

Прикладная эконометрика, 2018, т. 51, с. 102–125.

Applied Econometrics, 2018, v. 51, pp. 102–125.

М. А. Канева, В. М. Байдин¹

Гетерогенность ответов при самооценке здоровья россиян

Целью статьи является анализ феномена гетерогенности при самооценке здоровья для населения России в 2001–2015 гг. с выделением типов гетерогенности и ее социально-экономических детерминант, а также детерминант, связанных со здоровьем. Модель обобщенного порядкового пробита со случайными эффектами с учетом объективного индекса заболеваемости выявила, что детерминантами наблюдаемой гетерогенности в РФ для обоих полов являются семейное положение, доход, неполное среднее образование, проживание за пределами города и в отдельных регионах РФ, для женщин на разброс самооценок здоровья также влияет возраст. Для женщин детерминантами наблюдаемой гетерогенности также является среднее образование, а набор регионов, вносящих гетерогенность типа сдвига пороговых значений, различается для мужчин и женщин. Исследование показало, что прямое сравнение самооценок здоровья для различных групп населения без учета гетерогенности приводит к неверным выводам.

Ключевые слова: гетерогенность; самооценка здоровья; панельная регрессия; РМЭЗ; Россия.

JEL classification: I12; C21; C23.

1. Введение

Одним из наиболее широко используемых показателей в экономике здравоохранения является индикатор «самооценка здоровья», позволяющий респондентам на вопрос «Как Вы оцениваете свое здоровье?» выбрать ответ среди пяти категорий: «очень хорошее», «хорошее», «среднее», «плохое» и «очень плохое». Всемирная организация здравоохранения в 1996 г. рекомендовала к использованию показатель самооценки здоровья как один из основных индикаторов для мониторинга здоровья и качества жизни населения (Health interview surveys..., 1996). Простота вопроса и возможность оценить уровень здоровья индивида без непосредственного участия медицинского работника определили широкое распространение данной субъективной оценки на практике. Самооценка здоровья отражает физическое состояние респондента, а также психологический настрой и социальное благополучие (Au, Johnston, 2014; Cairney, 2000; Eriksson et al., 2001; Jylhä, 2009).

Однако субъективный характер данной меры оценки здоровья может привести к тому, что некоторые группы населения могут систематически занижать или, наоборот, завышать

¹ Байдин Валерий Михайлович — Научно-исследовательский финансовый институт министерства финансов Российской Федерации, Москва; vbaydin@nifi.ru.

Канева Мария Александровна — Институт экономической политики им. Е. Т. Гайдара, Институт экономики и организации промышленного производства СО РАН, Москва; mkaneva@gmail.com.

самооценку здоровья относительно ее объективного статуса, например, оценки самочувствия индивида врачом или же наличия у индивида хронических заболеваний. Причины для подобных систематических различий в самооценке здоровья могут быть разнообразными: национальные особенности граждан (например, более пессимистичные оценки жителей отдельных стран); поведение, преследующее определенную цель (например, желание получить пособие для нетрудоспособного населения) или же возрастные отличия. Важно, что наличие таких систематических различий приводит к тому, что прямые сравнения данных по самооценке здоровья между разными социально-экономическими группами без выделения факторов, являющихся источниками (детерминантами) гетерогенности, будут приводить к заведомо ложным выводам.

Феномен существования систематических различий при ответе на вопрос о самооценке здоровья для разных групп респондентов (мужчины / женщины, респонденты разных возрастных категорий, жители разных стран) получил название «гетерогенности (ответов)» при самооценке здоровья (Lindeboom, van Doorslaer, 2004).

Одним из наглядных примеров существования гетерогенности при самооценке здоровья является систематическое завышение оценок своего здоровья пожилыми респондентами по сравнению с другими возрастными группами, исследуемое в статье (Idler, 1996) на примере опроса жителей г. Нью-Хевен (США)². Среди возможных причин оптимистических оценок старших групп респондентов Idler предложила три основных: возрастные изменения, эффект когорты и эффект выживания. Действительно, возможно, с возрастом люди учатся игнорировать свои проблемы со здоровьем и адаптируются к хроническим заболеваниям (Livneh, Antonak, 2005), или же завышенная оценка собственного здоровья объясняется тем, что среди пожилого населения выживают самые здоровые индивиды? Наконец, возможно, что отношение к собственному здоровью меняется от поколения к поколению, и более старшие когорты склонны давать более оптимистическую оценку собственному здоровью. Какой бы ответ ни дал исследователь на данный вопрос, важна сама его постановка, а именно — признание феномена систематического различия в оценке здоровья старшим поколением, т. е. существование гетерогенности.

В настоящей статье авторы приводят результаты анализа феномена гетерогенности для российских респондентов на основе данных опроса РМЭЗ за 2001–2015 гг. с выделением социально-экономических детерминант гетерогенности и детерминант, связанных со здоровьем, а также определение типа гетерогенности. Основным методом является эконометрическое моделирование.

На основе количественных оценок авторы показывают, что в настоящее время в Российской Федерации основными социально-экономическими характеристиками гетерогенности самооценки здоровья для обоих полов являются: семейное положение, образование, доход, тип населенного пункта и регион проживания. Для женщин основной детерминантой наблюдаемой гетерогенности является наличие среднего образования, а набор регионов проживания, вносящих гетерогенность (типа сдвига пороговых значений), различается для мужчин

² Авторы обращают внимание на то, что возраст (в виде непрерывной переменной) обычно не рассматривается как возможная детерминанта гетерогенности, поскольку это означало бы необходимость сравнения респондентов одного возраста для учета гетерогенности по возрасту. Однако в литературе в качестве детерминант гетерогенности рассматриваются возрастные группы (например, моложе и старше 50 лет), отдельные возрастные интервалы для пожилого населения (Hirve et al., 2014) или квинтили возраста (Schneider et al., 2012).

и для женщин. Для обоих полов также наблюдается гетерогенность при самооценке здоровья в отдельные годы рассматриваемого периода.

Статья имеет следующую структуру. В разделе 2 приводится определение гетерогенности и описание двух ее типов. В разделе 3 представлен обзор литературы, моделей и методов исследования феномена гетерогенности. В четвертом разделе описываются данные, используемые в эконометрическом моделировании, и характеристики выбранных для анализа переменных. Раздел 5 посвящен обсуждению построения объективного индекса заболеваемости и его роли в эконометрическом анализе гетерогенности. В разделе 6 представлена построенная эконометрическая модель. Результаты моделирования приводятся и обсуждаются в разделе 7.

2. Феномен гетерогенности и ее типы

В данном разделе представлена модель гетерогенности, основой для которой стала работа (Lindeboom, van Doorslaer, 2004).

Любые вариации в самооценке здоровья могут быть разделены на вариации, связанные с изменениями в истинном здоровье, и на вариации, непосредственно связанные с самой мерой самооценки при неизменном истинном здоровье. Второй тип вариаций представляет собой гетерогенность в самооценке здоровья (Shmueli, 2003; Герри, Байдин, 2017). С помощью модели, приведенной в данной статье, можно определить и разделить между собой изменения, которые отражаются в самооценке здоровья, на те, что связаны с реакцией на перемены в истинном здоровье человека, и на те, что связаны с поведенческими факторами, влияющими на самооценку здоровья.

Пусть H^* — это ненаблюдаемое истинное здоровье индивида (латентная переменная), а H^S — его наблюдаемая субъективная оценка с категориями «отличное», «очень хорошее», «среднее», «плохое» и «очень плохое» здоровье. Представим, что H^* — это набор ненаблюдаемых индексов, покрывающих различные характеристики здоровья индивида, для простоты интерпретации в дальнейшем будем предполагать, что это один индекс. Как уже было сказано, H^* — это ненаблюдаемая величина, и, например, наблюдаемые болезни у человека через какой-то функциональный вид могут служить прокси переменной для истинного здоровья индивида (Linderboom, 2004). Этот факт будет использоваться ниже.

Для категориальной переменной H^S и соответствующей ей латентной переменной H^* можно определить порядковую пробит модель:

$$H^S = i \Leftrightarrow c_{i-1} < H^* \leq c_i, \quad i = 1, \dots, n, \quad (1)$$

$$H^* = f(H^0; \alpha) + X_2' \beta_2 + \varepsilon_2. \quad (2)$$

H^* в уравнении (2) зависит от X_2 — факторов, влияющих на здоровье (true health), например, окружающая среда, производственные травмы и т.д., H^0 — здоровье, которое в процессе времени подвергалось воздействию факторов и на момент анализа является истинным здоровьем. H^* можно представить в виде линейной комбинации (2), так что если не будет никаких внешних факторов, то H^* будет соответствовать уровню здоровья H^0 .

В упорядоченной пробит модели (1)–(2) лежит предположение о «параллельных линиях» (proportional odds assumption) (Сизова и др., 2017), которое означает, что при сравнении

вероятности выбора одной категории относительно другой не принимается во внимание третья категория. Это предположение соответствует независимости констант пороговых значений c_i от факторов модели. Данное важное свойство будет использовано впоследствии.

Ослабим требование о параллельности сдвигов. Теперь точки разграничений могут изменяться в зависимости от множества поведенческих факторов X_1 при оценке собственного здоровья. При формулировке модели на первом шаге предполагается, что множества X_1 и X_2 могут иметь одинаковые по смыслу компоненты. Тогда

$$c_i = g_i(X_1; \beta_{1i}) = X_1' \beta_{1i} + \varepsilon_1, i = 1, \dots, n - 1, c_0 = -\infty, c_n = \infty. \quad (3)$$

Подставив уравнения (2)–(3) в (1), получим обобщенную порядковую пробит модель (формулы (4)–(6) ниже). Данная модель отличается от модели (1)–(2) тем, что существует дополнительный набор факторов, влияющий на сдвиг точек разграничений, а именно, $f_1(\cdot)$ описывает поведение индивида при оценке собственного здоровья (reporting behavior).

С учетом (3) выражение (2) преобразуется в

$$H^S = i \Leftrightarrow g_{i-1}(X_1; \beta_{1,i-1}) - X_2' \beta_2 < f(H^0; \alpha) + \varepsilon_2 \leq g_i(X_1; \beta_{1i}) - X_2' \beta_2. \quad (4)$$

Гетерогенность в самооценке здоровья выражается в том, что соотношение между H^* и H^S не является постоянным и отличается для разных выборок населения. Поэтому прямые сравнения H^S для различных групп населения будут отражать как разницу в состоянии здоровья, так и различия в поведении при оценке собственного здоровья. Например, люди, регулярно занимающиеся спортом, могут оценивать свое здоровье выше, поскольку ведут более здоровый образ жизни, и их здоровье на самом деле лучше; регулярные медицинские осмотры снижают неопределенность в симптомах и также увеличивают самооценку здоровья.

Предположения модели относительно влияния факторов X_1 и X_2 на определенные ранее типы здоровья и его самооценки могут быть записаны в виде следующих формул, где ε_1 и ε_2 являются случайными величинами:

$$H^S = f_1(H^*, X_1, \varepsilon_1; \beta_1), \quad (5)$$

$$H^* = f_2(H^0, X_2, \varepsilon_2; \beta_2). \quad (6)$$

Рассмотрим модель (4)–(6) для различных комбинаций оценок точек разграничений. В данном разделе не рассматриваются способы оценки и возможные связанные с этим проблемы. Гетерогенность в самооценке здоровья может изменять как среднее, так и форму распределения самооценки здоровья (Lindeboom, Doorslaer, 2004). При параллельном сдвиге точек разграничений c_i от изменения в поведенческих факторах возникает гетерогенность типа «сдвиг индекса» (index shift)³ или гетерогенность *первого типа*. Если нельзя отвергнуть нулевую гипотезу (7), то можно утверждать, что наблюдается гетерогенность первого типа.

³ В дальнейшем в статье будет введено понятие «объективный индекс заболеваемости». Авторы обращают внимание на то, что понятие «сдвиг индекса» в тексте статьи относится только к типу распределения самооценки здоровья и не подразумевает сдвиг объективного индекса заболеваемости. Терминология «сдвиг индекса» и «объективный индекс заболеваемости» сохраняется авторами вслед за другими работами по гетерогенности в самооценке здоровья (Lindeboom, Doorslaer, 2004; Jürges, 2006; Pfarr et al., 2012; Schneider et al., 2012 и др.).

Нулевая гипотеза говорит о том, что влияние фактора на каждую точку разграничений одинаково. Если k_i — это оценка фактора для i -й точки разграничения, то

$$H_0 : k_i = \text{const}, \forall i. \quad (7)$$

Если же сдвиг точек разграничений не параллелен, т.е. меняется относительное положение точек между собой, то возникает гетерогенность типа «сдвиг пороговых значений» (cut-point shift) или гетерогенность *второго типа*. В данном случае, если нулевая гипотеза (7) отвергается, то возникает гетерогенность второго типа.

Точки разграничений могут сдвигаться параллельно, если для какой-то подгруппы населения характерно занижение своего здоровья (например, у юношей призывного возраста при осмотре в военкомате). Примером непараллельного сдвига может быть ситуация, когда в какой-то группе населения не принято использовать выражение «очень хорошо» при оценке собственного здоровья (например, из-за суеверий или национальных различий, как в случае жителей Германии (Jürges, 2006, p. 164)). Более детально об оценке типа гетерогенности будет сказано ниже. Далее в разделе 5 разбирается проблема идентификации типа гетерогенности и возможный способ ее решения.

На основе сказанного выше становится ясным, что в задачи исследователя, анализирующего гетерогенность, входит не только выявление социально-экономических детерминант гетерогенности, но также и определение типа гетерогенности для ее различных детерминант. В следующем разделе представлен обзор работ, обсуждающих феномен гетерогенности и ее типы.

3. Обзор литературы

Одной из первых работ по теме гетерогенности самооценки здоровья стало исследование Kerkhofs и Lindeboom (1995), где изучались систематические изменения в самооценке здоровья среди неработающих респондентов, которые занижали оценку собственного здоровья с целью получения пособия по безработице. Построенная авторами модель на основе метода правдоподобия показала, что статус индивида на рынке труда (работающий или безработный) действительно оказывает влияние на его самооценку здоровья. Хотя данная работа посвящена проблеме гетерогенности самооценки здоровья, термин «гетерогенность» в тексте еще не употреблялся.

Работа (Lindeboom, van Doorslaer, 2004) продолжает тему гетерогенности и изучает сдвиги индекса и пороговых значений на примере данных опроса National Population Health в Канаде. Для выявления сдвигов авторы строят модель на основе метода максимального правдоподобия, которая сравнивает пороговые значения пяти категорий ответа на вопрос о самооценке здоровья («очень хорошее», «хорошее», «среднее», «плохое» и «очень плохое») с пороговыми значениями объективного показателя здоровья респондентов. В качестве объективного показателя здоровья используется индекс HUI3⁴. Авторы показывают, что среди социально-экономических причин гетерогенности в оценке здоровья для населения Канады выделяются возрастные группы и пол, но не доход, образование или владение языками.

⁴ McMaster Health Utility Index — индекс полезности здоровья МакМастер.

Hernández-Quevedo et al. (2004) рассматривают гетерогенность при самооценке здоровья на примере панельных данных Британского панельного опроса домохозяйств (BHPS⁵) с 1991 по 2001 г. В девятой волне вопрос о самооценке здоровья был задан иначе, нежели в остальных волнах BHPS — традиционные пять категорий ответа⁶ были изменены на следующие: «отличное» (здоровье), «очень хорошее», «хорошее», «среднее» и «плохое». Перефразирование вопроса представляло собой, по мнению авторов, натурный эксперимент (natural experiment), позволяющий выявить систематические расхождения в самооценке здоровья для различных групп респондентов. Используя объединенные порядковые пробит (pooled ordered probit) модели и обобщенные порядковые пробит (generalised ordered probit) модели со случайными эффектами, авторы показали, что изменение формулировки вопроса связано со сдвигом индекса, однако данный тип гетерогенности не связан с социально-экономическими характеристиками респондентов, а связан только с переименованием категорий.

Исследованию влияния конкретного фактора среди широкого набора социально-экономических характеристик населения на гетерогенность самооценки здоровья жителей Франции, а именно дохода, посвящена работа (Etilé, Milcent, 2006). С помощью обобщенной порядковой пробит модели авторы показывают, что функция влияния дохода на гетерогенность является выпуклой: в то время как обеспеченные респонденты склонны переоценивать свое здоровье, представители низкодходных групп, наоборот, недооценивают свое здоровье, т. е. дают более пессимистические оценки своего здоровья относительно его объективного состояния.

Исследования (Jürges, 2006) и (Pfarr et al., 2012) описывают национальные особенности самооценки здоровья. Обе работы используют базу данных Survey of Health, Ageing and Retirement of Europe (SHARE), в которой содержится информация о состоянии здоровья жителей 10 стран Европы. Jürges (2006), анализируя ответы на вопрос о самооценке здоровья в опросе SHARE, проведенного в 2004 г., с общим числом респондентов 22731, выявил существенные различия в ответах жителей стран Европы. Согласно его исследованию, жители Швеции и Дании склонны переоценивать состояние своего здоровья при ответе на вопрос о самооценке, в то время как жители Германии склонны его недооценивать. Помимо межстрановых различий, Pfarr et al. (2012) выявили также наличие гетерогенности при самооценке здоровья, связанной с частотой обращения к врачу и с разными объективными уровнями физического и психического здоровья.

Наконец, статья (Schneider et al., 2012) является наиболее близким исследованием к настоящей работе. Ее цель — выявить социально-экономические детерминанты наблюдаемой гетерогенности и ее тип (сдвиг индекса или сдвиг пороговых значений) на основе панельного регрессионного анализа панельных данных ФРГ. Важными результатами работы является вывод о значительных гендерных различиях в гетерогенности. Социально-экономические детерминанты гетерогенности отличаются для мужчин и женщин. Одним из таких факторов является образование: жительницы Германии с разным уровнем образования склонны по-разному оценивать свое здоровье, однако среди мужчин гетерогенности по уровню образования не наблюдается.

⁵ BHPS — British Household Panel Survey, <https://www.iser.essex.ac.uk/bhps>.

⁶ «Очень хорошее» здоровье, «хорошее», «среднее», «плохое» и «очень плохое».

4. Используемые данные

Для анализа гетерогенности использовались данные Российского мониторинга экономического положения и здоровья НИУ ВШЭ (РМЭЗ), негосударственного лонгитюдного обследования домохозяйств⁷. Авторы использовали данные опроса РМЭЗ за 2001–2015 гг., т. е. работали с волнами с 10 по 24. Подробно об особенностях опроса РМЭЗ написано в (Kozyureva et al., 2016). Данные о числе респондентов в каждой волне приводятся в табл. 1.

Таблица 1. Общее число респондентов в волнах с 2001 по 2015 г.

Год	2001	2002	2003	2004	2005	2006	2007	2008
N	9290	9199	9036	8935	8248	10639	10272	9810

Год	2009	2010	2011	2012	2013	2014	2015
N	9511	16804	16976	17090	16087	15137	15118

В РМЭЗ задается вопрос о самооценке здоровья и даются 5 вариантов ответов — «очень плохое», «плохое», «среднее», «хорошее» и «очень хорошее». Однако варианты «очень плохое» и «очень хорошее» имеют малое число наблюдений. Так, лишь 1.98% из общего числа респондентов в 2001–2015 гг. оценили свое здоровье как «очень хорошее», а доля респондентов, оценивающих здоровье как «очень плохое», составила 1.84%. В связи с этим, а также с тем, что выбранная модель затрудняет интерпретацию для пяти категорий, авторы определяют переменную «самооценка здоровья» только с тремя категориями: «плохое», «среднее» и «хорошее», объединяя «плохое и очень плохое» в категорию «плохое», а «хорошее и очень хорошее» в категорию «хорошее».

Модель самооценки здоровья базируется на более ранних теоретических предположениях и эмпирических исследованиях о взаимосвязи состояния здоровья и социально-экономических характеристик. Согласно модели Гроссмана (Grossman, 1972), капитал здоровья снижается с возрастом, увеличивается с ростом дохода и с повышением уровня образования. В эмпирических исследованиях также было показано, что социально-экономические характеристики способны влиять на состояние здоровья (Канева, 2016). В них, помимо переменных модели Гроссмана — возраста (Cairney, 2002; Kim, 2015), образования (Simon et al., 2000; Schneider et al., 2012) и дохода (Lokshin, Ravallion, 2008) — учитываются семейное положение (Bos, Bos, 2007; Monden, Uunk, 2013), тип населенного пункта (Gerry, Papadopoulos, 2015) и регион проживания (Lyytikäinen, Kempainen, 2016). Относительно влияния семейного положения на самооценку здоровья не существует единого мнения: так, в исследовании (Bos, Bos, 2007) было показано, что разведенные женщины в целом чувствуют себя лучше, нежели замужние респондентки, и это объясняется тем, что в браке на женщину ложатся тяжелые обязанности по ведению хозяйства. Противоположный результат был получен

⁷ 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: <http://www.cpc.unc.edu/projects/rlms> и <http://www.hse.ru/rlms>.

исследователями из ЕС: Monden, Uunk (2013) показали, что разведенные женщины нередко впадают в депрессию. В России брак традиционно ассоциируется с наличием партнера, который может позаботиться о супруге в случае болезни последнего, а также на протяжении жизни способен оказывать моральную поддержку (Синельников, 2011). Согласно более ранним исследованиям авторов, самооценка здоровья значимо отличается для жителей, живущих в поселках городского типа, по сравнению с городскими жителями, обычно в лучшую сторону, а жители столицы обоих полов и мужчины из Северного Кавказа чувствуют себя лучше по сравнению с остальными категориями респондентов (Kaneva et al., 2018). Кроме того, работы по гетерогенности самооценки здоровья также показали, что именно эти социально-экономические индикаторы могут являться и часто являются (Pfart et al., 2012; Schneider et al., 2012) детерминантами гетерогенности, за исключением возраста, который как непрерывная величина традиционно не относится к характеристикам гетерогенности, оказывая при этом влияние на другие детерминанты. Приведенные выше аргументы объясняют включение описываемого набора социально-экономических характеристик в независимые переменные модели гетерогенности. Описательные характеристики набора переменных приведены в табл. 2.

Таблица 2. Описательные статистики зависимой и независимых переменных панельной пробит регрессии самооценки здоровья, 2001–2015 гг.⁸ Спецификация 1

Переменная	Определение	Мужчины		Женщины	
		Среднее	SD	Среднее	SD
Зависимая переменная					
Sah	Самооценка здоровья, категории 1–3 (хорошее, среднее, плохое здоровье)	1.706	0.644	1.926	0.654
Независимые переменные (количественные)					
Age	Возраст, непрерывная переменная	43.156	16.935	47.997	18.625
Lninc_r	Натуральный логарифм реального дохода за последние 30 дней в ценах 2000 г.	8.104	0.990	7.763	0.974
Ind5m	Объективный индекс заболеваемости для мужчин из шести диагностированных заболеваний	24.260	10.928		
Ind5f	Объективный индекс заболеваемости для женщин из шести диагностированных заболеваний			42.271	11.334
Ind6m	Объективный индекс заболеваемости для мужчин из девяти заболеваний	22.547	16.949		
Ind6f	Объективный индекс заболеваемости для женщин из девяти заболеваний			41.671	20.861
Фиктивные переменные					
Sing	1, если респондент холост / не замужем	0.167	0.373	0.127	0.333
Div	1, если респондент разведен (а)	0.049	0.215	0.119	0.323
Wid	1, если респондент является вдовой/вдовцом	0.031	0.173	0.206	0.404
Mard*	1, если респондент женат/замужем	0.754	0.431	0.548	0.498
Incsec	1, если у респондента неполное среднее образование	0.216	0.411	0.197	0.398

⁸ Описательные статистики для индекса 2 и периода 2012–2015 гг. предоставляются авторами по запросу.

Окончание табл. 2

Переменная	Определение	Мужчины		Женщины	
		Среднее	SD	Среднее	SD
Second	1, если у респондента среднее образование	0.381	0.486	0.271	0.445
Vocat	1, если у респондента начальное или среднее профессиональное образование	0.201	0.401	0.287	0.452
Higher*	1, если у респондента высшее образование	0.202	0.402	0.244	0.430
Pgt	1, если респондент проживает в поселке городского типа	0.062	0.242	0.062	0.241
Rural	1, если респондент проживает в сельской местности	0.258	0.438	0.245	0.430
Urb*	1, если респондент проживает в городе	0.679	0.467	0.693	0.461
North	1, если респондент проживает в Северном или Северо-Западном районе	0.062	0.240	0.068	0.252
Centr	1, если респондент проживет в Центральном или Центрально-Черноземном районе	0.189	0.391	0.190	0.392
Volga	1, если респондент проживает в Волго-Вятском или Поволжском районе	0.177	0.381	0.173	0.378
Cauca	1, если респондент проживает в Северо-Кавказском районе	0.139	0.346	0.128	0.334
Ural	1, если респондент проживает в Уральском районе	0.141	0.348	0.141	0.349
West	1, если респондент проживает в Западно-Сибирском районе	0.088	0.283	0.089	0.285
East	1, если респондент проживает в Восточно-Сибирском или Дальневосточном районе	0.093	0.291	0.089	0.285
Capit*	1, если респондент проживает в Москве или Санкт-Петербурге	0.112	0.316	0.121	0.326
Объем выборки		69539		98065	

Примечание. * обозначает базовую категорию.

В методологии выделения детерминант гетерогенности и определения ее типа возникает проблема идентификации, которая может быть решена через построение и использование объективного индекса заболеваемости (Linderboom, Doorslaer, 2004; Jürges, 2006; Pfarr et al., 2012; Schneider et al., 2012).

5. Проблема идентификации и построение объективного индекса заболеваемости

В модели гетерогенности, изложенной в разделе 2, приводится определение двух типов гетерогенности, при оценке которых возникает проблема идентификации. Напомним, что гетерогенность — это такие изменения в самооценке здоровья, которые не связаны с изменениями в истинном здоровье человека, а связаны с факторами, влияющими на его самооценку, т. е. с множеством факторов X_1 в модели (4)–(6). Модель определена таким образом, что именно от факторов X_1 зависят точки разграничений c_i в обобщенной порядковой пробит модели.

Для определения типа гетерогенности необходима проверка гипотезы (7). Оценив модель (4)–(6) и отвергнув нулевую гипотезу (7), можно сделать вывод о наличии гетерогенности типа сдвига точек разграничений.

Однако в том случае, если нулевая гипотеза (7) не отвергается в рамках данной модели, нельзя однозначно определить, принадлежит ли отдельный фактор множеству X_1 или же множеству X_2 , т. е. фактор сдвигает параллельно точки разграничений или же влияет на здоровье индивида (согласно модели (3)–(5), факторы линейно влияют на здоровье индивида). Например, плохая экология может влиять на здоровье человека (фактор из множества X_2), но данный фактор также может влиять и на сдвиг точек разграничений (фактор из множества X_1), т. к. в городах с плохой экологией проживают жители менее обеспеченных слоев населения, которые могут смотреть на будущее без оптимизма.

Именно в рамках данной модели в случае, если нулевая гипотеза, описываемая (7), не отвергается, возникает проблема идентификации. В этом случае исследователь не может однозначно сделать вывод о наличии или отсутствии гетерогенности, т. е. определить, вызваны ли изменения в самооценке здоровья изменением в истинном здоровье или же изменениями в поведении индивида при самооценке здоровья.

Lindeboom, van Doorslaer (2004) отмечают, что «на практике исследователь не наблюдает H^* , а наблюдает лишь набор более объективных мер, таких как различные проблемы со здоровьем, заболевания или связанные со здоровьем ограничения на выполнение каких-либо действий в течение дня». Для решения описанной проблемы они разделяют все объясняющие переменные на переменные, отражающие социально-экономический статус, и переменные, определяющие состояние здоровья. Они высказали предположение о том, что для определения типа сдвига необходима информация об объективном статусе здоровья респондента, построенного на основе возможных переменных из множества X_2 (например, данных о проблемах со здоровьем, диагностируемых заболеваниях, включая психические, а также индикаторов типа силы рукопожатия, индекса массы тела и др.). Для того чтобы решить проблему идентификации, различить оценки факторов из множеств X_1 и X_2 , «зафиксируем» ненаблюдаемую величину истинного здоровья через его прокси-индекс заболеваемости.

Предполагается, что фиксация истинного здоровья в прокси переменной для здоровья позволяет решить проблему идентификации, поскольку остаются только факторы из множества X_1 . Например, потребление алкоголя может как влиять на самооценку здоровья, так и ухудшать само здоровье. «Зафиксировав» здоровье человека через прокси переменную, мы оцениваем эффект от алкоголя на самооценку, т. е. отвечаем на вопрос «что будет с самооценкой здоровья при изменении количества выпиваемого спирта и неизменном здоровье?». В этом случае исходная модель (4)–(6) превратится в обобщенную порядковую пробит регрессию⁹, в которой по умолчанию не выполняются условия о «параллельных линиях» (proportional odds assumption), и в качестве одной из переменных берется прокси переменная для «истинного здоровья».

На основе построенного объективного индекса здоровья в эконометрической модели возможно определение факторов, которые влияют только на самооценку здоровья, при этом не влияя на истинное здоровье, т. е. возможен анализ факторов гетерогенности.

⁹ Эконометрическая обобщенная порядковая пробит модель описана в разделе 6 формулами (8)–(14), уравнения (12)–(14) описывают ограничения на ошибку модели.

Существует несколько методик построения объективного индекса здоровья или, как его называют в работах (Schneider et al., 2012; Pfarr et al., 2012), объективного индекса заболеваемости. Для построения индекса исследователи должны опираться на информацию, доступную для них в используемых данных, и понимать, что в основе индекса должны лежать заболевания, подтвержденные диагнозом медицинских специалистов, т. е. объективные показатели состояния здоровья. Ниже приведена авторская методика построения индекса.

Индекс может быть построен по ряду заболеваний, которые могут быть как диагностированы врачом, так и самодиагностированы пациентом. При этом основанием для включения в составляющие индекса являлась высокая вероятность заявленного заболевания у пациента. Очевидно, что под этот критерий попадали хронические заболевания, диагностируемые врачом. В РМЭЗ к ним относились шесть заболеваний: диабет, инфаркт миокарда, инсульт, анемия, туберкулез и гепатит. Данные об этих заболеваниях собирались каждый год с 2001 по 2015 г. Однако авторы также стремились учитывать заболевания, которые не входили в перечень диагностируемых медицинским специалистом, но самодиагностика которых обладает высокой верифицируемостью. По мнению экспертов, к таким заболеваниям из 16 хронических, перечисленных в РМЭЗ, относятся онкологические заболевания, аллергия и болезни суставов. Данные об этих заболеваниях доступны для периода 2012–2015 гг.

С учетом вышесказанного, объективный индекс заболеваемости строится следующим образом.

Вариант 1 (оригинальный индекс). В панельной пробит регрессии со случайными эффектами¹⁰ для 2001–2015 гг. в качестве независимых переменных X используются шесть диагностируемых врачом заболеваний у респондентов. Зависимой переменной Y в данной регрессии является бинарная переменная *prob*, которая равна 1, если у индивида были проблемы со здоровьем в последние 30 дней, и 0 в противном случае. Построение панельной пробит регрессии для переменной «проблемы со здоровьем» *prob* отдельно для женщин и для мужчин позволяет определить веса каждого из шести заболеваний в объективном индексе заболеваемости. При этом в качестве объективного индекса используются предсказанные значения $\hat{Y}^{11}$, которые масштабируются от 0 до 100 умножением значений на 100. Отдельно рассчитываются индексы для мужчин (*ind5m*) и женщин (*ind5f*). Индекс интерпретируется следующим образом: чем выше значение индекса, тем выше вероятность проблем со здоровьем.

Вариант 2 (широкой индекс). Процедура расчета индекса аналогична процедуре в варианте 1, но число заболеваний расширяется до 9, с тремя дополнительными самодиагностируемыми заболеваниями (онкология, аллергия, болезни суставов), а период времени сокращается до 2012–2015 гг. Отдельно рассчитываются индексы для мужчин (*ind6m*) и женщин (*ind6f*).

Описательные статистики болезней, а также регрессии, на основании которых построены оригинальные и широкие индексы для мужчин и женщин, предоставляются авторами по запросу.

¹⁰ Используется команда *xtprobit* в Stata 13, ошибки v_i в оцениваемом уравнении распределены как $N(0, \sigma_v^2)$.

¹¹ На основе команды *pu0* в Stata13.

6. Эконометрическая модель

В предыдущих разделах описана обобщенная порядковая пробит модель для оценки факторов гетерогенности в самооценке здоровья. Определены виды гетерогенности и обсуждена проблема идентификации, заключающаяся в невозможности выявить степень влияния фактора на истинное здоровье и на его самооценку при параллельном сдвиге точек разграничения. В разделе 5 для решения проблемы идентификации авторы предложили использовать в качестве прокси переменной сформулированный объективный индекс заболеваемости. Объективный индекс добавляется в множество независимых переменных регрессии и в дальнейшем описании модели не выделяется отдельно в формулах.

Для панельных данных Pfarr et al. (2011) разработали обобщенную порядковую пробит модель со случайными эффектами¹². Для представления панельной записи модели, в которой стандартными обозначениями являются i (индивид) и t (время), изменим обозначения категорий самооценки здоровья с i на j . Теперь зависимая переменная «самооценка здоровья» имеет три категории $j = 1, 2, 3$, а уравнение панельной регрессии записывается следующим образом:

$$y_{it}^* = x'_{it}\beta + \varepsilon_{it}, \quad (8)$$

$$\varepsilon_{it} = u_{it} + \alpha_i, \quad (9)$$

$$y_{it} = j \Leftrightarrow \tilde{c}_{j-1} + x'_{it}\gamma_{j-1} \leq y_{it}^* \leq \tilde{c}_j + x'_{it}\gamma_j, \quad j = 1, 2, 3, \quad (10)$$

$$c_j \leq c_{j+1} \quad \forall j, \quad c_0 = -\infty, \quad c_3 = +\infty, \quad (11)$$

$$E[\varepsilon_{it}] = 0, \quad (12)$$

$$\text{Var}[\varepsilon_{it}] = 1 + \sigma_\alpha^2, \quad (13)$$

$$\text{Corr}[\varepsilon_{it}, \varepsilon_{is}] = \rho = \frac{\sigma_\alpha^2}{1 + \sigma_\alpha^2}. \quad (14)$$

Как видно из формул (8)–(14), в панели, в отличие от кросс-секционной записи, оцениваемые вероятности будут включать индивидуальный эффект α_i . Использование панельных данных позволяет учесть два типа гетерогенности. Во-первых, это индивидуальная гетерогенность (индивидуальные ошибки, влияющие на самооценку здоровья), учет которой происходит при использовании спецификации со случайными эффектами обобщенной порядковой пробит регрессии. Во-вторых, это наблюдаемая гетерогенность, которая позволяет выявить сами факторы, влияющие на самооценку здоровья и оценки коэффициентов β и тип гетерогенности.

Обобщенная порядковая пробит модель эквивалентна $j - 1$ бинарным моделям пробит, где j — число категорий зависимой переменной «самооценка здоровья». Первая модель оценивает категорию 1 против категорий 2, ..., j ; вторая модель оценивает категории 1 и 2 против 3, ..., j ; модель $j - 1$ сравнивает категории 1, 2, ..., $j - 1$ и j . В случае для зависимой переменной «самооценка здоровья» имеем две новые категории для обобщенной порядковой

¹² Модуль *regoprob2* в пакете Stata 13, который основан на более ранней команде *regoprob* (Boes, 2006).

пробит модели. Первая категория — «среднее и плохое здоровье по сравнению с хорошим здоровьем», вторая — «плохое здоровье по сравнению с хорошим и средним здоровьем». Это аналогично оценке двух бинарных пробит моделей. В первой пробит модели зависимая переменная Y^1 равна 0, если индивид оценивает свое здоровье как хорошее, и 1 — в случае оценки здоровья как плохого или среднего. Во второй пробит модели зависимая переменная Y^2 принимает значения 0 в случае самооценки здоровья как среднего и хорошего и 1 — если индивид считает свое здоровье плохим.

Модель является итеративной. На первом шаге оценивается модель без ограничений, позволяющая варьировать все коэффициенты модели. На следующем шаге тест Вальда для каждой переменной оценивает, являются ли коэффициенты β различными для различных уравнений или нет. Для значимой переменной с наименьшим значением теста Вальда коэффициенты для двух категорий «среднее и плохое здоровье» и «плохое здоровье по сравнению с хорошим и средним здоровьем» при переменных считаются одинаковым для обеих категорий. Модель снова оценивается с этим новым ограничением. Процедура повторяется до тех пор, пока не определяются все коэффициенты β , являющиеся равными для всех категорий. Процедура завершается, когда тест Вальда на всю модель подтверждает выполнение нулевой гипотезы о параллельных линиях.

Описание панельной регрессии обобщенной порядковой пробит модели со случайными эффектами позволяет определить факторы, вносящие гетерогенность в самооценку здоровья. После включения в набор переменных объективного индекса заболеваемости значимость коэффициента при независимой переменной будет указывать на то, что этот фактор вносит гетерогенность в самооценку здоровья. В случае если для независимой переменной будет два коэффициента, это указывает на сдвиг пороговых значений, в противном случае — на сдвиг индекса. При анализе детерминант гетерогенности сосредоточимся на факторах, которые вносят гетерогенность типа сдвига пороговых значений, поскольку через непараллельные сдвиги они сильнее искажают самооценку здоровья относительно истинного здоровья.

7. Результаты и обсуждение

Таблицы 3 и 4 содержат результаты вычислений отдельно для мужчин и женщин¹³. Представлены две спецификации, отличающиеся объективным индексом заболеваемости. В спецификации 1 используется оригинальный индекс и период 2001–2015 гг. (ind5), в спецификации 2 — широкой индекс для 2012–2015 гг. (ind6). Таблица 3 содержит две спецификации для мужчин, а табл. 4 — для женщин. Для каждой спецификации представлены 2 столбца: первый (категория 1) содержит оценки для среднего и плохого здоровья относительно хорошего здоровья, второй столбец (категория 2) содержит оценки для плохого здоровья относительно среднего и хорошего. В том случае, если коэффициенты для первой и второй категорий равны, они представлены только в первом столбце.

Дополнительно для спецификаций 1 и 2 были рассчитаны средние предельные эффекты¹⁴. Они вычислены для каждой из трех категорий зависимой переменной «самооценка здоровья», т. е. представляют собой предельный эффект изменения независимой

¹³ Все расчеты проводились в Stata 13.

¹⁴ Предоставляются авторами по запросу.

Таблица 3. Обобщенная порядковая пробит регрессия со случайными эффектами, РМЭЗ для мужчин, для оригинального (спецификация 1) и широкого индекса (спецификация 2), зависимая переменная — самооценка здоровья

	Спецификация 1		Спецификация 2	
	Категория 1	Категория 2	Категория 1	Категория 2
Возраст	0.052*** (0.001)		0.052*** (0.001)	
Холост / не замужем	0.020 (0.028)	0.417*** (0.044)	0.037 (0.052)	0.615*** (0.084)
Разведен(а)	-0.064* (0.038)	0.254*** (0.050)	0.132* (0.076)	0.437*** (0.097)
Вдовец(ва)	-0.051 (0.063)	0.109** (0.053)	0.010 (0.093)	
Неполное среднее образование	0.264*** (0.032)	0.425*** (0.036)	0.408*** (0.051)	
Среднее образование	0.201*** (0.027)		0.222*** (0.045)	
Среднее профессиональное образование	0.139*** (0.028)		0.182*** (0.049)	
Объективный индекс заболеваемости (ind5m)	0.032*** (0.001)	0.029*** (0.001)	0.034*** (0.001)	0.028*** (0.001)
Логарифм реального дохода	-0.096*** (0.009)	-0.210*** (0.014)	-0.102*** (0.020)	-0.256*** (0.031)
Поселок городского типа (ПГТ)	-0.334*** (0.049)	-0.177** (0.061)	-0.464* (0.077)	-0.063 (0.102)
Село, деревня	-0.086** (0.029)	-0.167*** (0.037)	-0.154*** (0.042)	
Северный и Северо-Западный экономический район (ЭР)	0.271*** (0.052)		0.203** (0.085)	
Центральный	0.132** (0.040)		0.227*** (0.064)	
и Центрально-Черноземный ЭР	0.088** (0.042)		0.068 (0.067)	
Поволжский и Волго-Вятский ЭР	-0.204*** (0.045)	0.031 (0.053)	-0.276*** (0.071)	
Северо-Кавказский ЭР	0.074* (0.042)		0.141** (0.068)	
Уральский ЭР	0.216*** (0.050)	0.102* (0.058)	0.187** (0.081)	-0.052 (0.103)
Западно-Сибирский ЭР	0.215*** (0.049)	0.085* (0.058)	0.098 (0.075)	
Восточно-Сибирский и Дальневосточный ЭР	да			
Фиктивные переменные по годам	-1.879*** (0.097)	-3.820 (0.129)	-1.980*** (0.201)	-3.787*** (0.290)
Constant	0.519 (0.006)		0.596 (0.009)	
N	63539		23398	
Число групп	15499		9235	

Примечание. ***, **, * означает значимость на уровне 1, 5 и 10% соответственно; в скобках — стандартное отклонение. Во второй и четвертой колонках коэффициенты приводятся только в том случае, если коэффициенты по категориям различны.

Таблица 4. Обобщенная порядковая пробит регрессия со случайными эффектами, РМЭЗ для женщин, для оригинального (спецификация 1) и широкого индекса (спецификация 2), зависимая переменная — самооценка здоровья

	Спецификация 1		Спецификация 2	
	Категория 1	Категория 2	Категория 1	Категория 2
Возраст	0.056*** (0.001)	0.053*** (0.001)	0.053*** (0.001)	0.048** (0.002)
Холост / не замужем	-0.033 (0.025)	0.291*** (0.037)	-0.067 (0.047)	0.294** (0.075)
Разведен(а)	-0.022 (0.026)	0.149*** (0.031)	-0.003 (0.047)	0.154** (0.057)
Вдовец(ва)	0.076 (0.031)	0.233*** (0.026)	0.091 (0.057)	0.290*** (0.049)
Неполное среднее образование	0.252*** (0.030)	0.547 (0.031)	0.302*** (0.052)	0.698*** (0.056)
Среднее образование	0.236*** (0.023)		0.193*** (0.042)	0.314*** (0.052)
Среднее профессиональное образование	0.135*** (0.022)		0.191*** (0.038)	
Объективный индекс заболеваемости (ind5f)	0.025*** (0.001)		0.030*** (0.001)	0.025*** (0.001)
Логарифм реального дохода	-0.036*** (0.008)	-0.125*** (0.012)	-0.047** (0.016)	-0.157*** (0.025)
ПГТ	-0.415*** (0.047)	-0.061 (0.053)	-0.495*** (0.071)	0.018 (0.083)
Село, деревня	-0.105*** (0.027)	0.049 (0.030)	-0.093** (0.042)	0.066 (0.049)
Северный и Северо-Западный ЭР	0.378*** (0.049)		0.364*** (0.070)	
Центральный и Центрально-Черноземный ЭР	0.289*** (0.039)	0.212*** (0.041)	0.324*** (0.054)	
Поволжский и Волго-Вятский ЭР	0.285*** (0.041)	0.091** (0.043)	0.342*** (0.062)	-0.051 (0.068)
Северо-Кавказский ЭР	-0.099** (0.042)	0.010 (0.047)	-0.140** (0.061)	
Уральский ЭР	0.186*** (0.041)	0.102** (0.045)	0.263*** (0.062)	0.042 (0.070)
Западно-Сибирский ЭР	0.458*** (0.047)	0.157** (0.050)	0.404*** (0.072)	0.022 (0.082)
Восточно-Сибирский и Дальневосточный ЭР	0.395*** (0.046)	0.117** (0.051)	0.308*** (0.071)	-0.131 (0.083)
Фиктивные переменные по годам	да			
Константа	-2.105*** (0.078)	-4.616*** (0.105)	-2.483** (0.151)	-4.659 (0.226)
Rho	0.533 (0.005)		0.607 (0.008)	
N	98065		32978	
Число групп	19399		12538	

Примечание. ***, ** — значимость на уровне 1 и 5% соответственно; в скобках — стандартное отклонение. Во второй и четвертой колонках коэффициенты приводятся только в том случае, если коэффициенты по категориям различны.

переменной на вероятность ответа «хорошее здоровье», «среднее здоровье» и «плохое здоровье».

Из таблиц расчетов предельных эффектов можно сделать выводы о том, как изменения в истинном здоровье отражаются на изменениях в самооценке здоровья, и как эти результаты различаются для мужчин и для женщин. Предельные эффекты от изменения первоначального индекса здоровья на самооценку здоровья для всех категорий мужчин больше, чем для женщин. А тот факт, что разброс предельных эффектов по изменению в истинном здоровье для мужчин больше, чем для женщин, свидетельствует о том, что самооценка здоровья мужчин более чувствительна к изменению в истинном здоровье.

Жирным шрифтом в таблицах 3 и 4 выделены переменные, для которых нарушается предположение о «параллельных линиях» одновременно для спецификаций 1 и 2, а жирным курсивом — только для одной из спецификаций 1 или 2. Именно для этих переменных наблюдается гетерогенность второго типа — сдвиг точек разграничений. Про остальные значимые факторы можно сказать, что они вносят гетерогенность первого типа.

Анализ индекса заболеваемости показывает, что для мужчин с более высоким объективным индексом заболеваемости, т. е. имеющих более слабое истинное здоровье, в обеих спецификациях вероятность выбора среднего и плохого здоровья возрастает больше, нежели вероятность плохого здоровья (табл. 3). Для женщин в спецификации 1 с оригинальным индексом (ind5f) последний оказывает одинаковое влияние на категории 1 и 2, тогда как в спецификации 2 эффект истинного здоровья на самооценку женщин аналогичен эффекту на самооценку здоровья у мужчин в спецификации 1. Подчеркнем, что изменения по категориям для переменной «объективный индекс заболеваемости» не связаны с гетерогенностью, т. к. это те изменения в самооценке, которые не связаны с изменениями в истинном здоровье.

Для мужчин общими детерминантами наблюдаемой гетерогенности типа сдвига пороговых значений (т. е. переменными, для которых представлены значимые различные коэффициенты по категориям) в спецификациях с оригинальным индексом (ind5) и широким индексом (ind6) являются *категории «разведен» и «холост», доход, проживание в сельской местности и проживание в Западно-Сибирском и Дальневосточном районах.*

Статус «разведен» для мужчин увеличивает вероятность хорошей и плохой самооценки и уменьшает вероятность средней самооценки здоровья. Предельные эффекты составляют 0.014 для хорошего здоровья, -0.038 для среднего здоровья и 0.024 для плохого здоровья. Данный результат можно интерпретировать следующим образом. Расторжение брака способно как улучшить, так и ухудшить самочувствие, вторая ситуация возникает, например, когда развод связан с потерей партнера, способного позаботиться о супруге в случае его болезни, а также оказывать моральную поддержку на протяжении жизни (Синельников, 2011). В спецификации 2 в случае широкого индекса здоровья развод оказывает отрицательное значимое влияние на здоровье для обеих категорий.

Эконометрические расчеты показывают, что люди с неполным средним образованием занижают самооценку здоровья по сравнению с лицами, имеющими высшее образование (табл. 3). Общеизвестно, что уровень образования является прокси здоровья (Grossman, 1972), а более низкий уровень образования связан с недостаточным уровнем знания о течении болезней, их профилактике и лечении. Помимо того, что низкий уровень образования приводит к объективному снижению капитала здоровья, он приносит и психологический дискомфорт, связанный со снижением самооценки. Расчет предельных эффектов указывает

на снижение вероятности хорошего здоровья (-0.058) и рост вероятностей среднего (0.019) и плохого здоровья (0.040).

Доход положительно влияет на самооценку здоровья. Предельный эффект логарифма реального дохода положителен для хорошего здоровья (0.021 в спецификации 1 и 0.020 в спецификации 2 для мужчин) и отрицателен для плохого здоровья (-0.020 в спецификации 1 и -0.019 в спецификации 2 для мужчин), что означает, что рост дохода увеличивает вероятность самооценки здоровья как хорошего и снижает вероятность самооценки здоровья как плохого респондентами мужского пола. Возможно, респонденты связывают рост дохода с возможностью направлять дополнительные средства на лечение и медицинские препараты.

Эффект проживания в поселках городского типа и в сельской местности аналогичен эффекту роста дохода, однако оказывает значимый эффект на мужчин только в спецификации 1. Предельный эффект от проживания в селе/деревне на самооценку здоровья составляет 0.019 для хорошего здоровья и -0.016 для плохого здоровья. Сельская местность и поселки городского типа улучшают самооценку здоровья, что может быть вызвано в целом более благоприятной окружающей средой, темпом жизни, заставляющим постоянно двигаться, а также меньшей психологической нагрузкой.

Базовым регионом в исследовании являются города Москва и Санкт-Петербург. Жители всех остальных регионов, за исключением Северного Кавказа, занижают самооценку здоровья. Проживание в отдельных регионах также характеризуется сдвигом пороговых значений. Для спецификации 1 такими регионами являются: Северо-Кавказский, Западно-Сибирский, Восточно-Сибирский и Дальневосточный районы. Для жителей Северного Кавказа повышается вероятность оценить свое здоровье как хорошее. В целом для всех регионов, кроме Северного Кавказа, предельный эффект снижения вероятности самооценки здоровья как «хорошего» выше, чем предельный эффект роста вероятности оценки здоровья как плохого. Причины занижения самооценки здоровья могут быть различными и связанными как с более суровыми климатическими условиями (в Западно-Сибирском и Дальневосточном регионах), так и более низким уровнем доходов и качеством жизни по сравнению с Москвой и Санкт-Петербургом (Lyytikäinen, Kemppainen, 2016).

Отдельно стоит отметить постепенно возрастающий положительный эффект от фиктивных переменных для годов. Для мужчин в спецификации 1 данный эффект сильнее выражен для группы респондентов с плохим и средним здоровьем относительно респондентов с хорошим здоровьем (категория 1) после кризиса 2008–2009 гг., а именно, с 2011 по 2015 г. Эта тенденция хорошо иллюстрируется ростом предельных эффектов для вероятности самооценки здоровья как хорошего с 2011 по 2014 г. Данный эффект может быть связан с относительным ростом благосостояния страны, когда наблюдаемые экономические улучшения положительно сказываются на психологическом состоянии граждан РФ.

Для женщин в спецификации 1 для объективного индекса заболеваемости рассчитан один коэффициент, в то время как в спецификации 2 для более широкого индекса рассчитаны два коэффициента. С учетом того, что широкий индекс призван точнее учесть объективное состояние здоровья, авторы основывают интерпретацию детерминант наблюдаемой гетерогенности на спецификации 2. В этой спецификации детерминантами наблюдаемой гетерогенности вида сдвига пороговых значений для женщин являются *все три категории переменной «семейное положение» (холост / не замужем, разведен (а), вдовец (ва)), неполное среднее и среднее образование, доход, проживание за пределами города и в некоторых*

районах РФ (а именно, в Волго-Вятском, Уральском, Западно-Сибирском, Восточно-Сибирском и Дальневосточном). Рассмотрим эти факторы подробнее.

У женщин различие между ненаблюдаемым истинным здоровьем и самооценкой здоровья увеличивается с возрастом. При этом занижение самооценки здоровья с возрастом менее выражено для плохого здоровья относительно среднего и хорошего, нежели для плохого и среднего здоровья относительно хорошего (табл. 4).

Для одиноких и разведенных женщин коэффициент для категории 2 в спецификации 2 был значим и положителен, а коэффициенты для категории 1 были статистически незначимы и отрицательны (табл. 4). Анализ предельных эффектов позволил сделать вывод о том, что статусы «не замужем» и «разведена» увеличивают вероятность оценки здоровья и как плохого, и как хорошего, снижая вероятность оценки выбора в качестве ответа среднего здоровья, однако предельный эффект от роста вероятности плохого здоровья в 4 раза выше предельного эффекта от роста вероятности хорошего здоровья. Это указывает на то, что развод и вызванная им в ряде случаев депрессия (Monden, Uunk, 2013) отрицательно сказываются на самооценке здоровья женщин. Статус «вдова», характерный, как правило, для более старших респондентов, связан с ухудшением здоровья для обеих категорий, что подтверждают более ранние эмпирические исследования по России (Kaneva et al., 2018).

Для женщин с неполным средним и средним образованием растут вероятности оценить свое здоровье как «плохое или среднее» и вероятности оценить свое здоровье как «плохое», этот эффект более выражен для второй категории. Расчет предельных эффектов указывает на наличие градиента здоровья в образовании, поскольку предельный эффект для оценки здоровья как плохого выше для лиц с более низким уровнем образования (0.076 для неполного среднего и 0.034 для среднего).

Влияние дохода на самооценку у женщин аналогично влиянию дохода у мужчин, описанному выше, однако положительный эффект дохода более выражен для мужчин, поскольку их участие на рынке труда РФ выше. Для женщин эффект выражен слабее также и потому, что значительная доля респондентов состоит в браке и опирается на финансовую поддержку супругов.

Проживание в сельской местности значимо и положительно влияет на вероятность выбора категории «хорошее», в то время как эффект на категорию «плохое» здоровье является статистически незначимым, в отличие от мужчин, для которых самооценка здоровья повышается всегда. Расчет предельных эффектов показывает, что проживание в селе или деревне увеличивает вероятность выбора хорошего здоровья на 0.015, снижая вероятность выбора среднего здоровья на 0.022, а предельный эффект для категории плохое здоровье не является статистически значимым. Проживание в Волго-Вятском, Уральском, Западно-Сибирском, Восточно-Сибирском и Дальневосточном районах снижает самооценку здоровья женщин для категории 1 (плохое и среднее здоровье vs. хорошее здоровье) относительно респондентов из Москвы и Санкт-Петербурга.

Для женщин наблюдается рост самооценки здоровья для всех лет по сравнению с базовым годом в спецификации 1 и для 2014 и 2015 гг. в спецификации 2, при этом эффект улучшения самооценки возрастает с течением времени и максимален для 2015 г. в обеих спецификациях. Он также выше для первой категории респондентов по сравнению со второй.

Основной вывод из анализа детерминант наблюдаемой гетерогенности у мужчин и женщин заключается в том, что так же как и в работе (Schneider et al., 2012), набор переменных, определяющих гетерогенность типа сдвига пороговых значений в самооценке здоровья,

отличается для мужчин и для женщин. Среднее образование является детерминантой здоровья у женщин, но не у мужчин, а набор регионов, вносящих гетерогенность типа сдвига пороговых значений, различается для мужчин и женщин.

Для проверки робастности результатов авторы проверили расчеты по кросс-секционной модели для 2015 г. для мужчин и женщин для оригинального и широкого индексов¹⁵. Поскольку кросс-секционного аналога для обобщенной порядковой пробит регрессии со случайными эффектами не существует, рассчитывается обобщенная порядковая логит регрессия (*gologit2*), подробное описание которой можно найти в работах (Williams, 2006, 2016). Полученные результаты близки к спецификациям 1 и 2 и однозначно указывают на сдвиг пороговых значений для мужчин и женщин для обоих индексов. Отличиями от панельных расчетов является то, что для мужчин и оригинального индекса на разброс самооценок здоровья влияет возраст, но данный результат не подтверждается для широкого индекса. Также для мужчин изменяются региональные детерминанты гетерогенности типа сдвига пороговых значений. Для женщин, наоборот, возраст перестает влиять на разницу между истинным здоровьем и самооценкой здоровья, а семейное положение перестает быть детерминантой гетерогенности. Эти эффекты проявляются только в 2015 г. Остальные результаты панельных регрессий остаются справедливыми и в случае кросс-секционного анализа.

В качестве ограничения настоящего исследования необходимо отметить, что выводы о гетерогенности здоровья по тем или иным детерминантам могут зависеть от выбора инструментов для измерения объективного здоровья. Действительно, возможно построение различных индексов заболеваемости (они представлены, в частности, в Jürges (2006), Pfarr et al. (2012), Schneider et al. (2012)), и каждый из них аппроксимирует истинное здоровье по-разному, что приводит к разным коэффициентам в моделях обобщенного порядкового пробита со случайными эффектами, а также к разным результатам относительно равенства коэффициентов по категориям.

8. Заключение

Настоящая статья является первым исследованием по Российской Федерации, в котором анализируется феномен гетерогенности самооценки здоровья для мужчин и женщин. В качестве детерминант наблюдаемой гетерогенности типа сдвига пороговых значений для обоих полов авторами выделены следующие социально-экономические характеристики — статусы «разведен» и «холост» для мужчин, статусы «разведена», «не замужем» и «вдова» для женщин, доход, наличие неполного среднего образования, проживание за пределами города и в отдельных регионах. Для женщин с возрастом возрастает разница между истинным здоровьем и его самооценкой. В спецификации с периодами с 2001 по 2015 г. к детерминантам наблюдаемой гетерогенности второго типа также относятся временные периоды: для мужчин — 2009 и 2011–2015 гг., а для женщин — все годы, кроме 2002, 2003, 2005 и 2006 гг.

Выявление детерминант наблюдаемой гетерогенности позволило дать количественное подтверждение некоторым общеизвестным фактам. Так, например, регрессионный анализ показал, что мужчины на Северном Кавказе сообщают о более высокой самооценке

¹⁵ Таблица с результатами расчетов предоставляется авторами по запросу.

здоровья. Еще одним фактом является подтверждение существования образовательного градиента здоровья.

Полученные в исследовании результаты показывают, что прямые сравнения самооценок здоровья для отдельных социально-экономических групп и индивидов с разными наборами заболеваний в России невозможны. Любому сравнительному анализу самооценки здоровья россиян должны предшествовать исследования по выявлению детерминант гетерогенности.

Без тщательного анализа вопросов гетерогенности невозможны и международные сравнения самооценок жителей других стран с самооценкой россиян. Так, в 2014 г. Евростат представил информацию о самооценке здоровья респондентов в возрасте от 16 лет в 28 странах ЕС и шести других странах Европы (Статистика Евростата..., 2014). Аналитики Евростата отметили, что для жителей России характерна оценка своего здоровья как среднего, особенно среди женщин (53%). В то же время доля жителей России, оценивших свое здоровье как хорошее или очень хорошее, значительно меньше, чем аналогичный процент людей в развитых странах Европы. Причины низкого процента ответов россиян в категориях «хорошее» и «очень хорошее» здоровье могут быть напрямую связаны с гетерогенностью самооценки. Сравнение детерминант гетерогенности жителей ЕС и России может стать темой дальнейшего направления исследований. Еще одним направлением является включение в набор факторов, влияющих на самооценку здоровья, макроэкономических показателей, а именно ВРП на душу населения.

Результаты настоящего исследования могут использоваться для оценки уровней здоровья и заболеваемости в рамках разработки политики по оздоровлению населения и определения целей стратегий развития здравоохранения региональными и федеральными властями.

Благодарности. Авторы благодарят двух анонимных рецензентов и редактора журнала за помощь и ценные комментарии в процессе написания статьи.

Список литературы

Герри К. Д., Байдин В. М. (2017). Источники искажений в самооценке здоровья. *Фармаэкономика*, 10 (4), 31–36.

Канева М. А. (2016). Социально-экономические, поведенческие и психологические детерминанты самооценки здоровья россиян. *Национальные интересы: приоритеты и безопасность*, 6, 158–171.

Сизова И. Л., Леонова Л. А., Хензе А. (2017). Прекаритет занятости и доходов в России и Германии: самовосприятие наемными работниками. *Экономическая социология*, 18 (4), 14–59.

Синельников А. Б. (2011). Влияние брачного статуса на самооценку здоровья населения. *Социальные аспекты здоровья населения*, 20 (4), 9.

Статистика Евростата по самооценке здоровья. (2014). http://ec.europa.eu/eurostat/statistics-explained/index.php/Self-perceived_health_statistics.

Au N., Johnston D. W. (2014). Self-assessed health: What does it mean and what does it hide? *Social Science and Medicine*, 121, 21–28.

Boes S. (2006). GOPROBIT: Stata module to estimate generalized ordered probit models. Statistical Software Components S456603, Boston College Department of Economics, revised 06 Sep 2006.

Bos A. M., Bos A. J. (2007). The socio-economic determinants of older people's health in Brazil: The importance of marital status and income. *Ageing and Society*, 27, 385–405.

Cairney J. (2000). Socio-economic status and self-rated health among older Canadians. *Canadian Journal of Aging*, 19 (4), 456–478.

Eriksson I., Undén A.-L., Elofsson S. (2001). Self-rated health. Comparisons between three different measures. Results from a population study. *International Journal of Epidemiology*, 30, 326–333.

Etilé F., Milcent C. (2006). Income-related reporting heterogeneity in self-assessed health: Evidence from France. *Health Economics*, 15, 965–981.

Gerry C. J., Papadopoulos G. (2015). Sample attrition in RLMS, 2001–2010. Lessons for longitudinal analysis and an application in health. *Economics in Transition*, 23 (2), 425–458.

Grossman M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, 80 (2), 223–255.

Health interview surveys. Towards international harmonization of methods and instruments. (1996). Eds. A. De Bruin, H. S. J. Picavet and A. Nossikov. World Health Organization Publication, Copenhagen.

Hernández-Quevedo C., Jones A. M., Rice N. (2005). Reporting bias and heterogeneity in self-assessed health: Evidence from the British Household Panel Survey. *HEDG Working Paper 05/04*. York: University of York.

Hirve S., Verde E., Lele P., Juvekar S., Blomsted Y., Tollman S., Wall S., Chatterji S., Ng N. (2014). Evaluating reporting heterogeneity in self-rated health among adults aged 50 years and above in India: An anchoring vignettes analytical approach. *Journal of Aging and Health*, 26 (6), 1015–1031.

Jürges H. (2006). True health vs. response styles: Exploring cross-country differences in self-reported health. *Health Economics*, 16, 163–178.

Jylhä M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science and Medicine*, 69, 307–316.

Idler E. L. (1993). Age differences in self-assessments of health: Age changes, cohort differences, or survivorship? *Journal of Gerontology*, 48, S289–S300.

Kaneva M., Gerry C., Baidin V. (2018). The effect of chronic conditions and self-assessed health in Russia. *Scandinavian Journal of Public Health*, 1–11, doi: 10.1177/1403494817746283.

Kerkhofs M., Lindeboom M. (1995). Subjective health measures and state dependent reporting errors. *Health Economics*, 4, 221–235.

Kim Y. (2015). The dynamics of health and its determinants among elderly in developing countries. *Economics and Human Biology*, 19, 1–12.

Kozyreva P., Kosolapov M., Popkin B. M. (2016). Data resource profile: The Russian Longitudinal Monitoring Survey Higher School of Economics (RLMS-HSE) Phase II: Monitoring the Economic and Health Situation in Russia, 1994–2013. *International Journal of Epidemiology*, 45 (2), 395–401.

Lindeboom M., van Doorslaer E. (2004). Cut-point shift and index shift in self-reported health. *Journal of Health Economics*, 23, 1083–1099.

Livneh H., Antonak R. (2005). Psychosocial adaptation to chronic illness and disability: A primer for counselors. *Journal of Counseling and Development*, 83, 12–20.

Lokshin M., Ravallion M. (2008). Testing for economic gradient in health using subjective data. *Health Economics*, 17 (11), 1237–1259.

Lyytikäinen L., Kempainen T. (2016). Regional inequalities in self-rated health in Russia: What is the role of social and economic capital? *Social Science and Medicine*, 161, 92–99.

Monden C. W. S., Uunk W. J. G. (2013). For better and for worse: The relationship between union dissolution and self-assessed health in European panel data. *European Journal of Population*, 29, 103–125.

- Pfarr C., Schmid A., Schneider U. (2011). Estimating ordered categorical variables using panel data: A generalized ordered probit model with autofit procedure. *Journal of Economics and Econometrics*, 54 (1), 7–23.
- Pfarr C., Schmid A., Schneider U. (2012). Reporting heterogeneity in self-assessed health among older Europeans. *Health Economics Review*, 2, 21.
- Schneider U., Pfarr C., Shneider B. S., Ulrich V. (2012). I feel good! Gender differences and reporting heterogeneity in self-assessed health. *European Journal of Health Economics*, 13, 261–265.
- Shmueli A. (2003). Socio-economic and demographic variation in health and its measures: The issue of reporting heterogeneity. *Social Science and Medicine*, 57, 125–134.
- Simon J. G., van de Mheen H., van der Meer J. B. W., Mackenbach J. P. (2000). Socioeconomic differences in self-assessed health in a chronically ill population: The role of different health aspects. *Journal of Behavioral Medicine*, 23 (5), 399–420.
- Williams R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variable. *Stata Journal*, 6, 58–82.
- Williams R. (2016). Understanding and interpreting generalized ordered logit models. *The Journal of Mathematical Sociology*, 40 (1), 7–20.

Поступила в редакцию 27.07.2017;
принята в печать 17.07.2018.

Kaneva M., Baidin V. Heterogeneity in reporting self-assessed health of the Russians. *Applied Econometrics*, 2018, v. 51, pp. 102–125.

Maria Kaneva

Gaidar Institute for Economic Policy, Moscow, Russian Federation;
mkaneva@gmail.com

Valerii Baidin

Financial research institute of the Ministry of finance of the Russian Federation,
Moscow, Russian Federation;
vbaydin@nifi.ru

Heterogeneity in reporting self-assessed health of the Russians

The article aims to analyze a phenomenon of heterogeneity in self-assessed health for the Russian population from 2001 to 2015, to determine types of heterogeneity and to elicit socio-economic and health-rated determinants of heterogeneity. A generalized ordered probit model with random effects and objective diseases index indicated that the determinants of heterogeneity in the Russian Federation are for both genders: marital status, income, incomplete secondary education, living outside urban areas and region of settlement; age also affects the differences in self-reported health for women. Secondary education is also a driver of observed heterogeneity for women while different regions drive cut-point shift heterogeneity for males and females. The study confirms that direct comparison of self-assessed health for different population groups will lead to incorrect conclusions.

Keywords: heterogeneity; self-assessed health; panel regression; RLMS; Russia.

JEL classification: I12; C21; C23.

References

- Gerry C. J., Baidin V. M. (2018). Istochniki iskazhenij v samootsenke zdorov'ya. *Pharmaekonomika*, 10 (4), 31–36 (in Russian).
- Kaneva M. A. (2016). Sotsial'no-ehkonomicheskie, povedencheskie i psikhologicheskie determinanty samootsenki zdorov'ya rossiyan. *Natsional'nye interesy: priority i bezopasnost*, 6, 158–171 (in Russian).
- Sizova I. L., Leonova L. A., KHenze A. (2017). Prekaritet zanyatosti i dokhodov v Rossii i Germanii: samovospriyatie naemnymi rabotnikami. *Ekonomicheskaya sotsiologiya*, 18 (4), 14–59 (in Russian).
- Sinel'nikov A. B. (2011). Vliyanie brachnogo statusa na samootsenku zdorov'ya naseleniya. *Sotsial'nye aspekty zdorov'ya naseleniya*, 20 (4), 9 (in Russian).
- Statistics on self-assessed health of Eurostat. (2014). http://ec.europa.eu/eurostat/statistics-explained/index.php/Self-perceived_health_statistics.
- Au N., Johnston D. W. (2014). Self-assessed health: What does it mean and what does it hide? *Social Science and Medicine*, 121, 21–28.
- Boes S. (2006). GOPROBIT: Stata module to estimate generalized ordered probit models. Statistical Software Components S456603, Boston College Department of Economics, revised 06 Sep 2006.
- Bos A. M., Bos A. J. (2007). The socio-economic determinants of older people's health in Brazil: The importance of marital status and income. *Ageing and Society*, 27, 385–405.
- Cairney J. (2000). Socio-economic status and self-rated health among older Canadians. *Canadian Journal of Aging*, 19 (4), 456–478.
- Eriksson I., Undén A.-L., Elofsson S. (2001). Self-rated health. Comparisons between three different measures. Results from a population study. *International Journal of Epidemiology*, 30, 326–333.
- Etilé F., Milcent C. (2006). Income-related reporting heterogeneity in self-assessed health: Evidence from France. *Health Economics*, 15, 965–981.
- Gerry C. J., Papadopoulos G. (2015). Sample attrition in RLMS, 2001–2010. Lessons for longitudinal analysis and an application in health. *Economics in Transition*, 23 (2), 425–458.
- Grossman M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, 80 (2), 223–255.
- Health interview surveys. Towards international harmonization of methods and instruments. (1996). Eds. A. De Bruin, H. S. J. Picavet and A. Nossikov. World Health Organization Publication, Copenhagen.
- Hernández-Quevedo C., Jones A. M., Rice N. (2005). Reporting bias and heterogeneity in self-assessed health: Evidence from the British Household Panel Survey. *HEDG Working Paper* 05/04. York: University of York.
- Hirve S., Verde E., Lele P., Juvekar S., Blomsted Y., Tollman S., Wall S., Chatterji S., Ng N. (2014). Evaluating reporting heterogeneity in self-rated health among adults aged 50 years and above in India: An anchoring vignettes analytical approach. *Journal of Aging and Health*, 26 (6), 1015–1031.
- Jürges H. (2006). True health vs. response styles: Exploring cross-country differences in self-reported health. *Health Economics*, 16, 163–178.
- Jylhä M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science and Medicine*, 69, 307–316.
- Idler E. L. (1993). Age differences in self-assessments of health: Age changes, cohort differences, or survivorship? *Journal of Gerontology*, 48, S289–S300.

- Kaneva M., Gerry C., Baidin V. (2018). The effect of chronic conditions and self-assessed health in Russia. *Scandinavian Journal of Public Health*, 1–11, doi: 10.1177/1403494817746283.
- Kerkhofs M., Lindeboom M. (1995). Subjective health measures and state dependent reporting errors. *Health Economics*, 4, 221–235.
- Kim Y. (2015). The dynamics of health and its determinants among elderly in developing countries. *Economics and Human Biology*, 19, 1–12.
- Kozyreva P., Kosolapov M., Popkin B. M. (2016). Data resource profile: The Russian Longitudinal Monitoring Survey Higher School of Economics (RLMS-HSE) Phase II: Monitoring the Economic and Health Situation in Russia, 1994–2013. *International Journal of Epidemiology*, 45 (2), 395–401.
- Lindeboom M., van Doorslaer E. (2004). Cut-point shift and index shift in self-reported health. *Journal of Health Economics*, 23, 1083–1099.
- Livneh H., Antonak R. (2005). Psychosocial adaptation to chronic illness and disability: A primer for counselors. *Journal of Counseling and Development*, 83, 12–20.
- Lokshin M., Ravallion M. (2008). Testing for economic gradient in health using subjective data. *Health Economics*, 17 (11), 1237–1259.
- Lyytikäinen L., Kemppainen T. (2016). Regional inequalities in self-rated health in Russia: What is the role of social and economic capital? *Social Science and Medicine*, 161, 92–99.
- Monden C. W. S., Uunk W. J. G. (2013). For better and for worse: The relationship between union dissolution and self-assessed health in European panel data. *European Journal of Population*, 29, 103–125.
- Pfarr C., Schmid A., Schneider U. (2011). Estimating ordered categorical variables using panel data: A generalized ordered probit model with autotit procedure. *Journal of Economics and Econometrics*, 54 (1), 7–23.
- Pfarr C., Schmid A., Schneider U. (2012). Reporting heterogeneity in self-assessed health among older Europeans. *Health Economics Review*, 2, 21.
- Schneider U., Pfarr C., Shneider B. S., Ulrich V. (2012). I feel good! Gender differences and reporting heterogeneity in self-assessed health. *European Journal of Health Economics*, 13, 261–265.
- Shmueli A. (2003). Socio-economic and demographic variation in health and its measures: The issue of reporting heterogeneity. *Social Science and Medicine*, 57, 125–134.
- Simon J. G., van de Mheen H., van der Meer J. B. W., Mackenbach J. P. (2000). Socioeconomic differences in self-assessed health in a chronically ill population: The role of different health aspects. *Journal of Behavioral Medicine*, 23 (5), 399–420.
- Williams R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variable. *Stata Journal*, 6, 58–82.
- Williams R. (2016). Understanding and interpreting generalized ordered logit models. *The Journal of Mathematical Sociology*, 40 (1), 7–20.

Received 27.07.2017; accepted 17.07.2018.