблюдается всегда лишь один из двух исходов – или  $Y_i(1)$ , если объект попал в группу воздействия, или  $Y_i(0)$ , если он попал в контрольную группу. Следовательно, чтобы получить оценку среднего эффекта воздействия, необходимо применить особые методы (см. далее), которые позволили бы для каждого объекта оценить его второй (ненаблюдаемый) потенциальный исход, контролируя характеристики объектов.

## 1.2. Данные и их дескриптивный анализ

В работе используются данные РМЭЗ НИУ ВШЭ за 2009–2019 гг.<sup>2</sup> Анализ проводится отдельно по выборкам мужчин и женщин старше 18 лет для учета возможных гендер-

---

<sup>2</sup> «Российский мониторинг экономического положения и здоровья населения НИУ ВШЭ (RLMS – HSE)», проводимый Национальным исследовательским университетом «Высшая школа экономики» и ООО «Демоскоп» при участии Центра народонаселения Университета Северной Каролины в Чапел Хилле и Института социологии Федерального научно-исследовательского социологического центра РАН. (Сайты обследования RLMS-HSE: <https://rlms-hse.cpc.unc.edu> и <http://www.hse.ru/rlms>).

ных различий. По аналогии с работой [Kossova et al., 2021] были выделены шесть возрастных когорт, состоящих из людей, рожденных в 1935–1944 гг., 1945–1954 гг., 1955–1964 гг., 1965–1974 гг., 1975–1984 гг., 1985–1994 гг. Данное разбиение удобно, в частности, тем, что позволяет выделить довоенный и послевоенный периоды, а также период глобальных изменений в экономике России (начиная с 1985 г.). Кроме того, взятие десятилетних когорт дает возможность рассмотреть минимальные интервалы времени, за которые могли произойти какие-либо социально-экономические изменения, в том числе и в отношении образования, и при этом позволяет получить достаточно большое число наблюдений в каждой когорте.

Выбор временного периода (2009–2019 гг.) связан с ограничениями базы РМЭЗ, а также с пандемией коронавируса, учет которой мог бы значимым образом исказить результаты оценивания.

В качестве переменной воздействия в работе выступает переменная *diploma*, отражающая факт наличия диплома о высшем образовании. Она принимает значение единица, если у индивида есть оконченное высшее образование, и ноль в противном случае.

В качестве переменных исхода, т.е. целевых зависимых переменных, в исследовании используются бинарные переменные, которые отражают наличие у индивида различных хронических заболеваний<sup>3</sup>. На каждое из 10 рассмотренных в исследовании хронических заболеваний была сформирована соответствующая бинарная переменная исхода. В число этих 10 заболеваний вошли хронические заболевания сердца, легких, печени, почек, желудочно-кишечного тракта (ЖКТ), позвоночника, эндокринной системы, суставов, гипертоническая болезнь и хроническая аллергия. Также в качестве переменной исхода рассматривалось наличие ожирения ( $ИМТ > 30 \text{ кг/м}^2$ ) как одного из факторов риска развития определенных хронических заболеваний. Кроме того, проводилось оценивание среднего эффекта воздействия на вероятность наличия хотя бы одного любого хронического заболевания у индивида для проверки целесообразности использования агрегированных показателей здоровья в анализе эффектов воздействия на него.

Помимо этого, в качестве еще одной переменной исхода в работе выступила самооценка здоровья, принимавшая значение единица, если индивид оценивал свое здоровье как «Очень хорошее» или «Хорошее», и ноль в остальных случаях. Таким образом, в исследовании изучалось влияние высшего образования на различные аспекты здоровья, которые были учтены за счет 13 бинарных переменных исхода.

Переменные, используемые в качестве контрольных, представляют собой стандартный набор, который чаще всего учитывается в исследованиях, посвященных взаимосвязи образования и здоровья [Fonseca, Zheng, 2011; Viego, Temporelli, 2017; Коссова, Косорукова, 2023], и отражает социально-экономический статус человека и его индивидуальные характеристики. В частности, в состав контрольных переменных вошли возраст индивида (*age*), группа дамми-переменных, отражающих место окончания школы индивидом (*soblc* – дамми-переменная на окончание школы в Москве, Санкт-Петербурге или областном центре; *sgorod* – дамми-переменная на окончание школы в другом городе; базовая категория – окончание школы в поселке городского типа (ПГТ) или в селе), группа дамми-переменных, отражающих его текущее место проживания (*oblc*, *gorod*; базовая ка-

<sup>3</sup> Примечание: информация о наличии или отсутствии того или иного заболевания у индивида в базе РМЭЗ указывается со слов самого индивида.

тегория – проживание в ПГТ или в селе), наличие у индивида статуса инвалида (*invalid*), натуральный логарифм среднедушевого дохода домохозяйства, в котором проживает индивид, за последние 30 дней (*ln\_income*), продефлированного в соответствии с уровнем индекса потребительских цен к 2009 г.<sup>4</sup>, количество детей у индивида (*num\_of\_child*), бинарная переменная на семейное положение (*marriage* = 1, если индивид находится в официальном или гражданском браке, 0 иначе), частоту посещения врача (*doctor* = 1, если индивид посещает доктора 2 раза в год и чаще, 0 иначе), наличие работы (*work*), а также бинарные переменные, отражающие факт потребления индивидом алкоголя, факт курения и факт периодических занятий спортом (*alcohol* = 1, если индивид употребляет хотя бы иногда алкогольные напитки, включая пиво, 0 иначе; *smoking* = 1, если индивид курит в настоящее время, 0 иначе; *phys\_active* = 0, если индивид совсем не занимается физкультурой, 1 иначе).

Таблица 1.

**Число наблюдений в выборках по возрастным когортам  
и полу на данных РМЭЗ за 2009–2019 гг.**

	Число наблюдений	
	женщины	мужчины
Когорта людей, родившихся в следующие года:		
1935–1944	7699 (71%)	3093 (29%)
1945–1954	11920 (64%)	6706 (36%)
1955–1964	14132 (58%)	10043 (42%)
1965–1974	13041 (57%)	10015 (43%)
1975–1984	15541 (53%)	13523 (47%)
1985–1994	13371 (53%)	11830 (47%)

Согласно табл. 1, в каждой когорте женщины представлены в большей степени, чем мужчины. При этом чем старше когорта, тем разрыв между числом женщин и мужчин выше, что, вероятно, объясняется более высокой продолжительностью жизни женщин. Самой же многочисленной является когорта людей, родившихся в 1975–1984 гг.

Из данных о распределении высшего образования среди представителей различных когорт следует (см. рис. 1), что доля людей с высшим образованием по всей выборке колеблется от 22 до 35%. Для более старших когорт доля мужчин с высшим образованием немного выше доли женщин с высшим образованием. При этом для более молодых когорт ситуация меняется на противоположную и, начиная с когорты 1965–1974 гг., доля женщин с высшим образованием становится заметно выше, чем доля мужчин.

<sup>4</sup> Использовались данные Федеральной службы государственной статистики.

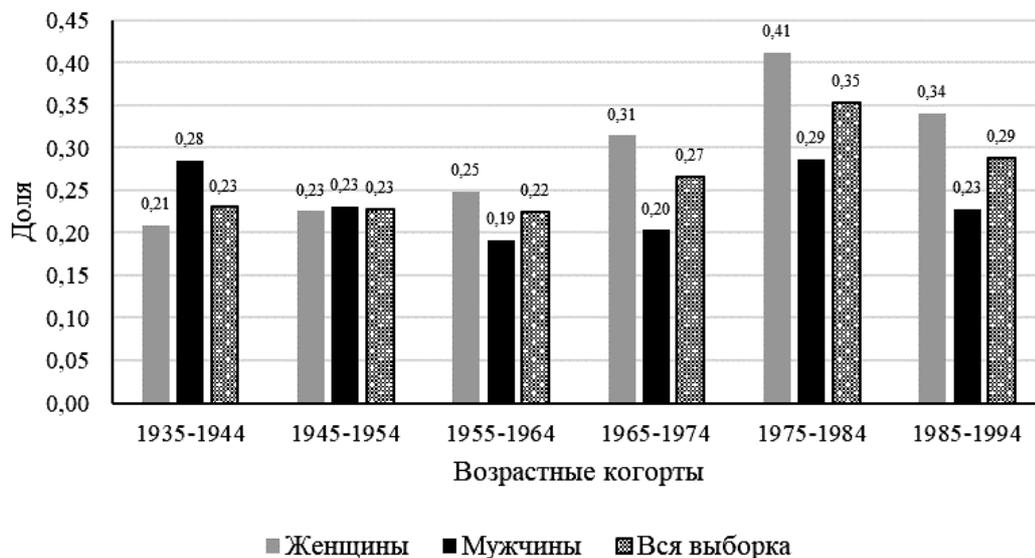


Рис. 1. Доли индивидов с высшим образованием по возрастным когортам и полу

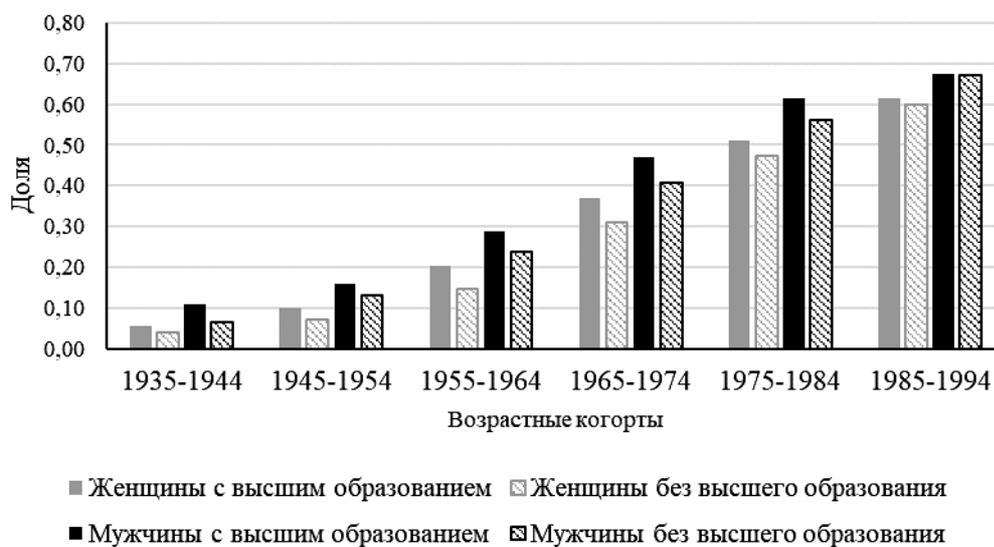
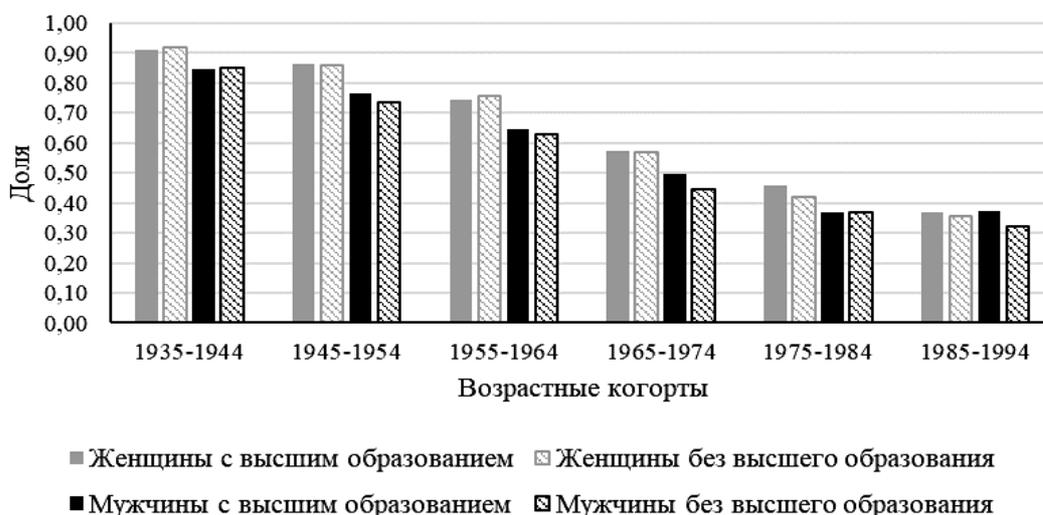


Рис. 2. Доли индивидов, оценивающих свое здоровье как «Хорошее» или «Очень хорошее», по возрастным когортам и полу

Согласно данным, проиллюстрированным на рис. 2, более молодые когорты вне зависимости от пола и наличия высшего образования чаще оценивают свое здоровье как

«Хорошее» или «Очень хорошее» по сравнению с более старшими когортами. Однако наличие данного тренда необязательно говорит о том, что с каждым следующим поколением население здоровеет. Обнаруженный тренд может быть связан с возрастом когорт, который необходимо контролировать в моделях. Во всех когортах доли мужчин и женщин с высшим образованием, оценивающих свое здоровье как хорошее или очень хорошее, выше соответствующих долей мужчин и женщин без высшего образования, что сигнализирует о наличии положительной взаимосвязи между высшим образованием и самооценкой здоровья. Однако статистическая значимость данной взаимосвязи будет проверена далее в исследовании.

Гистограммы, приведенные на рис. 3, демонстрируют, что чем моложе когорта, тем ниже доля индивидов, у которых есть хотя бы одно хроническое заболевание. При этом в рамках отдельных когорт заметно, что существенной разницы между высокообразованными индивидами и индивидами без высшего образования не наблюдается. Как предполагается в настоящем исследовании, отсутствие значимых различий может быть обусловлено разнонаправленным характером влияния высшего образования на вероятность различных хронических заболеваний, приводящим к взаимной компенсации эффектов с разными знаками в агрегированном показателе.



**Рис. 3.** Доли индивидов, имеющих хотя бы одно хроническое заболевание, по возрастным когортам и полу

На рис. 4 иллюстрируется распределение долей индивидов, имеющих ожирение, по возрастным когортам. Можно заметить, что ожирение более распространено среди старших когорт. При этом у женщин оно в целом наблюдается чаще, чем у мужчин, а доля женщин, имеющих ожирение, среди высокообразованных меньше доли женщин с ожирением среди тех, у кого нет высшего образования. По мужчинам же в основном наблюдается обратная ситуация.

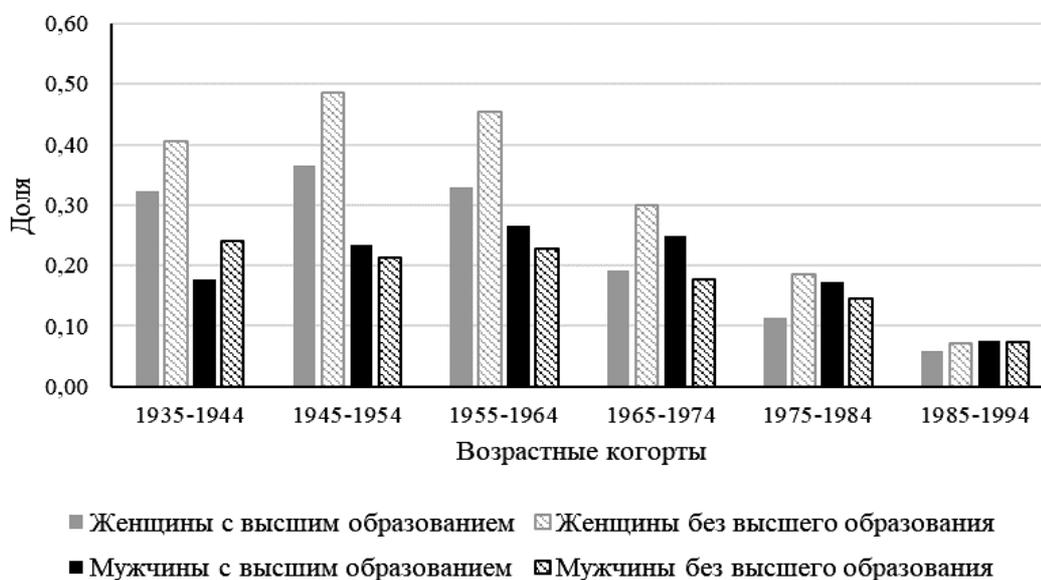


Рис. 4. Доли индивидов, у которых есть ожирение, по возрастным когортам и полу

Отдельные графики, иллюстрирующие распределение различных хронических заболеваний среди мужчин и женщин разных возрастных когорт, опускаются для краткости. Однако стоит отметить, что наиболее всего среди индивидов распространены гипертоническая болезнь, болезни сердца и ожирение. Менее всего распространена хроническая аллергия. При этом хронические заболевания, в основном, выявляются чаще у женщин, чем у мужчин. Вероятно, это может быть связано с тем, что, как правило, женщины более ответственно относятся к своему здоровью, с большей регулярностью посещают врача (см. табл. 2 и 3 далее), из-за чего заболевания и могут обнаруживаться у них с большей частотой.

В табл. 2 и 3 представлены описательные статистики используемых контрольных переменных. Согласно данным таблицам, среди женщин и мужчин с высшим образованием (по сравнению с женщинами и мужчинами без него) выше доля тех, кто окончил школу в Москве, Санкт-Петербурге или областном центре, состоит в официальном либо гражданском браке, посещает врача два раза в год или чаще, имеет работу и употребляет хотя бы иногда алкоголь; при этом ниже доля тех, кто проживает в городах (не в областных центрах) и курит. В среднем также высокообразованные женщины и мужчины обладают большим доходом, при этом у них меньше детей.

Как показывает дескриптивный анализ, между высокообразованными индивидами и индивидами, у которых нет высшего образования, существуют различия в распределении факторов, которые потенциально могут влиять и на вероятность наличия высшего образования, и на состояние здоровья, такие как, например, место проживания. Это, в свою очередь, подчеркивает важность проведения дальнейшего эконометрического анализа. Кроме того, наблюдаются очевидные различия между индивидами в разрезе возрастных когорт и пола, поэтому весь дальнейший анализ будет построен на соответствующих подвыборках.

**Таблица 2.**

**Описательные статистики контрольных переменных  
по всей выборке женщин**

Название переменной	Нет высшего образования (57844 наблюдения)		Есть высшее образование (23513 наблюдений)	
	среднее значение	стандартное отклонение	среднее значение	стандартное отклонение
age	48,069	16,7	43,932	15,244
oblс	0,374	0,484	0,55	0,497
gorod	0,282	0,45	0,259	0,438
soblс	0,263	0,44	0,423	0,494
sgorod	0,303	0,459	0,323	0,468
invalid	0,098	0,298	0,061	0,241
ln_income	8,998	0,665	9,352	0,617
num_of_child	1,514	1,019	1,278	0,845
marriage	0,59	0,492	0,659	0,474
doctor	0,519	0,5	0,563	0,496
work	0,515	0,5	0,741	0,438
alcohol	0,621	0,485	0,706	0,455
smoking	0,181	0,385	0,105	0,306
phys_active	0,201	0,401	0,345	0,476

**Таблица 3.**

**Описательные статистики контрольных переменных  
по всей выборке мужчин**

Название переменной	Нет высшего образования (45023 наблюдения)		Есть высшее образование (13358 наблюдений)	
	среднее значение	стандартное отклонение	среднее значение	стандартное отклонение
age	43,429	15,178	43,342	15,121
oblс	0,359	0,48	0,568	0,495
gorod	0,283	0,45	0,246	0,43
soblс	0,273	0,446	0,442	0,497
sgorod	0,296	0,457	0,318	0,466
invalid	0,086	0,28	0,064	0,244
ln_income	9,025	0,692	9,371	0,678
num_of_child	1,31	1,072	1,25	0,948
marriage	0,772	0,419	0,822	0,383
doctor	0,292	0,455	0,37	0,483
work	0,664	0,472	0,777	0,416
alcohol	0,763	0,425	0,795	0,404
smoking	0,593	0,491	0,362	0,48
phys_active	0,214	0,41	0,409	0,492

### 1.3. Многомерная рекурсивная пробит-модель (multivariate recursive probit model)

Для оценивания среднего эффекта воздействия высшего образования на здоровье в данном исследовании применялись как параметрический, так и непараметрический подходы, а именно многомерная рекурсивная пробит-модель [Cappellari, Jenkins, 2003; Mourifie, Meango, 2014; Коссова и др., 2023] и мэтчинг [Stuart, 2010] соответственно. Выбор методов обосновывается наличием проблемы самоотбора в данных, а также тем, что переменная воздействия и переменные исхода являются бинарными.

Ниже представлен общий вид многомерных рекурсивных пробит-моделей, оцениваемых в исследовании:

$$\left\{ \begin{array}{l} Y_i^* = \beta_0 + \beta_1 \text{diploma}_i + \beta_2 \text{alcohol}_i + \beta_3 \text{smoking}_i + \beta_4 \text{phys\_active}_i + \beta_5 \text{marriage}_i + \\ + z_i' \tilde{\beta} + \varepsilon_{1i} = x_i' \beta + \beta_1 \text{diploma}_i + \varepsilon_{1i}, \\ \text{diploma}_i^* = \alpha_0 + \alpha_1 \text{sobl}_i + \alpha_2 \text{sgorod}_i + \alpha_3 \text{invalid}_i + \alpha_4 \text{age}_i + \varepsilon_{2i} = x_i' \alpha + \varepsilon_{2i}, \\ \text{marriage}_i^* = \gamma_0 + \gamma_1 \text{mar\_share}_i + w_i' \tilde{\gamma} + \varepsilon_{3i} = x_i' \gamma + \varepsilon_{3i}, \\ \text{alcohol}_i^* = \delta_0 + \delta_1 \text{marriage}_i + \delta_2 \text{alcoh\_share}_i + w_i' \tilde{\delta} + \varepsilon_{4i} = x_i' \delta + \varepsilon_{4i}, \\ \text{smoking}_i^* = \lambda_0 + \lambda_1 \text{marriage}_i + \lambda_2 \text{smoke\_share}_i + w_i' \tilde{\lambda} + \varepsilon_{5i} = x_i' \lambda + \varepsilon_{5i}, \\ \text{phys\_active}_i^* = \tau_0 + \tau_1 \text{marriage}_i + w_i' \tilde{\tau} + \varepsilon_{6i} = x_i' \tau + \varepsilon_{6i}, \end{array} \right.$$

$$Y_i = \mathbf{1}\{Y_i^* \geq 0\}, \quad \text{diploma}_i = \mathbf{1}\{\text{diploma}_i^* \geq 0\}, \quad \text{marriage}_i = \mathbf{1}\{\text{marriage}_i^* \geq 0\},$$

$$\text{alcohol}_i = \mathbf{1}\{\text{alcohol}_i^* \geq 0\}, \quad \text{smoking}_i = \mathbf{1}\{\text{smoking}_i^* \geq 0\}, \quad \text{phys\_active}_i = \mathbf{1}\{\text{phys\_active}_i^* \geq 0\},$$

$$z_i = (\text{age}_i, \text{obl}_i, \text{gorod}_i, \text{invalid}_i, \ln\_income_i, \text{num\_of\_child}_i, \text{doctor}_i, \text{work}_i),$$

$$w_i = (\text{age}_i, \text{obl}_i, \text{gorod}_i, \text{invalid}_i), \quad i \in \{1, \dots, N\},$$

где  $Y_i$  – бинарная переменная исхода, отражающая значение конкретной целевой переменной здоровья;  $\text{diploma}_i$ ,  $\text{marriage}_i$ ,  $\text{alcohol}_i$ ,  $\text{smoking}_i$  и  $\text{phys\_active}_i$  – эндогенные бинарные переменные системы. Значения данных переменных определяются соответствующими латентными переменными  $Y_i^*$ ,  $\text{diploma}_i^*$ ,  $\text{marriage}_i^*$ ,  $\text{alcohol}_i^*$ ,  $\text{smoking}_i^*$  и  $\text{phys\_active}_i^*$ . Также  $w_i'$  – вектор-строка значений экзогенных переменных из уравнений для эндогенных переменных  $\text{marriage}_i$ ,  $\text{alcohol}_i$ ,  $\text{smoking}_i$  и  $\text{phys\_active}_i$ ;  $z_i'$  – вектор-строка значений контрольных переменных из уравнения для переменной исхода  $Y_i$ ;  $\text{sobl}_i$ ,  $\text{sgorod}_i$ ,  $\text{invalid}_i$ ,  $\text{age}_i$  – контрольные переменные из уравнения для эндогенной переменной воздействия  $\text{diploma}_i$ ;  $x_i'$  – общий вектор-строка значений всех объясняющих переменных (если в соответствующем уравнении нет какой-либо переменной, коэффициент при ней считается равным нулю);  $\beta$ ,  $\alpha$ ,  $\gamma$ ,  $\delta$ ,  $\lambda$ ,  $\tau$  – векторы-столбцы коэффициентов;  $\varepsilon_{ki}$  ( $k = 1, \dots, j$ ) –

случайные ошибки, имеющие многомерное нормальное распределение с нулевыми математическими ожиданиями и следующей ковариационной матрицей:

$$V = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1j} \\ \rho_{21} & \ddots & \ddots & \rho_{2j} \\ \vdots & \ddots & \ddots & \vdots \\ \rho_{j1} & \rho_{j2} & \cdots & 1 \end{pmatrix},$$

где  $j$  – число уравнений системы;  $\rho_{hq}$  – коэффициент корреляции между случайными ошибками  $\varepsilon_{hi}$  и  $\varepsilon_{qi}$ .

Система бинарных уравнений оценивалась методом максимального правдоподобия (ММП) отдельно для выборок мужчин и женщин, каждого из 13 показателей здоровья и представителей каждой из 6 возрастных когорт. Таким образом, всего в исследовании оценивалось  $2 \times 13 \times 6 = 156$  многомерных рекурсивных пробит-моделей. Переменные, отражающие доли населения в браке (*mar\_share*), доли населения, употребляющего алкоголь (*alcoh\_share*), и доли ежедневно курящих людей (*smoke\_share*) по регионам проживания и годам волн проведения опроса<sup>5</sup>, включались в оцениваемые многомерные пробит-модели с целью выполнения так называемого ограничения исключения (*exclusion restriction*).

По итогам оценивания системы бинарных уравнений по каждой подвыборке индивидов определенного пола и когорты рождения далее следующим образом оценивался средний эффект воздействия высшего образования на конкретную целевую переменную здоровья:

$$\widehat{ATE} = \frac{1}{N} \sum_{i=1}^N \left[ \hat{P}(Y_i(1) = 1 | x_i) - \hat{P}(Y_i(0) = 1 | x_i) \right]^6,$$

где  $Y_i(1)$  – потенциальный исход  $i$ -го индивида, если бы он получил высшее образование ( $diploma_i = 1$ );  $Y_i(0)$  – потенциальный исход  $i$ -го индивида, если бы он не получил высшее образование ( $diploma_i = 0$ );  $N$  – число наблюдений в соответствующей подвыборке индивидов определенного пола и когорты рождения.

Достоинство многомерной рекурсивной пробит-модели заключается в возможности учета ненаблюдаемых факторов, которые могут влиять и на вероятность наличия высшего образования, и на здоровье, за счет корреляций между соответствующими случайными ошибками системы. Кроме того, данная модель позволяет в явном виде учитывать эндогенность переменной воздействия (образования), а также возможную эндогенность таких переменных, как семейное положение индивида, потребление алкоголя, курение и физическая активность, с помощью включения в систему отдельных уравнений для данных переменных. При этом следует подчеркнуть, что технических отличий в оценивании рекурсивной и нерекурсивной бинарных систем нет, что принципиально отличает системы

<sup>5</sup> Использовались данные Федеральной службы государственной статистики и Комплексного наблюдения условий жизни населения.

<sup>6</sup> Подробности расчета см. в Приложении 2.

бинарных уравнений от систем с непрерывными зависимыми переменными. Как подчеркивается в [Greene, 2012], эндогенная природа некоторых переменных в правой части системы бинарных уравнений фактически может быть «проигнорирована» при нахождении параметров системы, поскольку проводится максимизация логарифмической функции правдоподобия, слагаемые в которой в рекурсивном случае отличаются от нерекурсивного только наличием разных констант в линейных индексах.

Тем не менее строгость функциональной формы и необходимость делать предположение о распределении случайных ошибок мотивируют, наравне с системами бинарных уравнений, использовать для оценивания средних эффектов воздействия непараметрические методы, например, мэтчинг.

#### 1.4. Мэтчинг (matching)

Идея процедуры мэтчинга состоит в том, чтобы для каждого объекта из одной группы найти наиболее похожий на него объект из противоположной группы на основе значений контрольных переменных. Благодаря данной процедуре исследователь достигает баланса между объектами из противоположных групп по всем переменным, за исключением переменной воздействия. Тогда исход подобранного объекта из противоположной группы можно использовать в качестве оценки ненаблюдаемого потенциального исхода объекта из рассматриваемой группы. Например, здоровье  $i$ -го индивида с высшим образованием  $Y_i(1)$  можно сравнить со здоровьем  $j$ -го индивида без высшего образования  $Y_j(0)$  и, если значения контрольных переменных (возраст, регион окончания школы и т.д.) у этих индивидов окажутся достаточно схожими, то, при определенных условиях,  $Y_j(0)$  можно рассматривать в качестве достаточно точной аппроксимации  $Y_i(0)$  и оценить эффект воздействия высшего образования на здоровье  $i$ -го индивида как  $Y_i(1) - Y_j(0)$ .

На данном этапе возникает вопрос, каким образом определять схожесть объектов. Наиболее распространенными метриками для этого выступают расстояние Махаланобиса (Mahalanobis distance) и мера склонности к воздействию (propensity score, PS). Для проверки устойчивости результатов в данном исследовании для определения близости объектов было решено использовать обе указанные метрики. Первая из них рассчитывается следующим образом:

$$r(u, l) = (u - l)' \Sigma^{-1} (u - l),$$

где  $u, l$  – векторы значений контрольных переменных двух объектов;  $\Sigma$  – ковариационная матрица контрольных переменных. Чем меньше расстояние Махаланобиса  $r$ , тем объекты считаются более схожими.

Вторая метрика, т.е. мера склонности к воздействию (propensity score), – это условная на контрольные переменные  $\tilde{X}_i$ , участвующие в процедуре мэтчинга, вероятность того, что объект будет подвергнут воздействию:  $P(\text{diploma}_i = 1 | \tilde{X}_i)$ . Схожими считаются

объекты с близкими значениями этой меры. Преимуществом данной метрики по сравнению с расстоянием Махаланобиса является тот факт, что вместо нескольких контрольных переменных для определения похожих объектов задействована всего одна величина. В работе для оценивания меры склонности используется логит-модель с тем же набором контрольных переменных, что и в уравнении для эндогенной переменной воздействия  $diploma_i$  в многомерной рекурсивной пробит-модели.

Качество баланса, достигнутого между группами по результатам проведения мэтчинга, можно проверить с помощью нескольких метрик. Основными из них являются абсолютная стандартизованная разность средних значений контрольных переменных между группой воздействия и контрольной группой (Absolute Standardized Mean Difference, ASMD), отношение дисперсий контрольных переменных (Variance Ratio, VR), среднее и максимальное значения расстояния между эмпирическими функциями распределения контрольных переменных (eCDF Mean и eCDF Max соответственно).

ASMD для непрерывной контрольной переменной  $X_j$  рассчитывается следующим образом:

$$ASMD_{X_j} = \left| \frac{\bar{X}_{j,T} - \bar{X}_{j,C}}{\sqrt{\frac{s_{X_j,T}^2 + s_{X_j,C}^2}{2}}} \right|,$$

где  $\bar{X}_{j,T}$  и  $\bar{X}_{j,C}$  – выборочное среднее значение переменной  $X_j$  в группе воздействия и контрольной группе соответственно;  $s_{X_j,T}^2$  и  $s_{X_j,C}^2$  – выборочные дисперсии переменной  $X_j$ .

ASMD для бинарной контрольной переменной  $X_j$  рассчитывается по формуле:

$$ASMD_{X_j} = \left| \frac{\hat{p}_T - \hat{p}_C}{\sqrt{\frac{\hat{p}_T(1 - \hat{p}_T) + \hat{p}_C(1 - \hat{p}_C)}{2}}} \right|,$$

где  $\hat{p}_T$  и  $\hat{p}_C$  – выборочное среднее значение переменной  $X_j$  в группе воздействия и контрольной группе соответственно.

Внимание уделяется тому, насколько указанные метрики изменились после проведения мэтчинга. В частности, близость ASMD, eCDF Mean и eCDF Max к нулю, а также VR к единице после процедуры мэтчинга будут говорить о достижении хорошего баланса между распределениями контрольных переменных в группе воздействия и контрольной группе и успешности проведенной процедуры.

Для примера, в Приложении 3 приводятся таблицы со значениями указанных метрик до и после проведения мэтчинга на основе расстояния Махаланобиса и PS-мэтчинга для когорты женщин 1945–1954 гг. Можно заметить, что после обеих процедур качество баланса между группой воздействия и контрольной группой заметно улучшилось, поскольку значения всех перечисленных метрик уменьшились.

По итогам проведения каждой из процедур мэтчинга по данным изначальных подвыборок индивидов разного пола и когорт рождения были получены новые, сбалансированные по контрольным характеристикам подвыборки. После этого с целью дополнительного контроля оценивание среднего эффекта воздействия по каждой сбалансированной после мэтчинга подвыборке производилось с использованием оцененных с помощью логит-модели условных вероятностей единичных потенциальных исходов в соответствии с формулой

$$\widehat{ATE} = \frac{1}{N} \sum_{i=1}^N [\hat{P}(Y_i(1) = 1 | x_i) - \hat{P}(Y_i(0) = 1 | x_i)],$$

где  $N$  – число наблюдений в соответствующей подвыборке индивидов определенного пола и когорты рождения.

Достоинство мэтчинга состоит в ослаблении предположения о распределении случайных ошибок и функциональной форме связи между переменными. Однако при его использовании возникает необходимость в соблюдении возможно более сильных допущений, а именно:

1) пересечение (overlap): не существует таких значений контрольных переменных, при которых можно точно сказать, в какую группу попадет объект, т.е.  $0 < P(\text{diploma} = 1 | \tilde{X}) < 1$ ;

2) стабильность эффекта воздействия на объект (SUTVA): на потенциальные исходы одного объекта не должно влиять воздействие на другие объекты и при этом на все объекты из группы воздействия должно быть оказано одинаковое воздействие;

3) несмешиваемость (unconfoundedness):  $(Y(0), Y(1)) \perp \text{diploma} | \tilde{X}$ , т.е. условно по контрольным переменным отбор объекта в группу не должен зависеть от его потенциальных исходов. Данное предположение фактически подразумевает, что (условно на контрольные переменные  $\tilde{X}$ ) не существует ненаблюдаемых характеристик, влияющих на значение переменной воздействия для объекта и его потенциальные исходы. Являясь близким по смыслу к предположению об экзогенности в стандартных регрессионных моделях, данное ограничение становится наиболее важным предположением мэтчинга и его ключевым недостатком в сравнении с многомерной пробит-моделью. Связано это с тем, что проверить его выполнимость на реальных данных не представляется возможным в силу отсутствия для этого формальных тестов, тогда как на практике оно довольно часто может нарушаться. Например, данное предположение может не выполняться при оценивании среднего эффекта воздействия высшего образования на вероятность наличия ожирения у индивида при неучете в процедуре мэтчинга переменной, отражающей уровень стресса, который индивид испытывал в процессе получения образования в школе. Данный фактор не поддается точному измерению и выражению в виде конкретной пе-

ременной, которая могла бы быть включена в процедуру мэтчинга. При этом уровень испытываемого индивидом стресса во время обучения в школе может значимым образом влиять как на наличие высшего образования у индивида (и, соответственно, попасть в случайную ошибку уравнения для переменной воздействия), так и на наличие у него ожирения (и попасть в случайную ошибку уравнения для переменной исхода, т.е. переменной наличия ожирения), приводя, тем самым, к нарушению предположения о несмещаемости.

## **2. Результаты оценивания эффекта воздействия высшего образования на здоровье и их обсуждение**

Как уже отмечалось ранее, оценивание средних эффектов воздействия проводилось отдельно для каждой когорты мужчин и женщин. Соответствующая гипотеза о наличии структурной неоднородности связи между образованием и переменными здоровья по полу и когорте рождения индивидов проверялась с помощью теста отношения правдоподобия (LR-тест). По результатам проведения данного теста были получены статистические свидетельства в пользу того, что для любой пары соседних когорт мужчин и женщин гипотеза о существовании единой зависимости отвергается на любом разумном уровне значимости. То есть между когортами наблюдаются статистически значимые различия, вследствие чего они не могут быть объединены и оценивание средних эффектов воздействия, действительно, необходимо проводить для каждой когорты мужчин и женщин в отдельности.

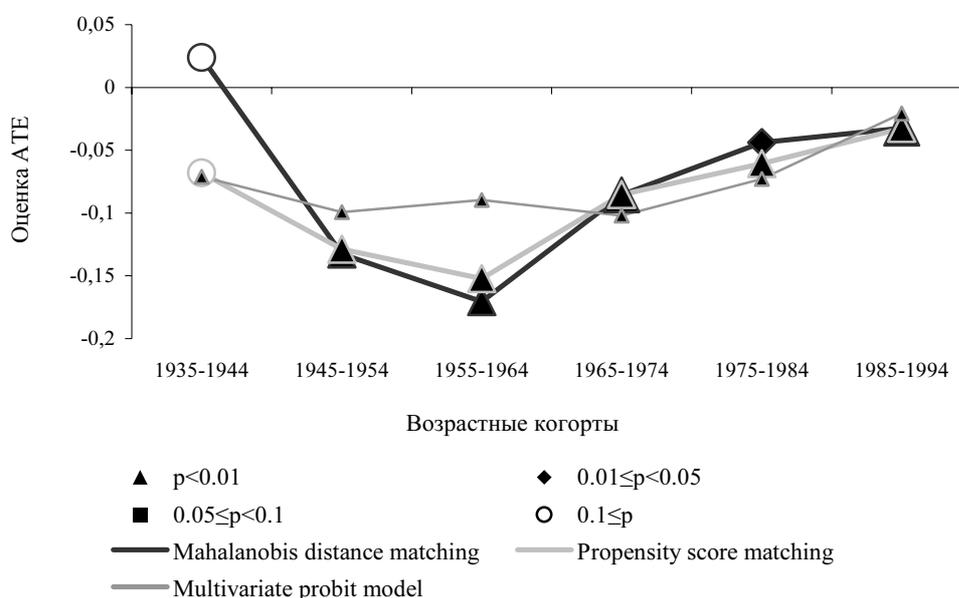
В силу невозможности в доступной для читателя форме обсудить результаты оценивания всех моделей для краткости в данном разделе будут представлены только наиболее интересные с точки зрения значимости и интерпретации результаты. Для наглядности они будут продемонстрированы на графиках динамики оценок АТЕ по возрастным когортам, где по вертикальной оси представлена оценка среднего эффекта воздействия высшего образования на определенный показатель здоровья, а по горизонтальной оси – период рождения индивидов (обозначение когорты). Для сравнения и определения степени устойчивости результатов к методу оценивания на графиках будут одновременно изображены результаты, полученные с помощью многомерной рекурсивной пробит-модели, мэтчинга на основе расстояния Махаланобиса и мэтчинга на основе меры склонности.

### **2.1. Результаты по выборкам женщин**

На рис. 5 представлены результаты оценивания среднего эффекта воздействия высшего образования на вероятность наличия ожирения у женщин разных возрастных когорт. Как видно из графика, наблюдаются устойчивые значимые отрицательные средние эффекты воздействия. Иными словами, согласно полученным оценкам, высшее образование для всех когорт женщин снижает вероятность иметь ожирение. Результат согласуется с работами предыдущих исследователей [Brunello et al., 2013; Grabner, 2009; Pisa et al., 2012] и может быть обусловлен тем, что люди с высшим образованием, в силу аккумулированных по результатам его получения знаний и навыков, могут формировать более здоровые привычки в питании и жизни в целом, что позволяет им с большей вероятностью поддер-

живать свой вес на нормальном уровне. При этом стоит обратить внимание, что, за исключением когорт преклонного возраста, чем моложе когорта, тем ниже по абсолютной величине средний эффект воздействия, что свидетельствует в пользу гипотезы о снижении отдачи от образования. Объясняется данный результат может тем, что сегодня повсеместно идет активное продвижение здорового образа жизни (ЗОЖ), и молодые люди в целом, независимо от их уровня образования, становятся более склонны к ведению ЗОЖ [Засимова и др., 2017], из-за чего различия в показателе наличия ожирения могут быть в меньшей степени связаны с образованием.

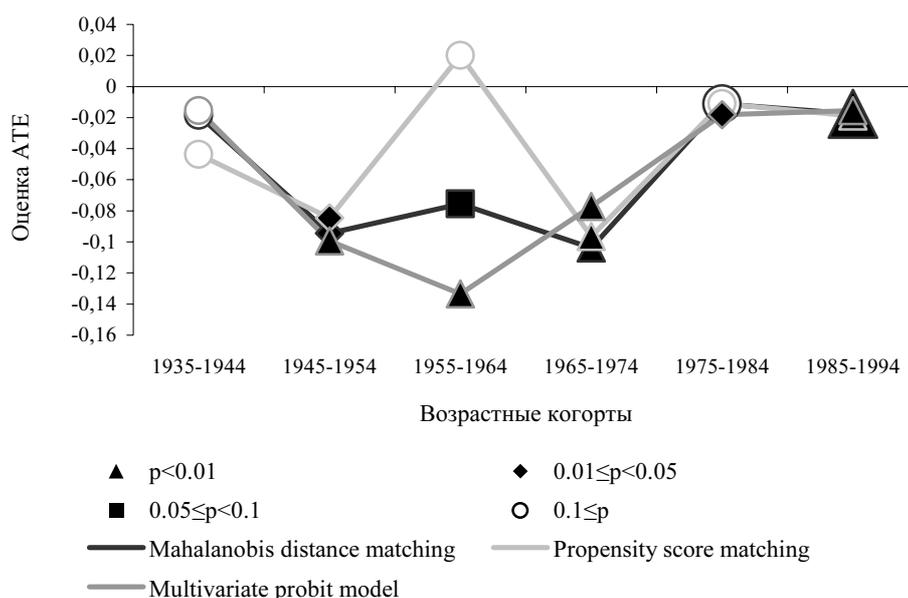
Здесь однако стоит отметить, что у мужчин в отношении ожирения подобной женщинам динамики эффекта не наблюдается (см. Приложение 1, рис. П1) – среди мужчин для большинства когорт средний эффект воздействия оказался незначимым, что, тем не менее, является также важным результатом, поскольку подтверждает необходимость проводить оценивание эффекта для мужчин и женщин в отдельности.



**Рис. 5.** Результаты оценивания средних эффектов воздействия высшего образования на вероятность наличия ожирения в динамике по возрастным когортам по выборке женщин

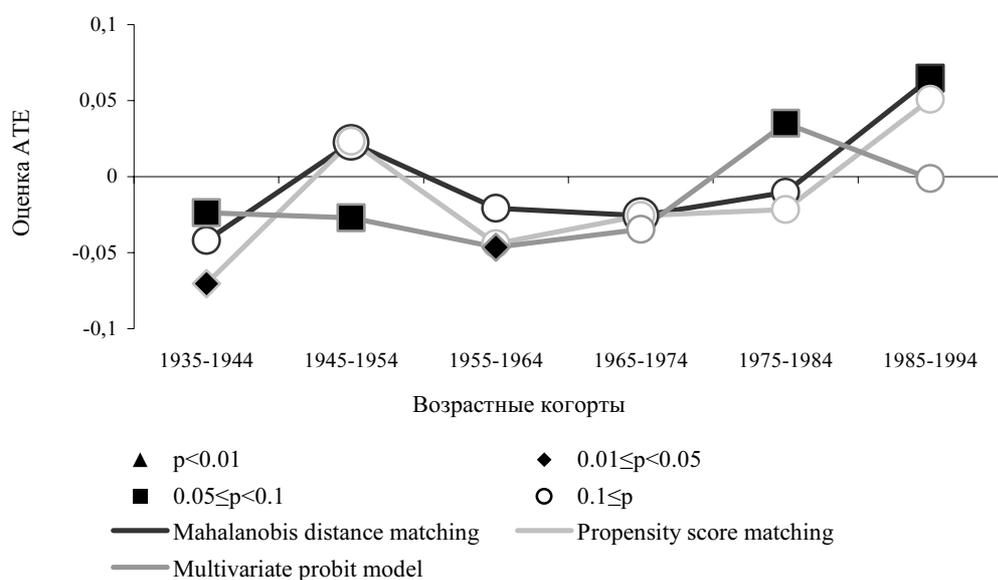
На рис. 6 представлена динамика оценок средних эффектов воздействия высшего образования на вероятность наличия гипертонической болезни у женщин. Почти для всех когорт наблюдаются значимые отрицательные средние эффекты воздействия, т.е. вероятность гипертонии при наличии высшего образования у женщин практически во всех когортах снижается, что также согласуется с результатами ряда предшествовавших исследований [Коссова, Косорукова, 2023; Fonseca, Zheng, 2011; Pisa et al., 2012; Viego, Temporelli, 2017]. При этом, как и на рис. 5, в динамике обнаруживается снижение оценок эффектов воздействия по модулю, начиная с когорты 1955–1964 гг., что вновь может свидетельст-

вывать в пользу справедливости гипотезы о снижении отдачи от образования. Стоит обратить внимание на то, что для самой молодой когорты наблюдаются уже весьма низкие по модулю оценки, тогда как для самой старшей когорты эффекты в принципе незначимы. Это может быть связано с тем, что хронические заболевания в целом и гипертония в частности сильно распространены среди пожилых людей, поэтому различия между высокообразованными женщинами и женщинами без высшего образования в крайней старшей когорте могут нивелироваться. В то же время среди представителей самой молодой когорты, наоборот, такие заболевания встречаются гораздо реже, из-за чего образование может пока не оказывать серьезного влияния на их вероятность.

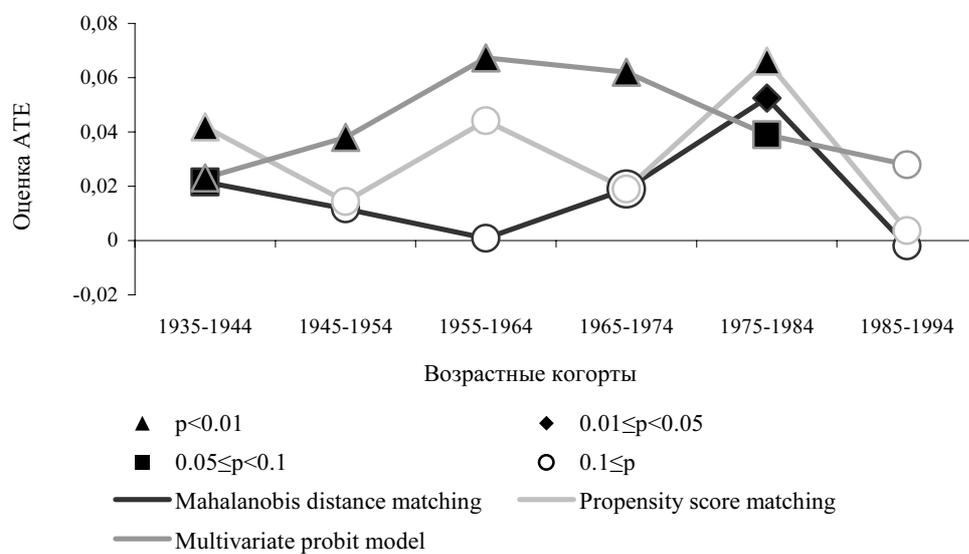


**Рис. 6.** Результаты оценивания средних эффектов воздействия высшего образования на вероятность наличия гипертонии в динамике по возрастным когортам по выборке женщин

Ранее в работе выдвигалось предположение о том, что использование агрегированных показателей здоровья при оценивании эффекта образования может быть не самым лучшим решением в силу возможного взаимного компенсирования эффектов. На рис. 7 видно, что это предположение подтверждается для женщин. Здесь показана динамика среднего эффекта воздействия высшего образования на вероятность наличия у женщины хотя бы одного хронического заболевания. Из рисунка можно заметить, что в основном средний эффект воздействия на данный агрегированный показатель здоровья для любой когорты оказывается незначимым. Вероятно, происходит компенсация эффектов на разные отдельные заболевания с противоположными знаками друг с другом, что в итоге приводит к получению незначимого результата. Отметим, что для мужчин в отношении наличия хотя бы одного любого хронического заболевания справедливы аналогичные выводы (см. Приложение 1, рис. П2).



**Рис. 7.** Результаты оценивания средних эффектов воздействия высшего образования на вероятность наличия хотя бы одного хронического заболевания в динамике по возрастным когортам по выборке женщин



**Рис. 8.** Результаты оценивания средних эффектов воздействия высшего образования на вероятность самооценки здоровья как «Хорошее» или «Очень хорошее» в динамике по возрастным когортам по выборке женщин

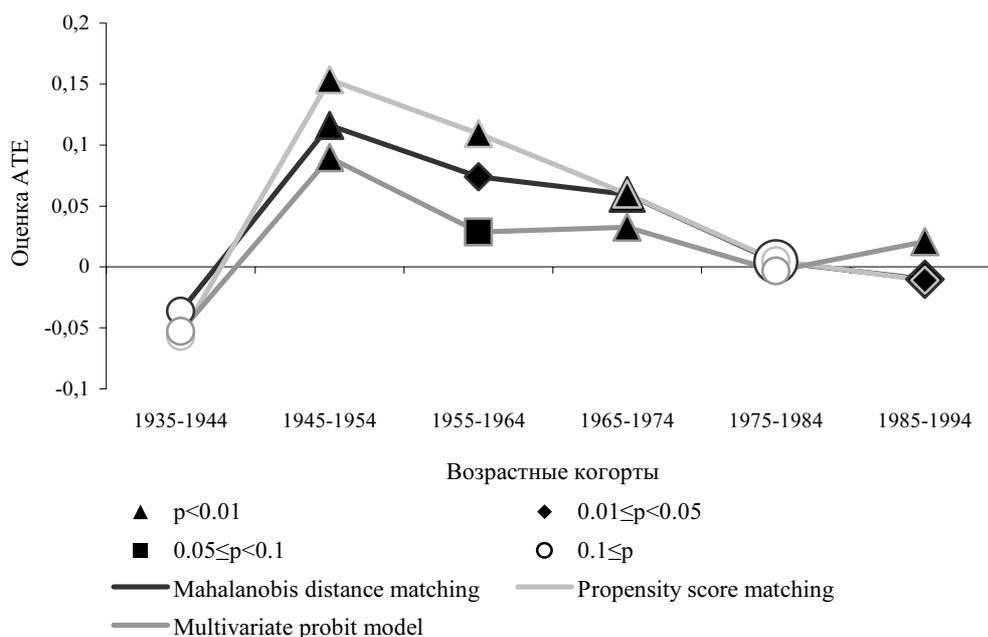
Заметим также, что, согласно рис. 8, наблюдаются положительные оценки среднего эффекта воздействия высшего образования на субъективную оценку здоровья. Иными словами, высшее образование повышает вероятность высокой самооценки здоровья среди женщин, что согласуется с результатами работ [Brunello et al., 2016; Hu, 2014; Silles, 2009]. Однако значимым практически во всех когортах эффект оказался только при оценивании его многомерной рекурсивной пробит-моделью. При этом отсутствие заметной динамики не дает статистических оснований утверждать, что эффект воздействия уменьшается или увеличивается при движении от старших когорт к младшим, что не позволяет сделать вывод о справедливости какой-либо из когортных гипотез. Так, например, согласно многомерной пробит-модели, наблюдается приблизительно гиперболическая динамика оценок с малыми по величине эффектами для крайних когорт. Этот результат может быть вызван тем, что представители самой старшей когорты, как правило, становятся более восприимчивы к различным заболеваниям, из-за чего эффект наличия высшего образования на вероятность высокой самооценки здоровья у них может быть не столь выражен по сравнению с соседними когортами. При этом, обращая внимание на снижение эффекта, начинающегося с когорты 1965–1974 гг., стоит отметить, что возраст многих представителей этой когорты в выборке составляет примерно 45–55 лет. Данный возраст можно обозначить как переходный в том смысле, что он характеризуется повышением рисков развития и/или обострения различных серьезных заболеваний. По этой причине индивиды в этом возрасте могут быть менее склонны оценивать свое здоровье на высоком уровне независимо от образования. Для самой же молодой когорты значимые эффекты не выявляются, что вновь может являться результатом распространения среди них здорового образа жизни, повышения общего уровня здоровья населения в целом и, следовательно, нивелирования различий между образованными и необразованными женщинами в данной когорте.

## 2.2. Результаты по выборкам мужчин

Результаты для мужчин оказались менее однозначными, чем для женщин. Согласно результатам, представленным на рис. 9, среди мужчин почти во всех когортах высшее образование значимо повышает вероятность наличия хронических заболеваний сердца. При этом для некоторых когорт наблюдается аналогичный результат и в отношении вероятности наличия гипертонии (рис. 10). С одной стороны, данный результат противоречит некоторым зарубежным исследованиям [Fonseca, Zheng, 2011; Pisa et al., 2012; Viego, Temporelli, 2017]. Однако, с другой стороны, важно обратить внимание на то, что среди представленных исследований отдельно для мужчин и женщин анализ проводился лишь в работе [Pisa et al., 2012] на данных по ЮАР. При этом было получено, что у высокообразованных женщин наблюдается более низкое кровяное давление, что согласуется с результатами настоящего исследования (рис. 6), тогда как у мужчин значимой разницы в показателях давления в зависимости от уровня образования обнаружено не было. Таким образом, полученный в настоящем исследовании результат подчеркивает гендерные и страновые различия в воздействии образования на вероятность гипертонии и здоровье сердечно-сосудистой системы.

Кроме того, следует отметить, что данный результат может являться следствием образа жизни, который ведут высокообразованные мужчины в России. Как было видно

из табл. 3, больше половины мужчин с высшим образованием проживают в крупнейших городах России и областных центрах. Существуют исследования [Табакаев, Артамонова, 2015; Sodinou et al., 2008], которые показывают, что, в отличие от ситуации в развитых странах, проживание на урбанизированных территориях в развивающихся странах сопровождается более высокими рисками развития артериальной гипертензии и сердечно-сосудистых заболеваний как следствий малоподвижного образа жизни, плохого питания, постоянного стресса и психоэмоционального напряжения.

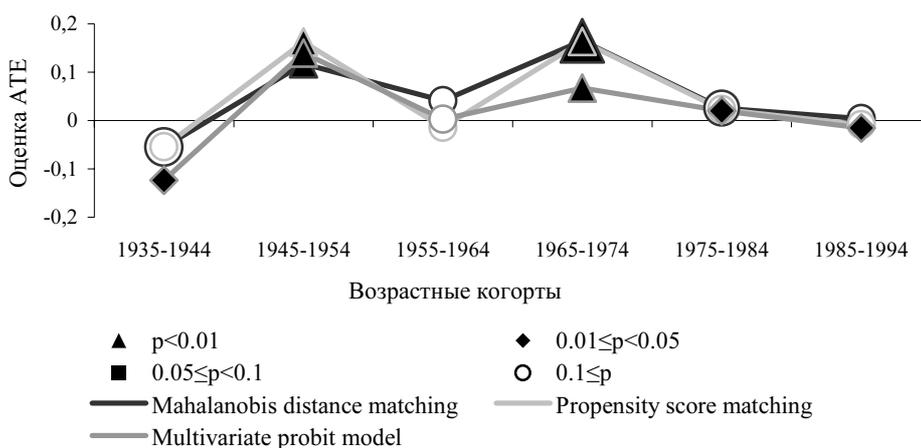


**Рис. 9.** Результаты оценивания средних эффектов воздействия высшего образования на вероятность наличия хронических заболеваний сердца в динамике по возрастным когортам по выборке мужчин

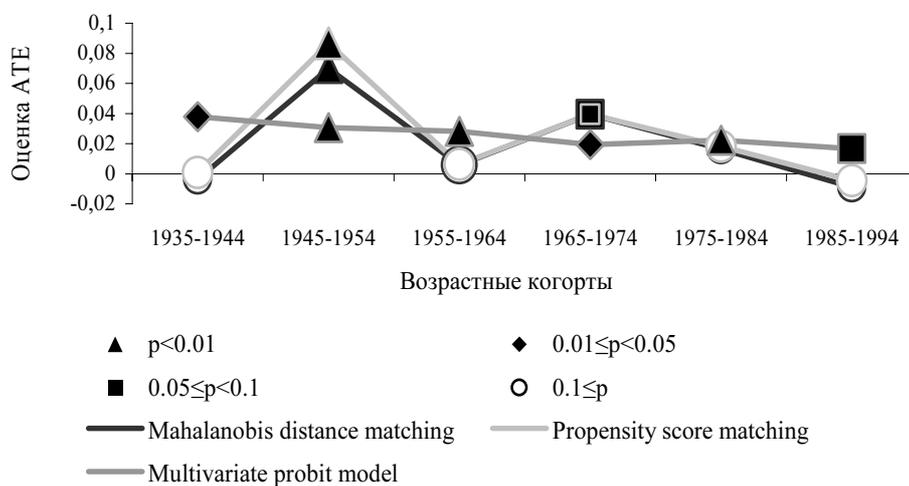
При этом на рис. 9 и 10 вновь наблюдается заметная когортная дифференциация полученных результатов оценивания, подчеркивающая ценность проведения исследования в разрезе возрастных когорт. В частности, для самой старшей когорты мужчин значимый средний эффект воздействия высшего образования на вероятность наличия заболеваний сердца не обнаруживается, тогда как для более молодых индивидов эффект становится уже значимым и различимым по когортам. Так, начиная с когорты мужчин, рожденных в 1945–1954 гг., наблюдается практически монотонное снижение оценок среднего эффекта воздействия на вероятность наличия заболеваний сердца, что в некоторой степени вновь подтверждает гипотезу о снижении отдачи от образования.

Помимо этого, у мужчин наблюдаются положительные средние эффекты воздействия высшего образования на вероятность наличия хронической аллергии практически во всех когортах (рис. 11). Как утверждается в работе [Dalstra et al., 2005], данный резуль-

тат возможно связан с длительным нахождением людей с высшим образованием в закрытых малопрветриваемых помещениях в силу специфики их работы, что может приводить к развитию хронической аллергии. При этом вновь наблюдается некоторое снижение оценок при движении от старших когорт к младшим, что свидетельствует в пользу справедливости гипотезы о когортном снижении отдачи от образования. Полученный в отношении хронической аллергии результат, кроме того, является дополнительным подтверждением необходимости проведения раздельного оценивания эффектов воздействия для мужчин и женщин, поскольку у последних не было обнаружено устойчивых эффектов воздействия на вероятность аллергии (см. Приложение 1, рис. ПЗ).

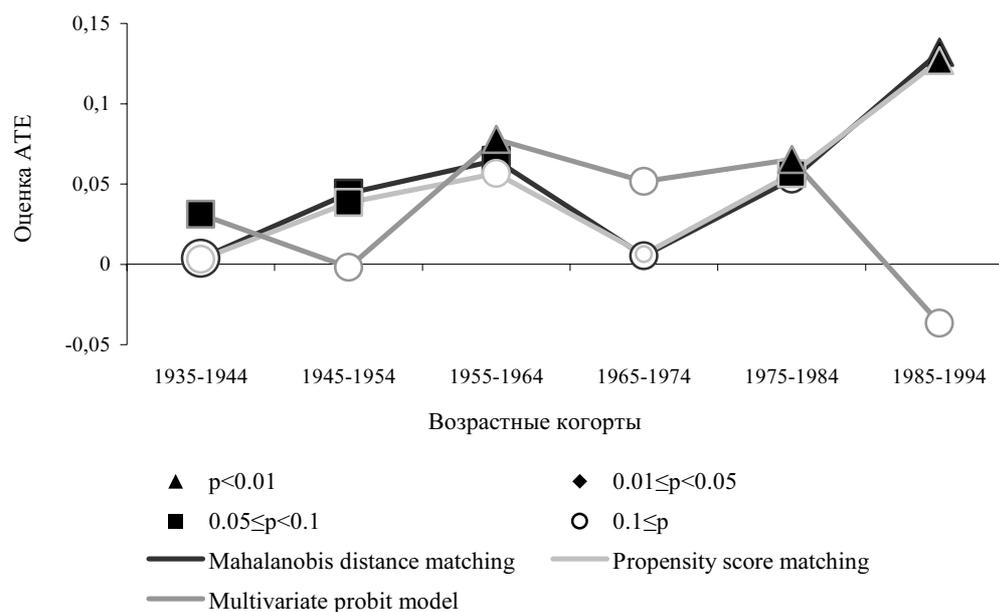


**Рис. 10.** Результаты оценивания средних эффектов воздействия высшего образования на вероятность наличия гипертонии в динамике по возрастным когортам по выборке мужчин



**Рис. 11.** Результаты оценивания средних эффектов воздействия высшего образования на вероятность наличия хронической аллергии в динамике по возрастным когортам по выборке мужчин

Наконец, рассмотрим средний эффект воздействия высшего образования на самооценку здоровья мужчин разных когорт (рис. 12). Как и для женщин, для мужчин обнаруживается неустойчивость значимости эффектов на данный показатель. Однако, если обратить внимание на значимые результаты, то почти для каждой когорты мужчин наблюдаются положительные оценки средних эффектов воздействия высшего образования на вероятность высокой самооценки здоровья, причем для более молодых когорт они выше. Подобный результат, вероятно, также отчасти связан с более бережным отношением высокообразованных молодых мужчин к своему здоровью, приводящим, как следствие, к оцениванию его на более высоком уровне.



**Рис. 12.** Результаты оценивания средних эффектов воздействия высшего образования на вероятность самооценки здоровья как «Хорошее» или «Очень хорошее» в динамике по возрастным когортам по выборке мужчин

### 3. Ограничения исследования

Настоящее исследование имеет определенные ограничения, которые важно учитывать при обсуждении его результатов. Прежде всего, следует отметить, что полученные в исследовании оценки средних эффектов воздействия связаны как со спецификой образования, так и со средним возрастом когорты, несмотря на то, что переменная «возраст» учитывалась при моделировании. В работе анализируются данные с 2009 по 2019 гг., и для индивидов из каждой когорты есть возможность рассмотреть только определенные 11 лет их жизни (например, от 45 до 55 лет для индивидов, рожденных в 1964 г.). Такая ограниченность данных не позволяет утверждать, что влияние полученного когортой образования является одинаковым на протяжении всей ее жизни, и требует делать по-

правку на средний возраст когорты при обсуждении эффектов воздействия образования на здоровье ее представителей. Так, отсутствие значимых средних эффектов воздействия образования для молодых и очень пожилых людей в настоящем исследовании может объясняться, что уже отмечалось ранее, естественными возрастными причинами. Однако при сравнении соседних когорт можно утверждать, что разница в эффектах воздействия связана именно с особенностями высшего образования, поскольку в таких когортах есть люди одного возраста.

Существуют и некоторые другие ограничения, связанные с особенностями базы данных РМЭЗ, на основе которой были сформированы исследуемые в работе выборки. В первую очередь, следует подчеркнуть, что ответы на вопросы о наличии того или иного заболевания индивиды, участвующие в опросах РМЭЗ, предоставляют самостоятельно на основе своих личных знаний о состоянии своего здоровья. Подобный механизм формирования данных по заболеваниям в базе приводит к тому, что в построенных для анализа в настоящем исследовании выборках могут находиться индивиды, ошибочно отвечающие «нет» на вопросы о наличии у них того или иного заболевания в силу отсутствия у них об этом информации. Таким образом, наличие заболевания у индивида предполагает знание о нем, что, в свою очередь, подразумевает, что индивид обращался в медицинские учреждения, проходил обследования и получил соответствующий диагноз. Из дескриптивных статистик, представленных в табл. 2 и 3, можно заметить, что среди индивидов с высшим образованием выше доля тех, кто посещает врача чаще, чем один раз в год. Это свидетельствует о том, что более высокая вероятность наличия определенных хронических заболеваний у высокообразованных индивидов и, в частности, заболеваний сердца, гипертонии и аллергии у мужчин с высшим образованием, может говорить о некоторой переоценке результатов в группе образованных индивидов и являться следствием большей осведомленности этих людей о своих проблемах со здоровьем по сравнению с индивидами без высшего образования, а не сигнализировать о лучшем здоровье последних.

Кроме того, важно отметить, что в опросниках РМЭЗ у каждого работающего индивида есть возможность выбрать одну из более чем 30 отраслей экономики, наиболее подходящую под его вид деятельности. Однако в настоящем исследовании было решено не включать в модели дамми-переменные на отрасли. В первую очередь, данное решение было мотивировано тем, что добавление такого количества дамми-переменных в многомерные рекурсивные пробит-модели значительно повысило бы вычислительную сложность и время оценивания моделей. В то же время какое-либо объединение отраслей в несколько групп не представлялось возможным в силу высокой степени разнородности отраслей. Стоит также заметить, что учет вида деятельности работающих индивидов привел бы к серьезному снижению размеров исследуемых выборок, поскольку далеко не все индивиды указывали в опросах отрасль, в которой они работают. В связи с этим необходимо иметь в виду, что некоторые из полученных в исследовании результатов оценивания средних эффектов воздействия высшего образования на здоровье могут быть опосредованы видом деятельности индивида. В частности, как отмечалось ранее в разделе 2.2, положительные эффекты воздействия образования на вероятность аллергии среди мужчин могут объясняться спецификой их работы, которую, в свою очередь, во многом мог определить их уровень образования.

Наконец, еще одна особенность базы РМЭЗ, которую необходимо упомянуть в контексте настоящего исследования, – отсутствие данных, связанных с наследственностью

индивидов, причем как с точки зрения здоровья, так и с точки зрения образования. Известно, что наследственность играет важную роль в развитии многих хронических заболеваний, таких как, например, сердечно-сосудистые заболевания, заболевания эндокринной системы, гипертония. В связи с этим пропуск в моделях контрольных переменных, отражающих наличие хронических заболеваний хотя бы у родителей индивида, может приводить к появлению смещения в оценках эффектов воздействия образования на здоровье. К аналогичным последствиям может привести и неучет переменных, отражающих уровень образования родителей индивида, поскольку также известно, что между образованием детей и образованием родителей существует значимая связь [Chevalier, 2004; Havari, Savegnago, 2014]. Иными словами, высшее образование может «наследоваться» параллельно с предрасположенностью к различным заболеваниям. В настоящем исследовании показатели здоровья родителей индивидов не учитывались в моделях в связи с отсутствием соответствующих данных в базе РМЭЗ. В то же время в базе имеются данные по образованию родителей индивидов, однако только в двух волнах опросов. Поскольку распространение данных за две волны на года, рассматриваемые в исследовании, приводило к значительному сокращению объема выборок, от использования данных по образованию родителей было решено отказаться.

Отдельное внимание стоит акцентировать и на ограничениях, связанных с методологией исследования, а именно на соблюдении предположения о несмешиваемости в методах мэтчинга. Поскольку на данный момент указанное предположение является нетестируемым, его выполнимость в нерандомизированных исследованиях может быть поставлена под сомнение, и настоящее исследование в данном случае не является исключением. В связи с этим один из главных методологических выводов данной работы, как будет указано далее, состоит в необходимости проведения оценивания эффектов воздействия с использованием нескольких различных методов одновременно.

### Заключение

В настоящем исследовании изучалась взаимосвязь образования со здоровьем российского населения. Несмотря на многочисленные свидетельства существования корреляционной связи между данными факторами, установление ее причинно-следственного характера является весьма непростой задачей из-за отсутствия рандомизации при выборе индивидами своей образовательной траектории. В работе учитывается самоотбор индивидов, приводящий к эндогенности переменной образования по отношению к здоровью, с помощью двух различных подходов – параметрического (многомерная рекурсивная пробит-модель) и непараметрического (мэтчинг). Новизна исследования заключается в оценивании влияния образования на здоровье отдельно для каждой когорты женщин и мужчин с целью проверки гипотезы о наличии структурной неоднородности связи между образованием и переменными здоровья по полу и когорте рождения индивидов и гипотезы о когортном снижении отдачи от образования на здоровье на российских данных.

По результатам проверки существования структурных различий в зависимостях между исследуемыми показателями по полу и когорте рождения индивидов гипотеза об однородности связи была отвергнута на любом разумном уровне значимости для каждой пары соседних когорт индивидов разного пола. Это свидетельствует о том, что поведение

мужчин и женщин в отношении своего здоровья значимо различается по когортам, что соответствующим образом отражается на уровне их здоровья и его самооценке. С методологической точки зрения данный результат подчеркивает, что оценивание средних эффектов воздействия образования на здоровье индивидов необходимо проводить по подвыборкам.

По некоторым показателям было выявлено, что, как и для европейских стран [Delaruelle et al., 2015] и США [Lynch, 2003], для России справедлива гипотеза о когортном снижении отдачи от образования на здоровье. Данный результат, однако, не указывает на отсутствие необходимости в получении высшего образования в России сегодня. Напротив, высокие по модулю эффекты воздействия для старших когорт и низкие для молодых могут свидетельствовать о кумулятивном эффекте высшего образования, проявляющемся в виде более осознанного поведения высокообразованных индивидов по отношению к своему здоровью на протяжении всей жизни и приводящем, как следствие, к более выраженному снижению вероятности заболеваний в позднем возрасте при наличии образования по сравнению с молодыми когортами. Полученный результат подчеркивает важность получения образования молодыми людьми и поддержания образования в России как сферы с определенными положительными экстерналиями на здоровье.

Для большинства когорт женщин были найдены статистические свидетельства в пользу того, что высшее образование значимо снижает вероятность возникновения ожирения и гипертонии. В то же время оно повышает вероятность наличия высокой самооценки здоровья. При этом для мужчин было обнаружено положительное воздействие высшего образования на вероятность наличия хронических заболеваний сердца, гипертонии, аллергии, а также, аналогично женщинам, на вероятность наличия высокой самооценки здоровья. Подчеркнем, однако, что увеличение вероятности указанных заболеваний для мужчин также не является результатом, свидетельствующим о необходимости прекращения получения мужчинами высшего образования, поскольку подобное поведение может повысить риски приобретения других заболеваний и тем самым привести к ухудшению здоровья. Данные результаты призваны обратить внимание российских мужчин с высшим образованием на ритм их жизни, питание и испытываемый ими уровень стресса, а также подчеркивают необходимость в дополнительной популяризации ЗОЖ среди людей, получающих высшее образование, например, путем организации специальных лекций о здоровье и правильном питании, создании дополнительных мест для занятий спортом в университетах.

С методологической точки зрения отсутствие причинно-следственной связи между некоторыми переменными здоровья и образованием указывает на важность использования методов оценивания, которые способны учитывать проблему самоотбора, поскольку иные методы могут ошибочно приводить к получению значимых результатов. Кроме того, разнообразие результатов работы как с точки зрения знаков эффектов, так и их значимости подчеркивает непригодность использования агрегированных показателей здоровья при анализе влияния на него образования из-за возможной взаимной компенсации разнонаправленных эффектов на разные показатели здоровья.

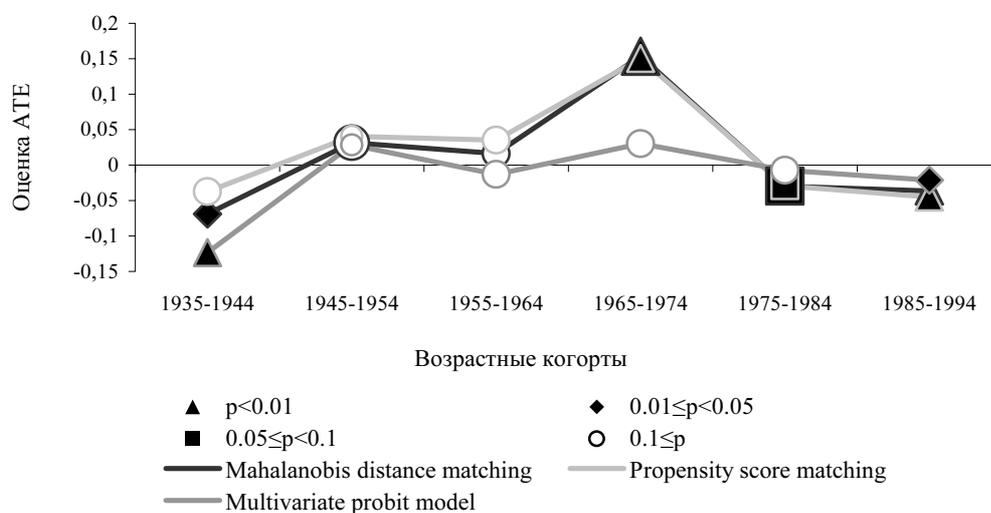
Наконец, говоря об используемых в данном исследовании методах оценивания, необходимо отметить, что, в отличие от мэтчинга, многомерная рекурсивная пробит-модель способна учесть возможное влияние ненаблюдаемых характеристик объектов на отбор в группы, в том числе влияющих одновременно на переменную воздействия и переменную исхода. В свою очередь плюсом мэтчинга по сравнению с многомерной пробит-

моделью является отсутствие необходимости в параметризации и предположении распределения случайных ошибок. Таким образом, отдавать предпочтение определенному методу оценивания не стоит, поскольку оба метода имеют как преимущества, так и недостатки, и для получения более надежных и устойчивых результатов необходимо проводить оценивание различными способами совместно.

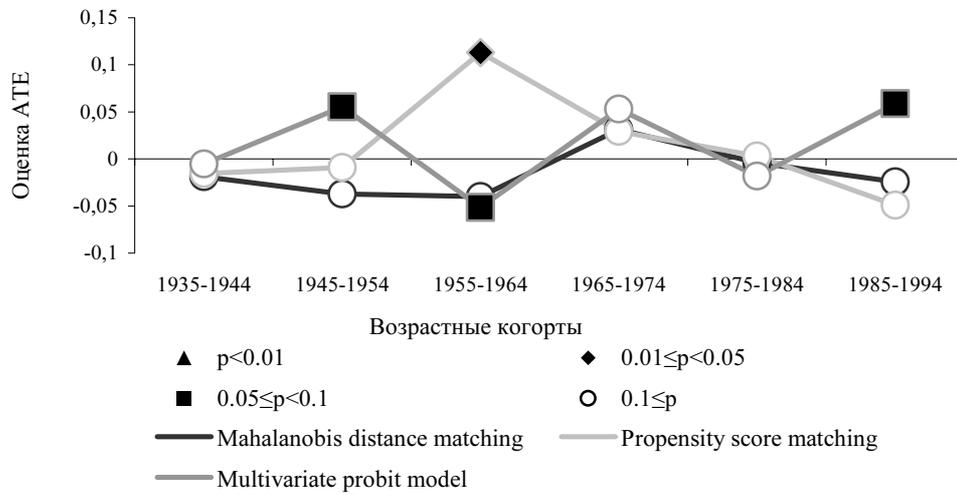
Принимая во внимание ограничения настоящей работы, будущие исследования могут быть посвящены проверке устойчивости ее результатов при использовании данных за больший промежуток времени, а также разработке подходов к проверке выполнимости предположения о несмешиваемости на реальных данных.

## Приложение 1.

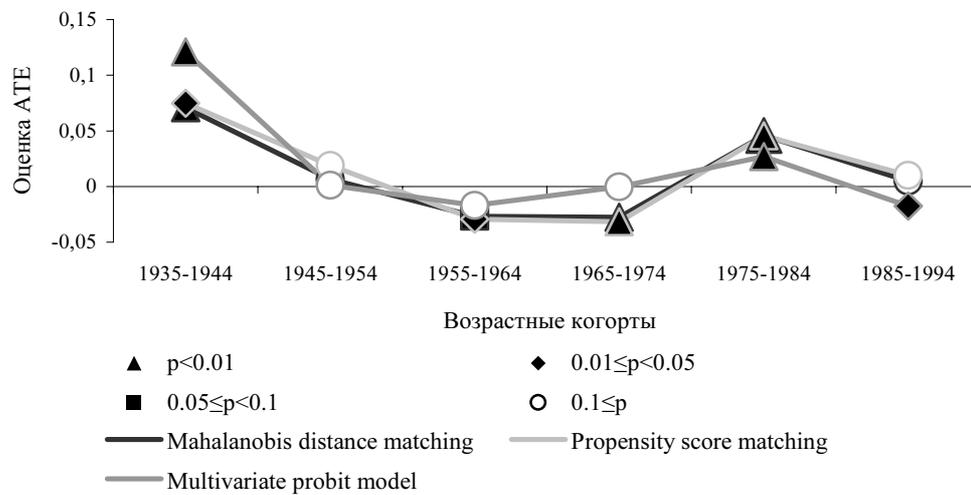
### Оценки средних эффектов воздействия высшего образования на вероятность наличия некоторых заболеваний в динамике по возрастным когортам, полученные по выборкам мужчин и женщин



**Рис. П1.** Результаты оценивания средних эффектов воздействия высшего образования на вероятность наличия ожирения в динамике по возрастным когортам по выборке мужчин



**Рис. П2.** Результаты оценивания средних эффектов воздействия высшего образования на вероятность наличия хотя бы одного хронического заболевания в динамике по возрастным когортам по выборке мужчин



**Рис. П3.** Результаты оценивания средних эффектов воздействия высшего образования на вероятность наличия хронической аллергии в динамике по возрастным когортам по выборке женщин

## Приложение 2.

### Оценивание АТЕ с помощью многомерной рекурсивной пробит-модели

Рассмотрим следующую 6-мерную рекурсивную пробит-модель:

$$\begin{cases} Y_i^* = x_i'\beta + \beta_1 \text{diploma}_i + \varepsilon_{1i}, \\ \text{diploma}_i^* = x_i'\alpha + \varepsilon_{2i}, \\ \text{marriage}_i^* = x_i'\gamma + \varepsilon_{3i}, \\ \text{alcohol}_i^* = x_i'\delta + \varepsilon_{4i}, \\ \text{smoking}_i^* = x_i'\lambda + \varepsilon_{5i}, \\ \text{phys\_active}_i^* = x_i'\tau + \varepsilon_{6i}, \end{cases}$$

$$\begin{aligned} Y_i &= \mathbf{1}\{Y_i^* \geq 0\}, \quad d_i = \text{diploma}_i = \mathbf{1}\{\text{diploma}_i^* \geq 0\}, \quad m_i = \text{marriage}_i = \mathbf{1}\{\text{marriage}_i^* \geq 0\}, \\ a_i &= \text{alcohol}_i = \mathbf{1}\{\text{alcohol}_i^* \geq 0\}, \quad s_i = \text{smoking}_i = \mathbf{1}\{\text{smoking}_i^* \geq 0\}, \\ ph_i &= \text{phys\_active}_i = \mathbf{1}\{\text{phys\_active}_i^* \geq 0\}. \end{aligned}$$

Логарифмическая функция правдоподобия для данной многомерной рекурсивной пробит-модели будет выглядеть следующим образом:

$$\begin{aligned} \ln L(\beta, \alpha, \gamma, \delta, \lambda, \tau, V) &= \\ &= \ln \prod_{i=1}^N \Phi_6((2Y_i - 1)(x_i'\beta + \beta_1 d_i), (2d_i - 1)x_i'\alpha, (2m_i - 1)x_i'\gamma, (2a_i - 1)x_i'\delta, (2s_i - 1)x_i'\lambda, (2ph_i - 1)x_i'\tau; V), \end{aligned}$$

где  $\Phi_6(\cdot)$  – функция распределения 6-мерного нормального распределения с нулевыми математическими ожиданиями и следующей ковариационной матрицей:

$$V = \begin{pmatrix} 1 & (2Y_i - 1)(2d_i - 1)\rho_{12} & \dots & (2Y_i - 1)(2ph_i - 1)\rho_{16} \\ (2Y_i - 1)(2d_i - 1)\rho_{12} & \ddots & \ddots & (2d_i - 1)(2ph_i - 1)\rho_{26} \\ \vdots & \ddots & \ddots & \vdots \\ (2Y_i - 1)(2ph_i - 1)\rho_{16} & (2d_i - 1)(2ph_i - 1)\rho_{26} & \dots & 1 \end{pmatrix}.$$

Тогда на основе оценок параметров модели  $(\hat{\beta}, \hat{\alpha}, \hat{\gamma}, \hat{\delta}, \hat{\lambda}, \hat{\tau}, \hat{V})'$ , полученных в ходе максимизации данной логарифмической функции правдоподобия, средний эффект воздействия АТЕ рассчитывается по следующей формуле:

$$\widehat{ATE} = \frac{1}{N} \sum_{i=1}^N \left[ \hat{P}(Y_i(1) = 1 | x_i) - \hat{P}(Y_i(0) = 1 | x_i) \right],$$

где

$$\begin{aligned} \hat{P}(Y_i(1) = 1 | x_i) &= \hat{P}(Y_i^*(1) > 0 | x_i) = \hat{P}(x_i' \hat{\beta} + \hat{\beta}_1 + \varepsilon_{1i} > 0 | x_i) = \hat{P}(-\varepsilon_{1i} < x_i' \hat{\beta} + \hat{\beta}_1 | x_i) = \\ &= \frac{\hat{\Phi}_5(x_i' \hat{\beta} + \hat{\beta}_1, (2m_i - 1)x_i' \hat{\gamma}, (2a_i - 1)x_i' \hat{\delta}, (2s_i - 1)x_i' \hat{\lambda}, (2ph_i - 1)x_i' \hat{\tau}; \hat{V}_5(1))}{\hat{\Phi}_4((2m_i - 1)x_i' \hat{\gamma}, (2a_i - 1)x_i' \hat{\delta}, (2s_i - 1)x_i' \hat{\lambda}, (2ph_i - 1)x_i' \hat{\tau}; \hat{V}_4)}, \end{aligned}$$

$$\begin{aligned} \hat{P}(Y_i(0) = 1 | x_i) &= \hat{P}(Y_i^*(0) > 0 | x_i) = \hat{P}(x_i' \hat{\beta} + \varepsilon_{1i} > 0 | x_i) = \hat{P}(-\varepsilon_{1i} < x_i' \hat{\beta} | x_i) = \\ &= \frac{\hat{\Phi}_5(x_i' \hat{\beta}, (2m_i - 1)x_i' \hat{\gamma}, (2a_i - 1)x_i' \hat{\delta}, (2s_i - 1)x_i' \hat{\lambda}, (2ph_i - 1)x_i' \hat{\tau}; \hat{V}_5(1))}{\hat{\Phi}_4((2m_i - 1)x_i' \hat{\gamma}, (2a_i - 1)x_i' \hat{\delta}, (2s_i - 1)x_i' \hat{\lambda}, (2ph_i - 1)x_i' \hat{\tau}; \hat{V}_4)}, \end{aligned}$$

$$\hat{V}_5(1) = \begin{pmatrix} 1 & (2m_i - 1)\hat{\rho}_{13} & \dots & (2ph_i - 1)\hat{\rho}_{16} \\ (2m_i - 1)\hat{\rho}_{13} & \ddots & \ddots & (2m_i - 1)(2ph_i - 1)\hat{\rho}_{36} \\ \vdots & \ddots & \ddots & \vdots \\ (2ph_i - 1)\hat{\rho}_{16} & (2m_i - 1)(2ph_i - 1)\hat{\rho}_{36} & \dots & 1 \end{pmatrix},$$

$$\hat{V}_4 = \begin{pmatrix} 1 & (2m_i - 1)(2a_i - 1)\hat{\rho}_{34} & \dots & (2m_i - 1)(2ph_i - 1)\hat{\rho}_{36} \\ (2m_i - 1)(2a_i - 1)\hat{\rho}_{34} & \ddots & \ddots & (2a_i - 1)(2ph_i - 1)\hat{\rho}_{46} \\ \vdots & \ddots & \ddots & \vdots \\ (2m_i - 1)(2ph_i - 1)\hat{\rho}_{36} & (2a_i - 1)(2ph_i - 1)\hat{\rho}_{46} & \dots & 1 \end{pmatrix}.$$

### Приложение 3.

#### Метрики качества баланса между распределениями контрольных переменных в группе воздействия и контрольной группе, достигнутого после проведения процедур мэтчинга

**Таблица П1.**  
Значения метрик качества баланса между группой воздействия и контрольной группой до проведения мэтчинга для женщин 1945–1954 гг.

Название контрольной переменной	ASMD	Variance Ratio	eCDF Mean	eCDF Max
age	0,0962	1,0453	0,0190	0,0326
soblc	0,3674	–	0,1683	0,1683
sgorod	0,0150	–	0,0070	0,0070
invalid	0,0996	–	0,0366	0,0366

*Примечание:* Variance Ratio, как правило, не рассчитывается для бинарных переменных, поскольку, являясь функцией от выборочных долей, не дает дополнительной информации о балансе при наличии ASMD и eCDF-статистик.

**Таблица П2.**  
Значения метрик качества баланса между группой воздействия и контрольной группой после проведения процедуры PS-мэтчинга для женщин 1945–1954 гг.

Название контрольной переменной	ASMD	Variance Ratio	eCDF Mean	eCDF Max
age	0,0413	1,0166	0,0082	0,0126
soblc	0,0007	–	0,0003	0,0003
sgorod	0,0014	–	0,0007	0,0007
invalid	0,0333	–	0,0122	0,0122

Таблица ПЗ.

**Значения метрик качества баланса между группой воздействия и контрольной группой после проведения процедуры мэтчинга на основе расстояния Махаланобиса для женщин 1945–1954 гг.**

Название контрольной переменной	ASMD	Variance Ratio	eCDF Mean	eCDF Max
age	0.0010	1.0016	0.0002	0.0033
soblc	0.0000	–	0.0000	0.0000
sgorod	0.0000	–	0.0000	0.0000
invalid	0.0000	–	0.0000	0.0000

\* \*  
\*

### СПИСОК ЛИТЕРАТУРЫ

- Ениколопов Р.* Оценивание эффекта воздействия // Квантиль. 2009. № 6. С. 3–14.
- Засимова Л.С., Колосницына М.Г., Красильникова М.Д.* Изменение поведения россиян в отношении здорового образа жизни (по результатам социологических опросов 2011 и 2017 гг. // Препринт № WP8/2017/02). М.: Изд. дом ВШЭ, 2017.
- Коссова Е.В., Косорукова М.А.* Оценивание влияния высшего образования на здоровье: сравнение многомерной рекурсивной пробит-модели и мэтчинга // Прикладная эконометрика. 2023. Т. 69. С. 65–90.
- Коссова Е.В., Слаболицкий И.С., Потанин Б.С.* Оценивание вероятностей, предельных эффектов и эффектов воздействия в иерархических системах бинарных уравнений // Вестник Московского университета. Серия 6: Экономика. 2023. Т. 58. № 4. С. 23–49.
- Табакаев М.В., Артамонова Г.В.* Урбанизация и сердечно-сосудистые заболевания в современном обществе // Российский кардиологический журнал. 2015. Т. 6. № 122. С. 94–99.
- Черкасов С.Н., Киртадзе И.Д., Камаев Ю.О., Олейникова В.С.* Влияние образования на заболеваемость и интенсивность потребления услуг здравоохранения в старших возрастных группах // Бюллетень Национального научно-исследовательского института общественного здоровья имени Н.А. Семашко. 2019. № 3–4. С. 129–139.
- Albarrán P., Hidalgo-Hidalgo M., Iturbe-Ormaetxe I.* Education and Adult Health: Is There a Causal Effect? // Social Science & Medicine. 2020. Vol. 249. P. 112830.
- Berger M.C., Leigh J.P.* Schooling, Self-Selection, and Health // The Journal of Human Resources. 1989. Vol. 24. № 3. P. 433–455.
- Brunello G., Fabbri D., Fort M.* The Causal Effect of Education on Body Mass: Evidence from Europe // Journal of Labor Economics. 2013. Vol. 31. № 1. P. 195–223.
- Brunello G., Fort M., Schneeweis N., Winter-Ebmer R.* The Causal Effect of Education on Health: What Is the Role of Health Behaviors? // Health Economics. 2016. Vol. 25. № 3. P. 314–336.

- Cappellari L., Jenkins S.P.* Multivariate Probit Regression Using Simulated Maximum Likelihood // The Stata Journal: Promoting Communications on Statistics and Stata. 2003. Vol. 3. № 3. P. 278–294.
- Chevalier A.* Parental Education and Child's Education: A Natural Experiment // SSRN Electronic Journal. 2004. (<http://dx.doi.org/10.2139/ssrn.553922>)
- Clark D., Royer H.* The Effect of Education on Adult Mortality and Health: Evidence from Britain // The American Economic Review. 2013. Vol. 103. № 6. P. 2087–2120.
- Davies N.M., Dickson M., Davey Smith G., Van Den Berg G.J., Windmeijer F.* The Causal Effects of Education on Health Outcomes in the UK Biobank // Nature Human Behaviour. 2018. Vol. 2. № 2. P. 117–125.
- Dalstra J.A. et al.* Socioeconomic Differences in the Prevalence of Common Chronic Diseases: An Overview of Eight European Countries // International Journal of Epidemiology. 2005. Vol. 34. № 2. P. 316–326.
- Delaruelle K., Buffel V., Bracke P.* Educational Expansion and the Education Gradient in Health: A Hierarchical Age-Period-Cohort Analysis // Social Science & Medicine. 2015. Vol. 145. P. 79–88.
- Dilmaghani M.* The Causal Effects of Education on Health over the Life Course: Evidence from Canada // Public Health. 2020. Vol. 186. P. 170–177.
- Eide E.R., Showalter M.H.* Estimating the Relation between Health and Education: What Do We Know and What Do We Need to Know? // Economics of Education Review. 2011. Vol. 30. № 5. P. 778–791.
- Fonseca R., Michaud P.-C., Zheng Y.* The Effect of Education on Health: Evidence from National Compulsory Schooling Reforms // SERIEs. 2020. Vol. 11. № 1. P. 83–103.
- Fonseca R., Zheng Y.* The Effect of Education on Health: Cross-Country Evidence // SSRN Electronic Journal. 2011. (<http://dx.doi.org/10.2139/ssrn.1924005>)
- Grabner M.* The Causal Effect of Education on Obesity: Evidence from Compulsory Schooling Laws // SSRN Electronic Journal. 2009. (<http://dx.doi.org/10.2139/ssrn.1505075>)
- Greene W.H.* Econometric Analysis. 7<sup>th</sup> ed., international ed. Harlow: Pearson, 2012. ISBN 978-0-273-75356-8.
- Hamad R., Elser H., Tran D.C., Rehkopf D.H., Goodman S.N.* How and Why Studies Disagree about the Effects of Education on Health: A Systematic Review and Meta-Analysis of Studies of Compulsory Schooling Laws // Social Science & Medicine. 2018. Vol. 212. P. 168–178.
- Havari E., Savegnago M.* The Causal Effect of Parents' Schooling on Children's Schooling in Europe. A New IV Approach // CEIS Working Paper № 315 // SSRN Electronic Journal. 2014. (<http://dx.doi.org/10.2139/ssrn.2435933>)
- Hu A.* The Health Benefits of College Education in Urban China: Selection Bias and Heterogeneity // Social Indicators Research. 2014. Vol. 115. № 3. P. 1101–1121.
- Kossova T., Kossova E., Sitnikova A., Sheluntcova M.* Evaluation of Changes in Alcohol Consumption: Evidence from Russia // Journal of Economic Studies. 2021. Vol. 48. № 7. P. 1336–1345.
- Lynch S.M.* Cohort and Life-Course Patterns in the Relationship between Education and Health: A Hierarchical Approach // Demography. 2003. Vol. 40. № 2. P. 309–331.
- Mehta P.D.* Control Variable in Research // International Encyclopedia of the Social & Behavioral Sciences. 2001. P. 2727–2730. (<https://doi.org/10.1016/B0-08-043076-7/00734-8>)
- Mourifie I., Meango R.* A Note on the Identification in Two Equations Probit Model with Dummy Endogenous Regressor // Economics Letters. 2014. Vol. 125. №3. P. 360–363.
- Pisa P.T., Behanan R., Vorster H.H., Kruger A.* Social Drift of Cardiovascular Disease Risk Factors in Africans from the North West Province of South Africa: The PURE Study // Cardiovascular Journal of Africa. 2012. Vol. 23. № 7. P. 371–388.
- Rubin D.B.* Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies // Journal of Educational Psychology. 1974. Vol. 66. № 5. P. 688–701.
- Rubin D.B.* Assignment to Treatment Group on the Basis of a Covariate // Journal of Educational Statistics. 1977. Vol. 2. № 1. P. 1–26.
- Schlotter M., Schwerdt G., Woessmann L.* Econometric Methods for Causal Evaluation of Education Policies and Practices: A Non-Technical Guide // Education Economics. 2011. Vol. 19. № 2. P. 109–137.
- Silles M.A.* The Causal Effect of Education on Health: Evidence from the United Kingdom // Economics of Education Review. 2009. Vol. 28. № 1. P. 122–128.

*Sodjinou R., Agueh V., Fayomi B., Delisle H.* Obesity and Cardio-Metabolic Risk Factors in Urban Adults of Benin: Relationship with Socio-Economic Status, Urbanisation, and Lifestyle Patterns // *BMC Public Health*. 2008. Vol. 8. № 84. P. 1–13.

*Stuart E.A.* Matching Methods for Causal Inference: A Review and a Look Forward // *Statistical Science*. 2010. Vol. 25. № 1. P. 1–21.

*Viego V., Temporelli K.* Socioeconomic Status and Self-Reported Chronic Diseases among Argentina's Adult Population: Results Based on Multivariate Probability Models // *Journal of Public Health Research*. 2017. Vol. 6. № 1. P. 55–63.

## Treatment Effect of Higher Education on Health in Age Cohort Models

**Mariia Kosorukova<sup>1</sup>, Elena Kossova<sup>2</sup>,  
Bogdan Potanin<sup>3</sup>, Aleksandr Shylo<sup>4</sup>**

<sup>1</sup> National Research University Higher School of Economics,  
20, Myasnitskaya str., Moscow, 101000, Russian Federation.  
E-mail: kosorukova.mariya@gmail.com

<sup>2</sup> National Research University Higher School of Economics,  
20, Myasnitskaya str., Moscow, 101000, Russian Federation.  
E-mail: ekossova@hse.ru

<sup>3</sup> National Research University Higher School of Economics,  
20, Myasnitskaya str., Moscow, 101000, Russian Federation.  
E-mail: bogdanpotanin@gmail.com

<sup>4</sup> National Research University Higher School of Economics,  
20, Myasnitskaya str., Moscow, 101000, Russian Federation.  
E-mail: alexandr.vas.shilo@gmail.com

The work is devoted to the estimation of the average treatment effect of higher education on health, assuming the existence of structural differences in the dependencies between the studied factors by gender and the birth cohort of individuals. The effects are estimated separately for each cohort of men and women using parametric (multivariate recursive probit

---

**Mariia Kosorukova** – postgraduate student of the Department of Applied Economics, Faculty of Economic Sciences.

**Elena Kossova** – PhD (Candidate of Sciences) in Physics and Mathematics, Associate Professor of the Department of Applied Economics, Faculty of Economic Sciences.

**Bogdan Potanin** – PhD (Candidate of Sciences) in Economics, Associate Professor of the Department of Applied Economics, Faculty of Economic Sciences.

**Aleksandr Shylo** – master's degree student, Faculty of Economic Sciences.

model) and nonparametric (matching) methods. The use of fundamentally different approaches makes it possible to verify the robustness of the results to violating the assumptions of the applied methods. During the study the statistical evidence in favor of the existence of significant negative average treatment effects of higher education on the probability of obesity and the probability of hypertension and significant positive effects on the likelihood of high self-esteem of health was revealed for women. For men, the statistical evidence in favor of the presence of positive average treatment effects on the probability of having heart diseases, hypertension, allergy, and high self-esteem of health was found. A significant differentiation of the average treatment effects depending on the age cohort was revealed. For certain types of diseases, the cohort diminishing hypothesis for education is confirmed. The conclusion is made about the necessity of studying the impact of education on individuals' health separately for men and women born in different time periods, as well as the inexpediency of using aggregated health indicators when analyzing the impact of education on health. The results of the study allow determining the nature of cohort changes in the impact of education on health of the Russian population and can be taken into account when forming state policy on education as a source of positive external effects on health.

**Key words:** education; health; treatment effect; age cohorts; matching; multivariate probit model.

**JEL Classification:** C14, C21, C31, C35, I19, I26.

\* \*

\*

### References

- Albarrán P., Hidalgo-Hidalgo M., Iturbe-Ormaetxe I. (2020) Education and Adult Health: Is There a Causal Effect? *Social Science & Medicine*, 249, p. 112830.
- Berger M.C., Leigh J.P. (1989) Schooling, Self-Selection, and Health. *The Journal of Human Resources*, 24, 3, pp. 433–455.
- Brunello G., Fabbri D., Fort M. (2013) The Causal Effect of Education on Body Mass: Evidence from Europe. *Journal of Labor Economics*, 31, 1, pp. 195–223.
- Brunello G., Fort M., Schneeweis N., Winter-Ebmer R. (2016) The Causal Effect of Education on Health: What is the Role of Health Behaviors? *Health Economics*, 25, 3, pp. 314–336.
- Cappellari L., Jenkins, S.P. (2003) Multivariate Probit Regression using Simulated Maximum Likelihood. *The Stata Journal: Promoting Communications on Statistics and Stata*, 3, 3, pp. 278–294.
- Cherkasov S., Kirtadze I., Kamaev Yu., Oleynikova V. (2019) Influence of Education on Morbidity and Intensity of Consumption of Health Services in Older Age Groups. *Bulletin of Semashko National Research Institute of Public Health*, 3–4, pp. 129–139.
- Chevalier A. (2004) Parental Education and Child's Education: A Natural Experiment. *SSRN Electronic Journal*. (<http://dx.doi.org/10.2139/ssrn.553922>)

- Clark D., Royer H. (2013) The Effect of Education on Adult Mortality and Health: Evidence from Britain. *The American Economic Review*, 103, 6, pp. 2087–2120.
- Davies N.M., Dickson M., Davey Smith G., Van Den Berg G.J., Windmeijer F. (2018) The Causal Effects of Education on Health Outcomes in the UK Biobank. *Nature Human Behaviour*, 2, 2, pp. 117–125.
- Dalstra J.A. et al. (2005) Socioeconomic Differences in the Prevalence of Common Chronic Diseases: An Overview of Eight European Countries. *International Journal of Epidemiology*, 34, 2, pp. 316–326.
- Delaruelle K., Buffel V., Bracke P. (2015) Educational Expansion and the Education Gradient in Health: A Hierarchical Age-Period-Cohort Analysis. *Social Science & Medicine*, 145, pp. 79–88.
- Dilmaghani M. (2020) The Causal Effects of Education on Health over the Life Course: Evidence from Canada. *Public Health*, 186, pp. 170–177.
- Eide E.R., Showalter M.H. (2011) Estimating the Relation Between Health and Education: What Do We Know and What Do We Need to Know? *Economics of Education Review*, 30, 5, pp. 778–791.
- Enikolopov R. (2009) Estimation of Treatment Effects. *Quantile*, 6, pp. 3–14. (In Russ.)
- Fonseca R., Michaud P.-C., Zheng Y. (2020) The Effect of Education on Health: Evidence from National Compulsory Schooling Reforms. *SERIEs*, 11, 1, pp. 83–103.
- Fonseca R., Zheng Y. (2011) The Effect of Education on Health: Cross-Country Evidence. *SSRN Electronic Journal*. (<http://dx.doi.org/10.2139/ssrn.1924005>)
- Grabner M. (2009) The Causal Effect of Education on Obesity: Evidence from Compulsory Schooling Laws. *SSRN Electronic Journal*. (<http://dx.doi.org/10.2139/ssrn.1505075>)
- Greene W.H. (2012) *Econometric Analysis*. 7<sup>th</sup> ed., international ed. Harlow: Pearson. ISBN 978-0-273-75356-8.
- Hamad R., Elser H., Tran D.C., Rehkopf D.H., Goodman S.N. (2018) How and Why Studies Disagree about the Effects of Education on Health: A Systematic Review and Meta-Analysis of Studies of Compulsory Schooling Laws. *Social Science & Medicine*, 212, pp. 168–178.
- Havari E., Savegnago M. (2014) The Causal Effect of Parents' Schooling on Children's Schooling in Europe. A New IV Approach. CEIS Working Paper no 315. *SSRN Electronic Journal*. (<http://dx.doi.org/10.2139/ssrn.2435933>)
- Hu A. (2014) The Health Benefits of College Education in Urban China: Selection Bias and Heterogeneity. *Social Indicators Research*, 115, 3, pp. 1101–1121.
- Kossova E., Kosorukova M. (2023) Estimation of the Treatment Effect of Higher Education on Health: Comparison of the Multivariate Recursive Probit Model and Matching. *Applied Econometrics*, 69, pp. 65–90. (In Russ.)
- Kossova T., Kossova E., Sitnikova A., Sheluntcova M. (2021) Evaluation of Changes in Alcohol Consumption: Evidence from Russia. *Journal of Economic Studies*, 48, 7, pp. 1336–1345.
- Kossova E.V., Slabolitskiy I.S., Potanin B.S. (2023) Estimating Probabilities, Marginal Effects and Treatment Effects in Recursive Bivariate Probit Models. *Lomonosov Economics Journal*, 58, 4, pp. 23–49. (In Russ.)
- Lynch S.M. (2003) Cohort and Life-Course Patterns in the Relationship between Education and Health: A Hierarchical Approach. *Demography*, 40, 2, pp. 309–331.
- Mehta P.D. (2001) Control Variable in Research. *International Encyclopedia of the Social & Behavioral Sciences*, pp. 2727–2730. (<https://doi.org/10.1016/B0-08-043076-7/00734-8>)
- Mourifie I., Meango R. (2014) A Note on the Identification in Two Equations Probit Model with Dummy Endogenous Regressor. *Economics Letters*, 125, 3, pp. 360–363.
- Pisa P.T., Behanan R., Vorster H.H., Kruger A. (2012) Social Drift of Cardiovascular Disease Risk Factors in Africans from the North West Province of South Africa: The PURE Study. *Cardiovascular Journal of Africa*, 23, 7, pp. 371–388.
- Rubin D.B. (1974) Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66, 5, pp. 688–701.
- Rubin D.B. (1977) Assignment to Treatment Group on the Basis of a Covariate. *Journal of Educational Statistics*, 2, 1, pp. 1–26.

Schlotter M., Schwerdt G., Woessmann L. (2011) Econometric Methods for Causal Evaluation of Education Policies and Practices: A Non-Technical Guide. *Education Economics*, 19, 2, pp. 109–137.

Silles M.A. (2009) The Causal Effect of Education on Health: Evidence from the United Kingdom. *Economics of Education Review*, 28, 1, pp. 122–128.

Sodjinou R., Agueh V., Fayomi B., Delisle H. (2008) Obesity and Cardio-Metabolic Risk Factors in Urban Adults of Benin: Relationship with Socio-Economic Status, Urbanisation, and Lifestyle Patterns. *BMC Public Health*, 8, 84, pp. 1–13.

Stuart E.A. (2010) Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science*, 25, 1, pp. 1–21.

Tabakaev M.V., Artamonova G.V. (2015) Urbanization and Cardiovascular Diseases in Modern Society. *Russian Journal of Cardiology*, 6, 122, pp. 94–99. (In Russ.)

Viego V., Temporelli K. (2017) Socioeconomic Status and Self-Reported Chronic Diseases Among Argentina's Adult Population: Results Based on Multivariate Probability Models. *Journal of Public Health Research*, 6, 1, pp. 55–63.

Zasimova L.S., Kolosnitsyna M.G., Krasilnikova M.D. (2017) *Changes in the Behavior of Russians with Regard to a Healthy Lifestyle* (Based on the Results of Sociological Surveys in 2011 and 2017) (Preprint no WP8/2017/02). Moscow: Higher School of Economics. (In Russ.)

Подписано в печать 23.06.2025. Формат 70×100 <sup>1</sup>/<sub>16</sub>.

Печ. л. 12,0. Тираж 150 экз. Заказ №

Отпечатано ООО «Компания Полиграфмастер»

125130, Москва, ул. Клары Цеткин, д. 28