

*Е.Л. Сороко\**

## **ИСТОЧНИКИ ИСКАЖЕНИЯ ИНФОРМАЦИИ В ТАБЛИЦАХ СМЕРТНОСТИ\*\***

Цель предлагаемой статьи – привлечь внимание к проблеме искажения информации, содержащейся в публикуемых таблицах смертности. Работа базируется на анализе результатов, содержащихся в Human Mortality Database, а также содержит теоретические положения, дающие более глубокое понимание проблемы и объяснение того, что может служить источником искажений. Для иллюстрации рассмотрим следующую задачу. Исследователя интересует, в чем состоит прогресс в укреплении здоровья и безопасности жизни детей младшего школьного возраста, и для этого предполагается использовать показатели из таблиц смертности. При попытке решить задачу для конкретной страны может возникнуть ряд проблем, существенно усложняющих ее решение. Вот некоторые из них: а) потеря точности в результате неправильного округления десятичных знаков при публикации, б) псевдодискретность возрастных показателей смертности, в) значительный относительный вклад случайных флуктуаций, г) псевдоравенство значений показателей, д) достижение «абсолютного нуля» вероятности смерти в некоторых возрастах, е) потеря свойств монотонности и непрерывности возрастных профилей, ж) псевдорост значений  $q_x$ .

### ***Общемировые тенденции и таблицы смертности***

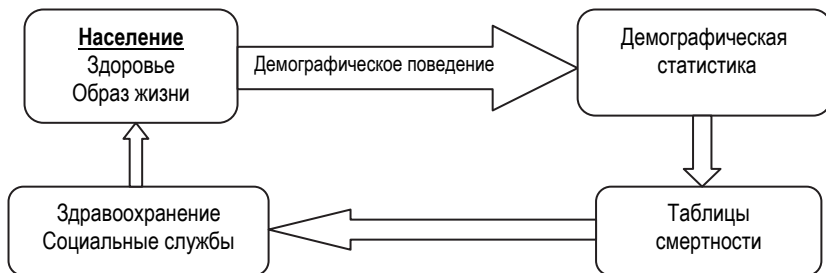
Прогресс человеческой цивилизации приводит к устойчивому сокращению смертности. Она снижается во всех возрастно-половых группах, происходит устойчивый неуклонный рост продолжительности жизни при рождении. Общемировые тенденции в этой сфере ярко описаны в работе Оппена и Вопеля (2002). Немалые изменения происходят в технологии сбора и обработки статистической информации о смертности. Учет и регистрация смертных случаев становится более полным и точным, совершенствуется в большинстве стран мира. А фиксация индивидуальных случаев смерти открывает новые горизонты применения все более тонких методов анализа. В качестве примера можно привести статью Школьниковова В.М.

\* Сороко Евгений Львович – к.ф.-м.н., с.н.с. Института демографии НИУ ВШЭ.

\*\* Исследование выполнено при поддержке РФФИ (грант № 14-07-00920).

с соавторами (2012). Однако, все-таки основным средством анализа смертности остаются таблицы смертности.

Информация, содержащаяся в таблицах смертности, служит основанием для совершенствования социальной и демографической политики, выработки приоритетов в области здравоохранения. Общая схема значения данной информации в перечисленных сферах можно представить на следующей схеме.



**Рис. 1.** Общая схема места демографической статистики в социальной политике

Таким образом, в системе управления социальной сферой таблицы смертности выполняют функцию обеспечения соответствующих служб информацией о результатах демографического анализа с адекватными показателями. Как и в любой сложной системе, при передаче информации возможны шумы, задержки и разного рода искажения. Возникает вопрос о том, какого рода искажения могут помешать использованию таблицы смертности.

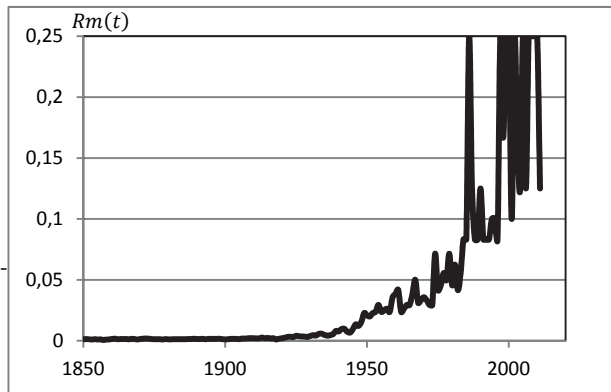
К настоящему времени сложился состав таблицы смертности, содержащий, как минимум: Возраст  $x$ , Коэффициент смертности  $m_x$  в возрасте  $x$ , Число доживающих до возраста  $x$ , Число умирающих в возрасте от  $x$  до  $x + 1$ , Вероятность смерти в возрасте  $x$   $q_x$ , Вероятность дожития до следующего возраста  $x + 1$   $p_x$ , Число живущих в возрасте  $x$   $L_x$ , Число человеко-лет жизни в возрасте  $x$  и старше. Ожидаемая продолжительность жизни в возрасте  $x$ . Вот уже полстолетия расчет таблиц смертности производится с использованием вычислительных машин, что позволяет хранить все исходные данные таблиц, делать промежуточные вычисления и подсчет результирующих показателей без каких-либо сомнений в их точности. Но именно после этого появляется первое противоречие таблиц смертности. Это противоречие между точностью показателей и вычислений при расчете таблицы смертности и при ее публикации. По

негласному согласию демографов со всего мира значения показателей  $m_x$  и  $q_x$  приводятся со значением только пяти знаков после запятой. К сожалению, затруднительно найти серьезное обоснование того, что их точность ограничивается только пятью знаками. Значения показателей  $m_x$  и  $q_x$  с точностью 5 знаков для публикации таблицы смертности населения развитой страны середины XX века можно рассматривать как достаточную.

Но сегодня это уже не так. Чтобы оперировать конкретными цифрами, предлагается рассмотреть несколько показателей. Ошибка округления  $q_x R(q_x) = |q_x - q_x^{LT}|$  представляет собой абсолютную величину разницы точного значения  $q_x$  и опубликованного в таблице смертности. Для округления до 5 значащих цифр после запятой по правилам округления максимальная величина этой ошибки  $\max R(q_x)$  составит 0,000005. С первого взгляда эта величина столь незначительна, что не может повлиять на результат демографического анализа показателей таблицы смертности. Это, конечно именно так для таких важных показателей, как числа доживших и продолжительности жизни. Поэтому рассмотрим также относительную максимальную величину этой ошибки  $Rel(q_x(t)) = \max R(q_x(t)) \div q_x(t)$ . Например, у шведских девочек в 2010 г. в возрасте 9 лет величина  $q_9 = 0,00002$ , и максимальная относительная ошибка, вызванная округлением, может достичь уже 25% ( $0,000005 \div 0,00002 \times 100$ ). Реальная ошибка округления будет, конечно, меньше (6%).

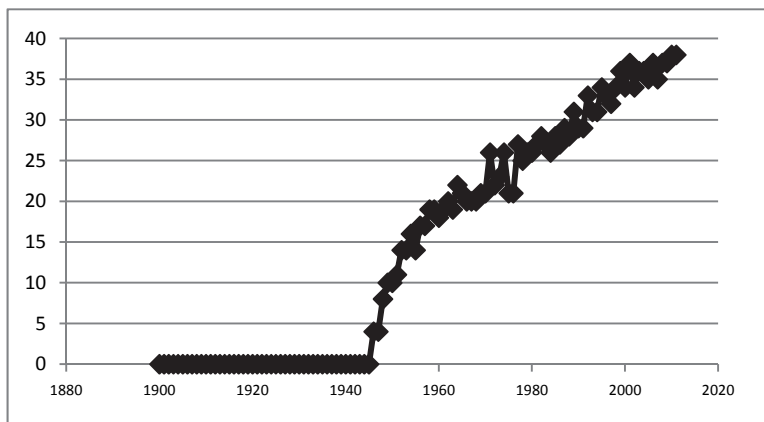
Динамика показателя максимальной величины по возрасту относительной ошибки  $Rm(t) = \max_x Rel(q_x(t))$  в качестве примера для женского населения Швеции выглядит следующим образом (рис. 2).

**Рис. 2.**  
Динамика максимальной относительной ошибки  $Rm(t)$ , Швеция, женщины, 1850–2011.  
*Источник:* HMD.



Существенным является рост максимальной относительной ошибки после 1950 г. Впервые она достигает величины 25% уже в 1986 г, а на рубеже веков встречается все чаще, не опускаясь последние 10 лет ниже 12,5%.

С чем связано данное явление, становится отчетливо ясным, если сделать подсчет числа возрастных групп, в которых значение  $q_x$  крайне мало, например, менее 0,00005. Динамика этого показателя представлена на рис. 3.



**Рис. 3.** Число возрастных групп, в которых  $q_x < 0,00005$ , Швеция, женщины, 1900–2011.

**Источник:** НМД.

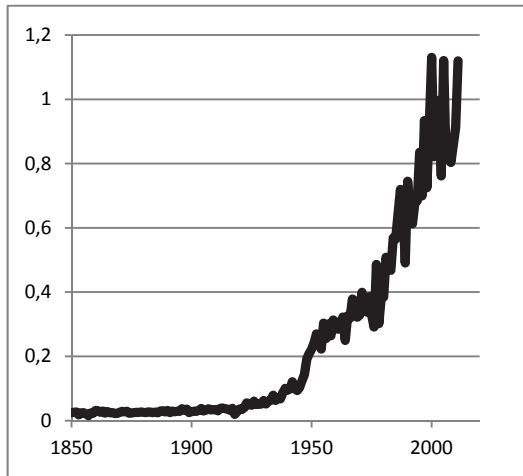
Таким образом, для шведских женщин превышает сейчас уже одну треть длины всего возрастного интервала  $[0, \omega]$  в таблицах смертности, где  $\omega = 110$  в НМД.

Рассмотрим теперь не максимальную относительную ошибку, а динамику средней ошибки  $Mean_{x=1, \dots, 79}(R(q_x(t) \div q_x))$  для возрастных групп от 1 года до 79 лет (рис. 4).

Таким образом, если до середины XX века средняя относительная ошибка между «истинным» значением  $q_x$  и ее значениями из таблиц смертности пренебрежимо мала, то к началу XXI века она не только демонстрирует устойчивый рост, но и превышает уже 1%.

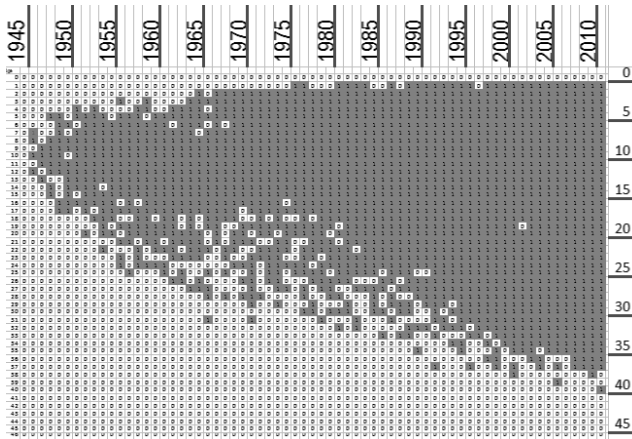
Вполне естественное предположение, что в настоящее время соотношение  $q_x < 0,00005$  характерно лишь для групп младших детских возрастов, не подтверждается, этот возрастной интервал стреми-

тельно расширяется и на взрослое население, а наиболее ярко процесс расширения таких возрастов, показывает Лексис-диаграмма, рис. 5.



**Рис. 4.** Средняя относительная ошибка округления  $q_x$ , Швеция, женщины, 1850–2011, %

*Источник:* HMD.



**Рис. 5.** Лексис-диаграмма динамики возрастных групп с  $q_x$  меньше 0,00005 в возрастах от 0 до 45 лет, Швеция, женщины, 1945–2011, выделены темным

*Источник:* HMD.

### Псевдодискретность $m_x$ и $q_x$

Проблема возможных искажений значений показателей в таблицах смертности не сводится только к ошибкам округления. Большие сомнения возникают также при рассмотрении спектра значений, например, возрастных коэффициентов смертности. Теоретические модели предполагают непрерывность их возрастного профиля. Типичный вид возрастного профиля  $m_x$  представлен на рис. 6. Для нашего анализа отметим две важнейших качественных особенности: монотонность зависимости от возраста и непрерывность кривой.

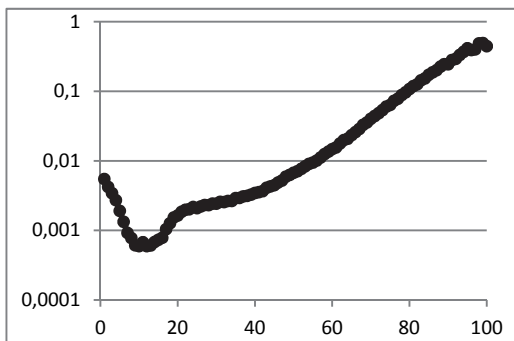


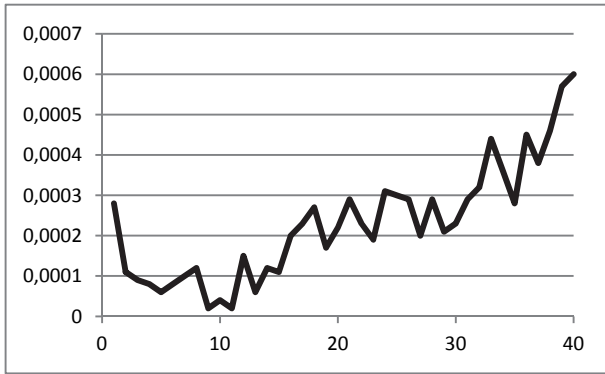
Рис. 6. Типичный вид возрастного профиля  $m_x$ , Япония, женщины, 1955

*Источник:* НМД.

Сегодняшнее население Швеции демонстрирует отсутствие такой непрерывности. Значения показателя, реально наблюдаемые в таблицах смертности, выстраиваются как бы ступеньками. Так, за все время статистики смертности, значение 0,00002 наблюдалось 12 раз, 0,00004 – 16 раз, а 0,00003 – лишь дважды. На возрастном интервале 1–40 лет возрастной профиль имеет, представленный на рис. 7

Он демонстрирует отсутствие монотонной зависимости от возраста, которую можно условно назвать «пилообразность», и не только. Значения показателей в соседних возрастных группах не только не показывают близости друг к другу, но и демонстрируют соотношения, превышающие 7. Для выяснения причины эффекта отвлекусь пока от Швеции и рассмотрим условную страну, в которой в течение ряда лет рождалось 67 000 детей ежегодно. Для режима смертности из 1, 2, 3 смертей в отдельной группе девочек возраста  $x$  в течение года значения показателей  $m_x$  составят 0,00003, 0,00006 и 0,00009. Таким образом, спектр их возможных значений окажется

узким (всего три), что является результатом совместного действия факторов неизменности масштаба воспроизводства, округлений до 5 знаков, а также малым числом случаев. Почему данное свойство предлагается назвать «псевдо»? Причина в том, что такая дискретность значений вовсе не связана со способом измерения процесса и не отражает демографический смысл данной характеристики. Это лишь результат того, что в данном населении происходит одновременное действие указанных факторов.



**Рис. 7.** Возрастной профиль  $m_x$  для возрастов от 1 года до 40 лет, Швеция, женщины, 2010

**Источник:** HMD.

Еще один пример влияния округления значений показателей при низком уровне смертности это возможность появления в когортных таблицах смертности неожиданных случаев «псевдороста». Для демонстрации этого явления в чистом виде предлагается использовать модельный пример. Рассмотрим вначале две последовательных когорты родившихся численностью 41 и 40 тысяч. Допустим, что отсутствуют случайные колебания чисел умерших, и в каждой из этих когорт происходит ровно 1 смерть в год. Но, если в первой когорте вероятность смерти окажется при этих предположениях 0,00002, то в следующей уже 0,00003, то есть вырастет якобы на 50%, в то время как ее реальная интенсивность отличается всего на 2,5%. Любопытно, что если рассмотреть еще одну следующую когорту из 39 тысяч родившихся, то возникнет эффект «псевдоравенства». При тех же числах умерших интенсивность смертности в этой когорте по сравнению с предыдущей вырастет на 2,5%. Однако, в когортной табли-

це появится совершенно неизменное значение  $q_x = 0,00003$ . Другая иллюстрация «псевдоравенства» – многократное повторение одного и того же значения  $q_x$  в разные периоды и различных возрастах, которое может быть воспринято формально математически как строгое равенство: Однако, если вспомнить связь  $q_x$  и  $m_x$ , а последнего с исходными данными демографической статистики  $D_x$  и  $E_x$ , то точное равенство может быть достигнуто лишь в случае одновременно двойного равенства соответствующих  $D_x$  и  $E_x$ : и, или их точной пропорциональности. Вероятность достижения подобного равенства крайне мала, а, например, для населения Швеция за 260 лет пока не наблюдалось.

### *А достигим ли «абсолютный ноль»?*

В физике ответ на этот вопрос уже давно решен. В отличие от нуля Цельсия, к абсолютному нулю Кельвина можно приблизиться сколь угодно близко, но нельзя достичь точно. А что же в демографии? Данный вопрос вполне естественно возникает, когда в таблице встречается нулевое значение  $q_x$ . Дело в том, что формально демографически, все вычисления сделаны правильно и математически точно. Однако, имеет ли содержательный смысл нулевое значение  $q_x$ ? Попробуем в этом разобраться. В качестве примера рассмотрим конкретный случай, состоящий в том, что в 2006 году не умерло ни одной шведской девочки в возрасте 7 лет. Можно допустить, что забота родственников и современный уровень медицины позволяет в этом возрасте полностью избежать смерти из-за болезни. Однако нельзя согласиться, что в течение всего года ни одна девочка не оказалась в качестве пассажира в автомобиле, ни разу не переходила дорогу и не садилась за руль велосипеда. А раз она могла быть в принципе участницей ДТП, нулевая вероятность смерти исключается. Сейчас ее можно оценить для детей в возрасте от 5 до 14 лет на уровне 0,0000078 (4 случая, экспозиция 511860) (HMD; UN DATA). Даже при безупречном выполнении всех правил безопасности остаются «внешние» причины этой внешней причины смерти. К ним можно отнести, например, туман, гололед, действия третьих лиц (за рулем другой машины-источника ДТП). Можно, конечно, допустить также, что как сами семилетние девочки, так и их родители особенно внимательны к здоровью ребенка и его поведению на улице. Но даже в этом случае, величину  $q_x$  вряд ли возможно оценить меньше,



чем треть средней величины  $q$  для детей 5–14 лет. В итоге оценка нижней границы величины  $q_x$  в том случае, когда точные расчеты дают ноль, составит (в данном случае)  $q_7 \sim 0,0000026$ .

Отметим, что использование данной величины в таблице смертности не может существенно повлиять на значения интегральных показателей дожития и продолжительности жизни – они, если и изменятся, то лишь в 6–7 знаке после запятой. Значение же  $d_7$  вообще не изменится (останется равным 0). Также сохранится и значение  $l_x$ .

Можно предложить и другую оценку снизу вероятности смерти в случае «абсолютного нуля» в виде величины  $1 / (3 \cdot E_x)$  (экспозиции, числа прожитых человеко-лет в населении под риском). Это будет несколько заниженная, достаточно осторожная оценка  $q_x$ , так как в этом случае величина «пуассоновского» нуля с запасом превышает половину (0,72).

Исходя из общемировых масштабов, может быть также дана еще одна нижняя оценка  $q_x$  вместо «абсолютного нуля». Для этой цели необходимо принять во внимание такие события, которые могут явиться причиной смерти, как землетрясения, удары молнии, наводнения, цунами и другие природные катаклизмы, теракты, а также техногенные катастрофы. Наступления событий, вызванных указанными причинами, практически не зависит от витального поведения человека, уровня его жизни и здоровья. Так как число таких событий на земном шаре исчисляется ежегодно сотнями и даже тысячами, то получаем оценку  $q_x$  снизу в случае «0»: она не может быть меньше  $10^{-7}$  ( $\sim 700/7 \times 10^9$ ).

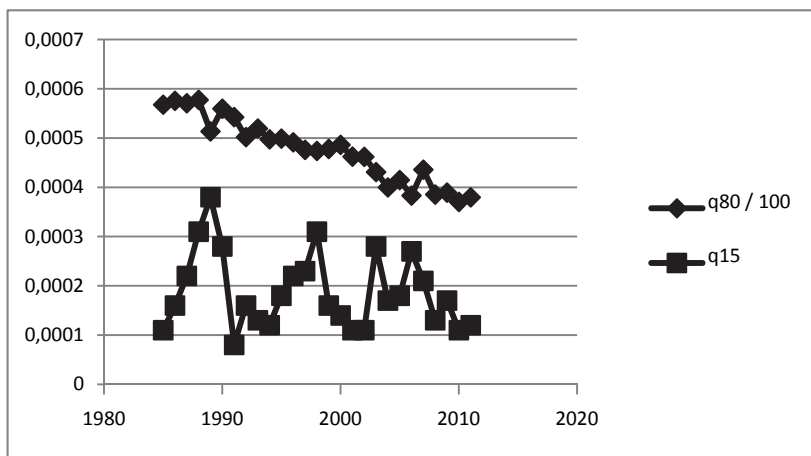
На определенную опасность использования в демографических моделях и расчетах нулевого числа умерших указывают также Toson и Baker (2003). Она состоит в том, что в случае нулевого числа умерших оказывается заниженной также его дисперсия. С содержательной точки зрения такая ситуация, видимо, не может отражать реальные свойства населения. Действительно, чисто формально математически нулевая вероятность для населения в целом означает, что ее значение равно нулю и для любого субнаселения. Таким образом, такие значения будут кардинально расходиться с современными представлениями о неравенстве перед лицом смерти и гетерогенности населения.

И, наконец, последний довод в пользу необходимости избежать публикации  $q = 0$  – это невозможность построить возрастные про-

фили с использованием давно уже устоявшегося логарифмического масштаба.

### *Случайные флуктуации в числе умерших*

По мере снижения смертности в развитых странах все чаще наблюдаются противоречия между общемировыми тенденциями и частными проявлениями в отдельных возрастах. Они связаны со случайными флуктуациями чисел умерших, которые находят свое отражение и в колебаниях возрастных показателей смертности. Для иллюстрации сравним динамику вероятности смерти для шведских женщин 15 и 80 лет.



**Рис. 8.** Динамика вероятности смерти для двух возрастных групп 15 и 80 лет, Швеция, женщины, 1985–2011, масштаб старшей группы уменьшен в 100 раз

*Источник:* HMD.

Если  $q_{80}$  в 1985–2011 гг. отчетливо демонстрирует устойчивую тенденцию к снижению почти в полтора раза, то для  $q_{15}$  характерны многократные значительные повышения, вызванные случайными флуктуациями, а генеральный тренд снижения остается почти незаметным. Более того, существует опасность того, что неопытный студент или недобросовестный исследователь сделают ошибочный вывод о направлении изменения показателя за 25 лет, если сделает сравнение лишь двух значений – начального (0,00011) и конечного (0,00012).

В проблеме флуктуаций чисел умерших нет, конечно, ничего нового. Особенно остро данная проблема встает для малой численности населения страны или отдельных возрастных групп. В случае старших возрастов, например, в Human Mortality Database применяются методы сглаживания демографических показателей (см. Methods Protocol HMD). Для населения малых территорий большие проблемы возникают не только для отдельных возрастов, но даже для важнейших интегральных показателей, таких как ожидаемая продолжительность жизни  $e_x$ , и прежде всего, при рождении ( $e_0$ ) (см., например, Bravo, Malta (2010), Scherbov, Ediev (2011)). Здесь не предполагается обсуждать подходы к решению проблемы достоверности показателей в таблице смертности для малых территорий по причине их достаточной разработанности. Однако, пример сегодняшнего населения Швеции позволяет сделать предположение о постепенном приближении его к категории так называемого населения малых территорий. Но причина этого оказывается несколько иной.

Дело не в малом числе испытаний в терминах теории вероятностей, а в малом риске наступления смерти. При очень маленькой вероятности смерти, если даже число испытаний большое, но недостаточное, закон больших чисел перестает работать и «мгновенная» (в данном возрасте и периоде) частота все хуже отражает истинное (реальное) значение вероятности  $q_x$ . Особенно подозрительной выглядит  $q = 0$ . На рубеже веков данная проблема становится особенно актуальной, поскольку число таких «мгновений» стремительно растет. Однако, что именно нужно сделать, остается пока нерешенным вопросом. Привлечение внимания к данной проблеме и является задачей данной статьи.

### **Обсуждение**

В рамках данной статьи предполагалось не останавливаться на тех хорошо известных искажениях, которые связаны с качеством демографической статистики (проблемы возрастной аккумуляции) или случаев, когда основной причиной случайных флуктуаций демографических показателей является малая величина численности населения. Методы возможного решения возникающих проблем уже давно разрабатываются в демографии. Внимание же предлагается уделить другим остающимся источникам возможных искажений. Отметим, что в отличие от сложнейших технических систем, их причиной не являются помехи, шумы и другие искажения и потери

информации. Если подвести итоги описанных в статье искажений, то можно выделить следующие четыре группы искажений: округление значений показателей, «псевдодискретность», «абсолютный ноль», а также флуктуации, связанные с очень низким значением вероятности смерти.

В чем могут состоять возможные пути решения затронутых проблем? Важнейшее направление – публикация значений  $q_x$  с точностью не менее 7 знаков после запятой, что позволит, по крайней мере, частично, решить проблемы больших относительных ошибок  $m_x$  и  $q_x$  при публикации, их псевдодискретности и псевдоравенства. Пользователи и авторы НМД могут, конечно, возразить, что имеется возможность использования  $m_x$  в разделе НМД Death rates, содержащем в отличие от таблиц смертности значения с 6 знаками после запятой. Это, конечно, так, но может служить лишь временным решением проблемы, так как дает уточнение лишь на порядок, кроме того, НМД содержит данные не по всем странам мира и не для всех периодов. Для возрастных групп с наиболее низкими значениями  $q_x$  с большими относительными флуктуациями чисел умерших требуется применение алгоритмов сглаживания по возрасту и времени основных показателей. Для решения проблемы «абсолютного нуля» может быть предложен вариант замены «0» на величину порядка  $1/(3 \cdot E_x)$ , а также учет неформальных факторов смертности в более широких возрастных группах, в том числе по причинам смерти.

### *Благодарности*

Автор выражает глубокую признательность Д.А. Жданову, А.И. Михальскому и В.Н. Новосельцеву за ценные соображения и замечания по предварительной версии статьи.

### **Литература**

1. Bravo J.M., Malta J. Estimating life expectancy in small population areas // United Nations Statistical Commission and Economic Commission for Europe, Conference of European statisticians, Statistical office of the European Union (Eurostat) Joint Eurostat / UNECE Work Session on Demographic Projections (28–30 April 2010, Lisbon, Portugal) WP 10, 23 April 2010 – <http://www.unecce.org/fileadmin/DAM/stats/documents/ece/ces/ge.11/2010/wp.10.e.pdf>
2. Human Mortality Database – <http://www.mortality.org/>
3. Oeppen J., Vaupel J.W. Broken limits to life expectancy // Science. 2002. Vol. 296. P. 1029–1031.

---

4. Scherbov S., Ediev D.M. Significance of life table estimates for small populations: Simulation-based study of standard errors // Demographic Research. 2011. № 24(22). P. 527–550.

5. Shkolnikov V.M., Andreev E.M., Jdanov D., Jasilionis D., Kravdal Ø., Vågerö D., Valkonen T. Increasing absolute mortality disparities by education in Finland, Norway and Sweden, 1971–2000 // Journal of Epidemiology and Community Health. 2012. Vol. 66, № 4. P. 372–378.

6. UN DATA. Deaths by cause of death, age and sex (WHO data) – <http://data.un.org/Data.aspx?d=POP&f=tableCode%3A105>

7. Toson B., Baker A. Life expectancy at birth: methodological options for small populations // National Statistics Methodological Series. 2003. № 33.