

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

УДК 577.73

Анализ продолжительности жизни человека в исторической перспективе

А.И. Михальский^{1,*} , В.Н. Анисимов² , Г.М. Жаринов³ 

¹Институт проблем управления имени В.А. Трапезникова, Российская академия наук,
Россия, 117997, г. Москва, ул. Профсоюзная, д. 65;

²Национальный медицинский исследовательский центр онкологии имени Н.Н. Петрова,
Россия, 197758, г. Санкт-Петербург, пос. Песочный, ул. Ленинградская, д. 68;

³Российский научный центр радиологии и хирургических технологий имени академика А.М. Гранова,
Россия, 197758, г. Санкт-Петербург, пос. Песочный, ул. Ленинградская, д. 70

*e-mail: ipuran@yandex.ru

Рассматривается проблема корректной интерпретации данных о среднем возрасте смерти людей в различные исторические периоды на основании информации, представленной в Википедии и включенной в базу данных «Личности Энциклопедии». Отмечается специфика таких данных, связанная с цензурованием и характером их сбора, приводящим к селективности. Предложен подход, учитывающий эти обстоятельства, что позволяет уточнять получаемые закономерности о продолжительности жизни и расширить исторический диапазон исследования.

Ключевые слова: продолжительность жизни, средний возраст смерти, Википедия, база данных «Личности Энциклопедии», цензурование, селективность, вероятность дожития

Исторический тренд продолжительности жизни человека вызывает пристальный интерес ученых различных специальностей, поскольку на основании этого тренда можно судить о факторах, способствующих долголетию, об эффективности социально-экономических мер по улучшению жизни людей, о влиянии прогресса в медицине. В опубликованных ранее работах [1, 2] на основании данных о странах, имевших наивысшую продолжительность жизни в конкретный исторический период, показан ее неуклонный рост, начиная с середины XIX в. При этом подчеркивается, что ни одна из существующих демографических теорий не предсказывала такого явления, делая ошибочные заключения о «видовом пределе» продолжительности жизни человека [2].

Развитие средств извлечения и хранения информации привело к накоплению больших массивов данных и, в частности, данных о продолжительности жизни человека в различные исторические эпохи. Такая информация концентрировано содержится в различных энциклопедических справочниках, информационных ресурсах, разных разделах Википедии. В отличие от обезличенных статистических данных эти ресурсы позволяют получить персонифицированную информацию, содержащую даты жизни, род занятий, агрегированные причины смерти. Использование такой информации позволяет выделять дополнительные факторы, порой неожиданные, влияю-

щие на продолжительность жизни [3–5], однако обоснованность подобных заключений требует использования методов анализа и сопоставления, учитывающих особенности сбора данных.

В демографии важной характеристикой состояния населения является продолжительность жизни. Эта характеристика отражает влияние на здоровье населения состояния окружающей среды, уровня медицинского обслуживания, социально-экономической ситуации. Различают ожидаемую продолжительность жизни (ОПЖ) при рождении, которую вычисляют по интенсивности смертности, рассчитанной по агрегированным данным о числе смертей в различных возрастных группах, и средний возраст смерти (СВС), который можно вычислить, сопоставляя даты рождения и смерти конкретных людей.

ОПЖ при рождении вычисляется на основании, демографических данных населения по формуле [6]:

$$e_0 = \frac{1}{l_0} \sum_{x=0}^{\omega-1} L_x, \quad (1)$$

где L_x — численность живущих в возрастном интервале от « x » до « $x + 1$ »; e_0 — ОПЖ при рождении; l_0 — число родившихся; $\omega - 1$ — предельный возраст, до которого доживает последний человек из поколения родившихся. Показатель ОПЖ имеет смысл теоретического среднего значения продол-

жительности жизни, рассматриваемой как случайная величина, распределение которой задается величинами интенсивности смертности в различных возрастных группах для людей одного года рождения — когорте рождения. Требование использовать когортные данные является существенным ограничением при вычислении ОПЖ, так как при этом необходима информация о числе рождений и смертей как минимум за 100 предыдущих лет. Подобная информация доступна лишь в редких случаях — например, для Швеции [7].

На практике для вычисления ОПЖ используют не реальные когортные данные, а данные для членов «синтетической когорты», составленной из лиц, рожденных в разные годы. Расчет ОПЖ при рождении производится по формуле:

$$e_0 = \sum_{k=0}^{\infty} \prod_{x=0}^k (1 - q_x), \quad (2)$$

где q_x — доля умерших за год среди лиц возраста x на начало года, k — номер одногодичной возрастной группы. Значения q_x берутся из таблиц смертности [6]. Рассчитанная таким образом оценка ОПЖ близка к величине ОПЖ, рассчитанной по когортным данным, если интенсивность смертности зависит только от возраста, но в каждой возрастной группе постоянна во времени. Такое население называют стабильным и часто используют в теоретических рассуждениях в демографии. В реальности влияние времени на интенсивность смертности в возрастных группах существенно — оно может как увеличиваться, так и уменьшаться. Для учета совокупного влияния возраста и времени используют математические демографические модели [8], которые позволяют учесть влияние времени на интенсивность смертности, но доля субъективизма в выборе модели этого влияния остается.

СВС при рождении $E_0(y)$ в группе из N человек, рожденных в году y , является эмпирической оценкой ОПЖ при рождении и вычисляется по формуле:

$$E_0(y) = \frac{1}{N} \sum_{i=1}^N t_i, \quad (3)$$

где t_i — возраст смерти i -го человека. СВС получается в результате прямых вычислений без использования предварительных допущений о характере распределения возраста смерти. В то же время эта характеристика не учитывает наличие еще живых на данный момент лиц, занижая реальный СВС, который получился бы, если учесть возраст смерти в будущем. При исследовании временных трендов СВС вычисляют для групп лиц одного года рождения, и полученная оценка характеризует когорту рождения. ОПЖ и СВС могут рассчитываться и для лиц, уже достигших возраста x . Тогда говорят

про ОПЖ и СВС в возрасте x . При этом в формулах (1–3) рассматриваются лишь такие люди.

Нами собрана база данных из Википедии и других энциклопедических источников, содержащая более 700 тыс. записей дат рождения и смерти, охватывающая период рождения от I века н.э. до настоящего времени. Понятно, что информация, внесенная в энциклопедические источники, подвергается селекции, отражая не все население, а наиболее выдающихся людей. Тем не менее, большой объем данных позволяет учесть подобные искажения и извлечь полезную информацию об исторических трендах в продолжительности жизни, а также о влиянии на продолжительность жизни широкого круга профессиональных занятий. Статья посвящена обсуждению методологии учета специфики данных, собранных из энциклопедических источников, обсуждению полученных результатов и их сопоставлению с опубликованными данными об исторических трендах в продолжительности жизни людей [1].

Материалы и методы

База данных «Личности Энциклопедии». Наиболее надежными источниками данных о продолжительности жизни являются результаты переписи населения, данные различных регистров [9, 10], результаты специальных долгосрочных обследований [11]. Для обработки и интерпретации таких данных разработаны специальные методы, учитывающие специфику сбора данных (регистрация по обращаемости, опрос, инструментальные наблюдения и т. п.), особенности населения (возрастной, этнический состав, социально-экономические различия), а также цели исследования. Недавно появились демографические базы, содержащие предобработанные данные по фертильности и продолжительности жизни человека почти во всех странах мира (HFD. <http://www.humanfertility.org>, HMD. <http://www.mortality.org>).

В отличие от специализированных баз данных, формируемых под руководством специалистов, запущенная в январе 2001 г. Википедия является самым крупным и наиболее популярным в настоящее время справочником в Интернете, основанном на принципе свободного сбора документированного материала. Существует ряд правил, которым должны удовлетворять публикуемые документы. За соблюдением этих правил следят редакторы Википедии. По объему сведений и тематическому охвату Википедия считается самой полной энциклопедией из когда-либо создававшихся за всю историю человечества. На май 2018 г. Википедия содержала более 40 млн статей. Интернет-сайт Википедии является пятым по посещаемости сайтом в мире и широко используется в качестве справочника.

В рамках Википедии стали доступны сведения обо всех сколько-нибудь значимых личностях.

В 2008 г. нами был начат систематический сбор информации, содержащейся в Википедии, в 2019 г. база данных «Личности Энциклопедии» была зарегистрирована в Реестре баз данных Роспатента. Объем базы ежегодно увеличивался на 60–100 тыс. человек и на момент написания этой статьи составлял более 726 тыс. персоналий.

За прошедшие годы с помощью материалов базы «Личности Энциклопедии» был опубликован ряд статей, в которых мы исследовали продолжительность жизни и некоторые причины смерти представителей различных, преимущественно творческих, профессий и ученых, включенных в Википедию [3, 4, 12–16].

Увеличивающийся с каждым годом объем базы позволяет выявлять новые, не всегда очевидные и ожидаемые, закономерности. Одновременно с этим неизбежно встает вопрос о том, насколько корректно материалы Википедии отражают настоящие демографические процессы. Данная статья является попыткой ответить на этот вопрос.

Специфика данных по продолжительности жизни, содержащихся в исторической базе «Личности Энциклопедии». Данные, собранные в базе данных, охватывают широкий исторический период от I века н. э. до настоящего времени. В силу особенностей процедур появления информации в Википедии данные по продолжительности жизни подвержены искажениям, которые необходимо учитывать при их интерпретации. На рис. 1 представлены значения СВС мужчин, рожденных с 1850 по 2000 гг. и занесенных в базу данных.

На рисунке выделяются два периода — с 1850 по 1924 гг. и с 1925 по 2000 гг. В первом периоде СВС держится на уровне 68 лет и, начиная с 1899 г., увеличивается до 77 лет. Начиная с 1925 г., СВС по данным из базы неуклонно снижается до 40 лет к 1970 г., опускаясь до 17 лет в 2000 г. Такое падение СВС очевидно не отражает демографические тенденции и связано со спецификой данных, представленных в исторической базе «Личности Энциклопедии».

Если оставить в стороне вопрос о точности сообщенных в энциклопедических изданиях дат рождения и смерти, то при интерпретации результатов анализа собранных данных следует учитывать наличие двух существенных ограничений: цензурирования и селекции данных. Цензурирование возникает в случае, если дата смерти человека неизвестна. В настоящем исследовании лица с неуказанной датой смерти, но рожденные в XX в., рассматриваются как случаи цензурирования на дату последней ревизии исторической базы данных. Лица, рожденные до XX в., дата смерти которых неизвестна, исключались из анализа. Селекция данных является следствием того, что в энциклопедию вносятся сведения не о случайно выбранных из населения людях, а о в каком-то

смысле выдающихся личностях. Учет указанных обстоятельств необходим для корректной интерпретации данных, собранных в базе «Личности Энциклопедии».

В ряде случаев удается учесть и нивелировать влияние селекции на оценку продолжительности жизни. Примером служит изучение продолжительности жизни нобелевских лауреатов [17]. Анализ показывает, что лауреаты живут дольше, чем население страны, в которой они проживают. Это объясняется тем, что поскольку Нобелевскую премию присуждают за открытия, сделанные в возрасте 30–40 лет, а то и позже, то все лауреаты дожили до достаточно зрелого возраста совершения открытия, до которого не доживает часть прочего населения. Отсюда становится ясно, что сравнивать надо не ОПЖ при рождении, а ОПЖ в возрасте совершения открытия. Сравнение дожития нобелевских лауреатов с дожитием населения США, представителями которого являются многие лауреаты, не выявило статистически значимых различий [17].

Подобное обстоятельство учитывалось и в исследовании Е.М. Андреева и Д.А. Жданова продолжительности жизни российских академиков (<http://www.demoscope.ru/weekly/2007/0283/tema03.php>). Оказалось, что ОПЖ в возрасте 50 лет членов Российской академии наук и мужчин России в 1901–1930 гг. статистически не различались, но ОПЖ членов Российской академии наук и мужчин Швеции различались очень сильно. Затем следует провал в ОПЖ академиков и мужчин России, а с 1970 по 2007 гг. ОПЖ в возрасте 50 лет членов Российской академии статистически не отличается от ОПЖ мужчин в том же возрасте, но уже не в России, а в Швеции.

Для корректной интерпретации данных из базы «Личности Энциклопедии» о продолжительности жизни важно учесть механизм селекции — выбора персоналий, включаемых в базу данных. Относительно этого механизма можно строить различные гипотезы. Мы предполагаем, что человек должен, как минимум, дожить до некоторого возраста, чтобы быть включенным в Википедию. В дальнейшем будем считать этот возраст равным 20 годам. Существование более сложных правил, связанных, возможно, с историческими социально-экономическими трендами и профессиональными предпочтениями, учесть не представляется возможным из-за отсутствия соответствующей информации.

Учет цензурирования. Для учета наличия цензурирования при расчете дожития применяется непараметрическая оценка вероятности дожития — множительная оценка кривой смертности Каплана–Майера, вычисляемая по формуле [18]:

$$S(t_j) = \prod_{i=1}^j \left(1 - \frac{d_i}{r_i} \right), \quad (4)$$

где t_0, t_1, \dots, t_j — последовательность лет, для которых вычисляются значения вероятности дожития до времени t_j (функции дожития); r_i — число людей, доживших до года t_i ; d_i — число людей, умерших между годами t_{i-1} и t_i . Особенностью формулы (4) является то, что убыль людей между годами t_{i-1} и t_i складывается из числа умерших за этот период людей d_i и числа цензурированных людей, выбывших из наблюдения.

Для получения оценки ОПЖ при наличии цензурирования в математической демографии используется величина математического ожидания продолжительности жизни, равная площади под кривой дожития — доли людей, проживших дольше текущего возраста. Кривую дожития строят, используя непараметрическую оценку Каплана–Майера, рассчитанную по имеющимся данным с учетом цензурирования.

Учет влияния селекции данных по продолжительности жизни, включенных в базу «Личности Энциклопедии». Прежде всего, интересно провести сравнение СВС лиц, включенных в базу данных, с ОПЖ при рождении, рассчитанной демографами. В [1] приведены исторические данные о средней продолжительности жизни в странах, в которых на текущий год она была максимальна. Показано, что ОПЖ при рождении среди мужчин в упомянутых странах, начиная с 1840 г., монотонно растет, увеличиваясь за год на три месяца. Тенденция роста ОПЖ для лиц, рожденных после 1925 г., сохраняется, но СВС, рассчитанный по данным базы «Личности Энциклопедии», монотонно снижается с уровня 75 лет в 1925 г. до 17 лет в 2000 г., а в странах–рекордсменах ОПЖ в 2000 г. достигает уровня 77,8 лет. Похожая картина наблюдается и при анализе данных по продолжительности жизни женщин. Строго говоря, сравнение СВС, рассчитанного по базе «Личности Энциклопедии», с ОПЖ в странах–рекордсменах некорректно, поскольку СВС рассчитывается для когорт — людей одного года рождения, а ОПЖ вычисляется путем создания синтетической когорты. Однако на качественном уровне такое сравнение кажется допустимым.

Возможный механизм селекции и цензурирования данных, занесенных в Википедию, можно сформулировать в виде двух необходимых условий:

- 1) человек должен быть взрослым — достигнуть возраста 20 лет;
- 2) человек должен умереть до даты последнего обновления базы.

Условие 2 подкрепляется тем фактом, что из 149119 человек, родившихся после 1925 г. и занесенных в историческую базу «Личности Энциклопедии» до 13 марта 2020 г. (момент последнего обновления) живыми оказались только 5949 человек, а умерли 143170.

СВС, рассчитанный по формуле (3), является случайной величиной, математическое ожидание

которой для лиц, рожденных в году y и умерших в возрасте между a_0 и a_1 , задается формулой:

$$E_{a_0, a_1}(y) = \frac{(a_0 + e_{a_0}) \times S_y(a_0) - (a_1 + e_{a_1}) \times S_y(a_1)}{S_y(a_0) - S_y(a_1)}, \quad (5)$$

где $S_y(a_0)$ и $S_y(a_1)$ — значения вероятности дожития до возраста a_0 и a_1 при рождении в году y соответственно; e_{a_0} и e_{a_1} — значения ОПЖ в возрасте a_0 и a_1 соответственно при рождении в году y . Заменив в формуле (5) математическое ожидание СВС его эмпирической оценкой $E_3(y)$, вычисленной по формуле (3), можно с учетом механизма селекции, описанного выше, по данным из базы «Личности Энциклопедии», рассчитать ОПЖ при рождении для лиц, рожденных в году y :

$$E_{sel}(y) = \frac{(S_y(a_0) - S_y(a_1)) \times E_3(y) + (a_1 + e_{a_1}) \times S_y(a_1)}{S_y(a_0)}, \quad (6)$$

где $E_3(y)$ — величина СВС при наличии механизма селекции среди людей, родившихся в году y ; $S_y(a)$ — вероятность дожития до возраста a при рождении в году y .

Справедливость формулы (6) проверялась в ходе «компьютерного эксперимента» по данным о продолжительности жизни мужчин в Швейцарии, содержащимся в базе данных по смертности человека НМД. Цель эксперимента заключалась в проверке того, что по когортным данным о среднем возрасте смерти для людей, доживших до 20 лет (предполагаемое необходимое условие включения в Википедию), по формуле (6) можно рассчитать ОПЖ при рождении. Швейцария была выбрана, поскольку для этой страны в НМД наиболее полно представлены данные за период рождения с 1925 по 2000 гг. Данные имеют вид набора «period life tables» — таблиц смертности, в которых данные по смертности, ОПЖ в различных возрастных группах берутся из когорт различных годов рождения. Кроме того, Швейцария не участвовала в изучаемый период в войнах и, по нашему мнению, имеющиеся данные в малой степени подвержены искажениям. По данным НМД вычислялась ОПЖ в возрасте 20 лет в зависимости от года рождения для лиц, рожденных в интервале 1925–2000 гг. и удовлетворяющих условиям включения в Википедию: прожить 20 лет и умереть до 2000 г. Последнее условие означает, что полученное значение ОПЖ совпадает с теоретическим значением наблюдаемого СВС $E_3(y)$. Результаты вычисления получившейся зависимости $E_3(y)$ от года рождения представлены на рис. 2 светлыми квадратами. Виден эффект снижения СВС во времени, аналогичный наблюдаемому на рис. 1 для данных из базы «Личности Энциклопедии».

Для применения формулы (6) необходимо задать вероятность $S_y(a)$ дожития до возраста a при рождении в году y , а также значения ОПЖ e_a

в возрасте a . Значения вероятности $S_y(a)$ брались из таблиц смертности последовательно изменяя год рождения y начиная с 1925 г. Поскольку предполагается, что в Википедию включаются в основном люди старше 20 лет, то при вычислениях рассматривались значения a не ниже этого возраста и, таким образом, исключалось влияние изменений в детской смертности. Аналогично бралась из таблиц и величина ОПЖ в возрасте $a - e_a$. Восстановленные значения СВС с учетом селекции

для лиц, рожденных в 1925 г., показаны на рис. 2 крестиками. Там же черными кружками показаны значения ОПЖ при рождении для мужчин, вычисленные непосредственно по данным продолжительности жизни мужчин в Швейцарии, рожденных в интервале 1925–2000 гг., но без условия смерти до 2000 г. На рис. 2 видно совпадение рассчитанных по формуле (6) значений СВС со значениями ОПЖ при рождении, рассчитанными по демографическим данным.

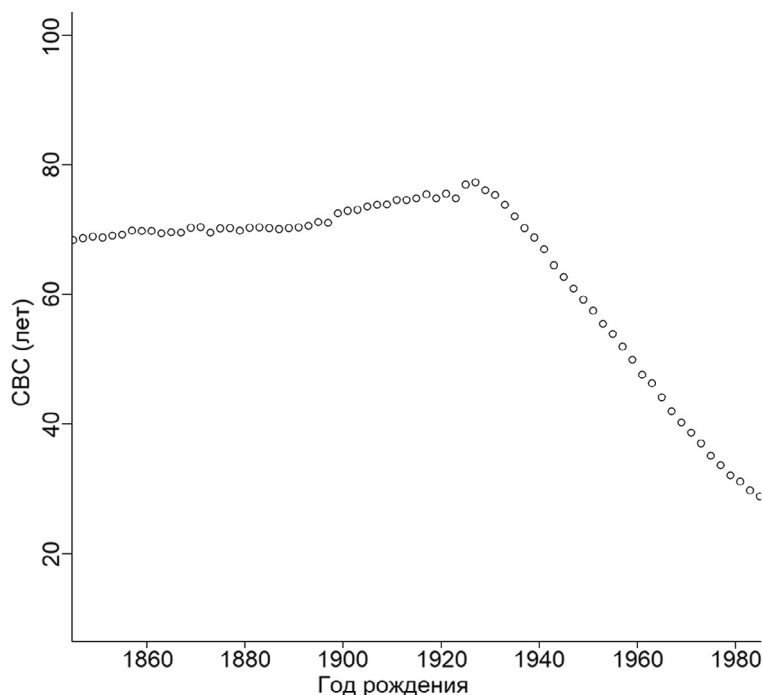


Рис. 1. СВС мужчин по данным из исторической базы «Личности Энциклопедии».

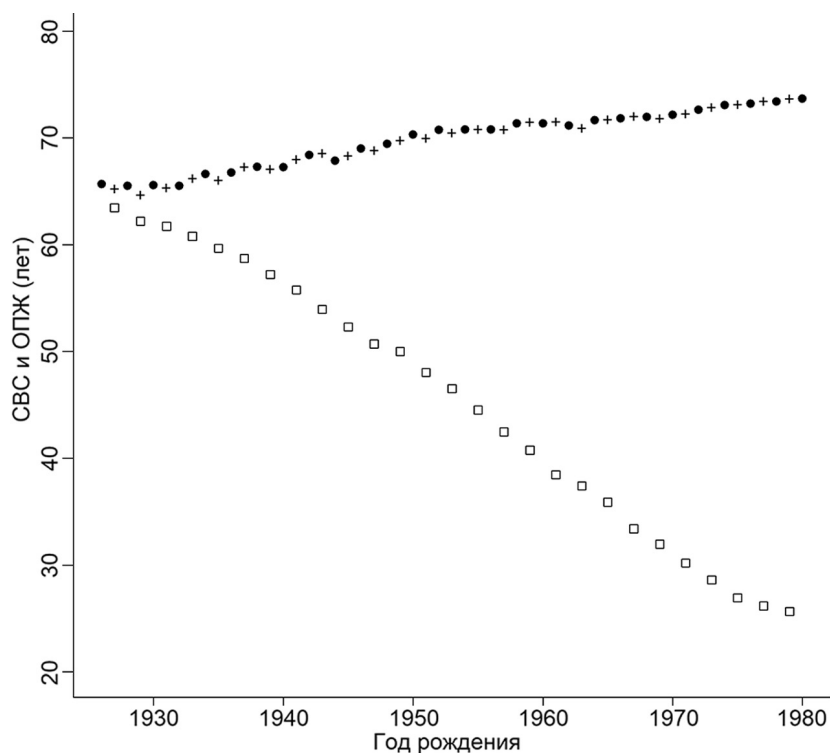


Рис. 2. ОПЖ при рождении мужчин в Швейцарии (темные кружки), СВС мужчин в Швейцарии после селекции по условиям, предполагаемым в статье, (светлые квадраты), восстановленные по формуле (6) значения ОПЖ при рождении (+).

Результаты и обсуждение

Учет селекции. Формулу (6) применили к данным из исторической базы «Личности Энциклопедии» для лиц, рожденных после 1924 г. Светлые символы на рис. 3 показывают результаты расчета СВС для лиц, родившихся с 1850 по 2000 гг. Крестиками показаны результаты расчета с учетом селекции по формуле (6) ОПЖ при рождении после 1924 г. При расчетах в качестве функции дожития использовали оценку функции дожития Каплана–Майера, построенную по записям базы для когорты лиц, рожденных в 1925 г. Люди, дожившие до 2010 г., считались цензурированными в возрасте 85 лет. Таким образом, предполагая, что для лиц, упомянутых в Википедии и рожденных после 1925 г., характер дожития будет такой же, как и в когорте 1925 г. рождения, получили прогноз ОПЖ при рождении, учитывающий селективный характер данных Википедии о СВС. Достоверность этого прогноза можно будет проверить, когда в Википедии появятся более полные данные о дате смерти лиц, рожденных в период 1925–2010 гг. Наши вычисления, скорее всего, дают нижнюю границу ОПЖ, поскольку используют характеристики дожития в когорте 1925 г. рождения, игнорируя существующий тренд к увеличению продолжительности жизни.

Сопоставление СВС и ОПЖ. Приведенные в статье формулы позволяют оценить, насколько СВС отличается от ОПЖ при рождении. Если механизм селекции отсутствует, то при вычислении СВС принимаются во внимание все рожденные люди, то есть $a_0 = 0$, а $S_y(a_0) = 1$. Возраст a_1 можно положить равным максимальному, вероятность

дожития до которого $S_y(a_1)$ равна нулю. Формула (5) преобразуется в выражение:

$$E_0(y) = e_0,$$

то есть различия между СВС и ОПЖ при рождении нет.

Если вычисление СВС проводится только среди людей, доживших до возраста a_0 , как при селекции данных Википедии, то формула (5) принимает вид:

$$E_{a_0}(y) = a_0 + e_{a_0}.$$

В этом выражении величина $a_0 + e_{a_0}$ равна ОПЖ при рождении для людей, доживших до возраста a_0 . Таким образом, СВС, вычисленный по данным Википедии, превосходит ОПЖ при рождении всего населения не более чем на 20 лет. Причем можно предположить, что порог селекции по возрасту a_0 тем выше, чем дальше в историю мы углубляемся и различие между СВС и ОПЖ при рождении всего населения увеличивается.

Интересно сравнить с этой точки зрения оценки СВС, построенные по данным исторической базы «Личности Энциклопедии», и ОПЖ при рождении в различные годы в странах-рекордсменах [1]. Разница между СВС и ОПЖ изменится от примерно 25 лет в 1850 г. до примерно 15 лет в 1920 г. Это, возможно, является результатом снижения детской смертности, ведущим к росту ОПЖ при рождении. СВС, рассчитанный по данным базы «Личности Энциклопедии», нечувствителен к этому явлению, поскольку вычисляется на основании данных жизни людей, переживших детский возраст.

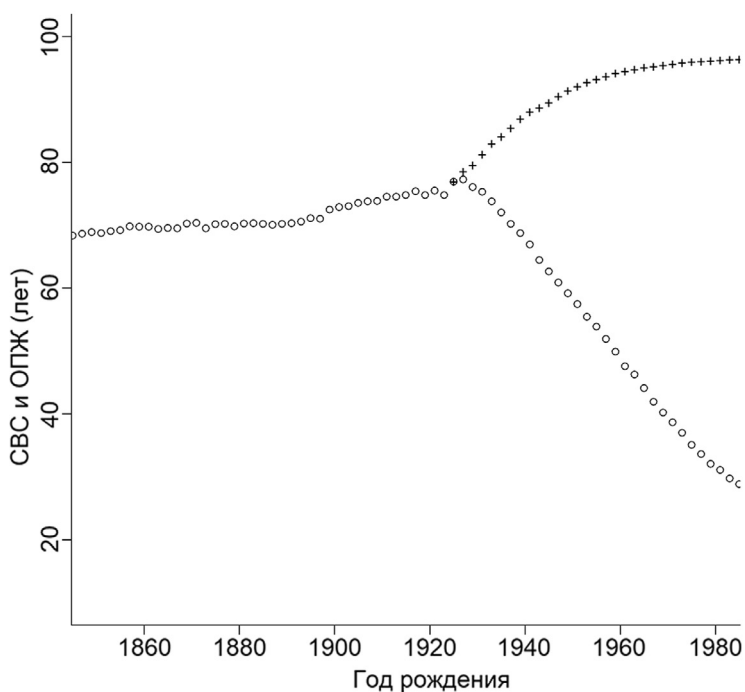


Рис. 3. СВС мужчин по данным из исторической базы «Личности Энциклопедии» (светлые символы), восстановленные по формуле (6) значения ОПЖ при рождении мужчин, рожденных после 1924 г. (+).

Мы рассмотрели проблему применимости данных Википедии, представленных в базе «Личности Энциклопедии», при оценке исторических трендов продолжительности жизни. Поскольку данные в Википедии не являются результатом случайного выбора из фиксированного населения, то при изучении динамики изменения продолжительности жизни необходимо учитывать наличие цензурирования и селекции. В ранее опубликованной работе [19] авторы поднимают вопрос о необходимости учета указанных недостатков данных из Википедии и демонстрируют результат такого учета. В настоящей статье даны подробные рекомендации и приведены формулы для такого учета, применимые при анализе любых демографических, социологических, экономических данных о длительности событий, подверженных селекции и цензурированию. Существенной особенностью приведенных в статье формул является необходимость использования оценок вероятности дожития именно в изучаемых

группах людей — женщин, мужчин, лиц одиноким профессий и т.д. Точность таких оценок резко снижается с уменьшением численности группы. С другой стороны, увеличение численности группы ведет к возрастанию гетерогенности данных. При оценке вероятности дожития возможны допущения — например, о малости влияния временного фактора на возрастную зависимость дожития, а также возможно использование рассчитанных демографических характеристик из, например, базы данных по смертности человека НМД.

Авторы выражают глубокую признательность программисту А.С. Чиркову, создавшему систему извлечения и обработки информации, без которой выполнение этой работы было бы невозможным. Исследование выполнено при финансовой поддержке Министерства здравоохранения Российской Федерации. Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Oeppen J., Vaupel J.W. Broken limits to life expectancy // *Science*. 2002. Vol. 296. N 5570. P. 1029–1031.
2. Vaupel J.W., Carey J.R., Christensen K., Johnson T.E., Yashin A.I., Holm N.V., Iachine I.A., Kannisto V., Khazaeli A.A., Liedo P., Longo V.D., Zeng Y., Manton K.G., Curtsinger J.W. Biodemographic trajectories of longevity // *Science*. 1998. Vol. 280. N 5365. P. 855–860.
3. Zharinov G.M., Anisimov V.N. Lifespan, longevity, and incidence of violent death in writers and poets // *Adv. Gerontol.* 2016. Vol. 6. N 4. P. 265–271.
4. Zharinov G.M., Anisimov V.N. Music and longevity // *Adv. Gerontol.* 2014. Vol. 4. N 4. P. 283–289.
5. Kenny D.T., Asher A. Life expectancy and cause of death in popular musicians: is the popular musician lifestyle the road to ruin? // *Med. Probl. Perform. Ar.* 2016. Vol. 31. N 1. P. 37–44.
6. Thomas R.K. Concepts, methods and practical applications in applied demography. Cham: Springer, 2018. 333 pp.
7. Yashin A.I., Begun A.S., Boiko S.I., Ukraintseva S.V., Oeppen J. New age patterns of survival improvement in Sweden: do they characterize changes in individual aging? // *Mech. Ageing Dev.* 2002. Vol. 123. N 6. P. 637–647.
8. Lee R.D., Carter L.R. Modeling and forecasting U. S. mortality // *J. Am. Stat. Assoc.* 1992. Vol. 87. N 419. P. 659–671.
9. Вальков М.Ю., Карпунов А.А., Коулман М.П., Аллемани К., Панкратьева А.Ю., Потехина Е.Ф., Валькова Л.Е., Гржибовский А.М. Популяционный раковый регистр как ресурс для науки и практического здравоохранения // *Экол. чел.* 2017. № 5. С. 54–62.
10. Kyvik K.O., Christensen K., Skytthe A., Harvald B., Holm N.V. The Danish twin register // *Dan. Med. Bull.* 1996. Vol. 43. N 5. P. 467–470.
11. Kagan A., Dawber T.R., Kannel W.B., Revotskie N. The Framingham study: a prospective study of coronary heart disease // *Fed. Proc.* 1962. Vol. 21. N 4. Pt. 2. P. 52–57.
12. Anisimov V.N., Zharinov G.M. Life span and longevity in representatives of creative professions // *Adv. Gerontol.* 2014. Vol. 4. N 1. P. 83–94.
13. Zharinov G.M., Mikhalsky A.I., Neklasova N.Y., Anisimov V.N. Longevity and some causes of death in musicians of the 20th Century // *Adv. Gerontol.* 2020. Vol. 10. N 2. P. 193–203.
14. Анисимов В.Н., Жаринов Г.М. Продолжительность жизни ученых различных специальностей // *Природа*. 2014. № 5. С. 51–52.
15. Anisimov V.N., Zharinov G.M. Mean age of death and longevity for male scientists of different specialties // *Moscow Univ. Biol. Sci. Bull.* 2016. Vol. 71. N 4. P. 193–198.
16. Anisimov V.N., Zharinov G.M. Mean age of death, longevity and incidence of violent death among various elite sportsmen categories // *J. Geront. Geriatr. Res.* 2017. Vol. 6: 6.
17. Анисимов В.Н., Михальский А.И. Старее ли Нобелевский лауреат? Математический анализ возраста и продолжительности жизни лауреатов Нобелевской премии за 1901–2003 гг. // *Усп. геронтол.* 2004. № 15. С. 14–22.
18. Cox D.R., Oakes D. Analysis of survival data. N.Y.: Chapman and Hall, 1984. 212 pp.
19. Michalski A.I., Zharinov G.M., Anisimov V.N. Capacities and limitations of the use of data from Wikipedia for analysis of human life expectancy // *Adv. Gerontol.* 2021. Vol. 11. N 1. P. 1–7.

Поступила в редакцию 20.04.2021 г.

После доработки 20.06.2021 г.

Принята в печать 12.07.2021 г.

RESEARCH ARTICLE

Analysis of human life span through history

A.I. Mikhalski^{1,*} , V.N. Anisimov² , G.M. Zharinov³ ¹*V.A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences,
65 Profsoyuznaya st., Moscow, 117997, Russia;*²*N.N. Petrov National Medical Research Center of Oncology,
68 Leningradskaya st., Pesochny, Saint Petersburg, 197758, Russia;*³*A.M. Granov Russian Research Center for Radiology and Surgical Technologies,
70 Leningradskaya st., Pesochny, St. Petersburg, 197758, Russia***e-mail: ipuran@yandex.ru*

The problem of correct interpretation of data on the mean age of death of people in different historical periods based on information presented in Wikipedia and included in the “Personalities of the Encyclopedia” database is considered. The specificity of these data, associated with censoring and the nature of their collection, leading to selectivity, is noted. An approach is proposed that takes into account these circumstances, which makes it possible to clarify the obtained patterns of life expectancy and expand the historical range of research.

Keywords: *life span, mean age at death, Wikipedia, “Personalities of the Encyclopedia”, censoring, selectivity, survival probability*

Funding: The research was carried out with the financial support of the Ministry of Health of Russian Federation.

Сведения об авторах

Михальский Анатолий Иванович — докт. биол. наук, ст. науч. сотр., гл. науч. сотр. Института проблем управления им. В.А. Трапезникова РАН. Тел.: 8-495-198-17-20; e-mail: ipuran@yandex.ru; ORCID: <https://orcid.org/0000-0003-1998-1315>

Анисимов Владимир Николаевич — чл.-корр. РАН, докт. мед. наук, проф., зав. научным отделом канцерогенеза и онкогеронтологии Национального медицинского исследовательского центра онкологии им Н.Н. Петрова Минздрава России. Тел.: 8-812-439-95-34; e-mail: aging@mail.ru; ORCID: <https://orcid.org/0000-0002-3683-861X>

Жаринов Геннадий Михайлович — докт. мед. наук, проф., гл. науч. сотр. отдела лучевых и комбинированных методов лечения Российского научного центра радиологии и хирургических технологий имени академика А.М. Гранова. Тел.: 8-812-439-66-44; e-mail: asatur15@mail.ru; ORCID: <https://orcid.org/0000-0002-6034-2040>