

ИССЛЕДОВАНИЕ МЕТОДОВ АНАЛИЗА ВРЕМЕНИ ДО СОБЫТИЯ ПРИ ОБРАБОТКЕ ДЕМОГРАФИЧЕСКИХ ДАННЫХ

Г. Г. Рапаков, В. А. Горбунов

Вологодский государственный университет

Поступила в редакцию 25.10.2015 г.

Аннотация. В статье исследованы алгоритмы и методы анализа времени до события. Проведено сопоставление непараметрического и параметрического подходов при обработке данных демографических наблюдений. Параметрический метод дает возможность эффективно выявлять аномалии в демографических показателях и хранить данные типа времени жизни в информационных системах. Результаты анализа времени до события использованы для поддержки принятия управленческих решений в муниципальной концепции активного долголетия.

Ключевые слова: анализ времени до события, метод Каплана–Мейера, параметрический метод, поддержка решений, активное долголетие.

Annotation. The article presents research results of the survival analysis algorithms and methods. Comparison of nonparametric and parametric approaches was introduced. The parametric method for time-to-event analysis provides effective storage of survival times data for information systems and demographic anomaly detection. Results were used for decision support of municipal active longevity project.

Keywords: time-to-event analysis, Kaplan and Meier method, parametric method, decision support, active longevity.

ВВЕДЕНИЕ

Данные типа времени жизни, с которыми оперирует анализ выживаемости (survival analysis), характеризуются системным событием. Первоначально, в задачах страхования и здравоохранения в качестве него выступала продолжительность жизни. В настоящее время в связи с расширением сферы приложений традиционного анализа выживаемости в литературе все чаще используется термин – событийный анализ (time-to-event analysis – анализ времени до события). Время наступления события T изучается в задачах страхования и экономики, социологии и демографии, медицины и надежности. Случайная величина T является зависимой переменной, неотрицательное значение которой определяется объясняющими переменными – вектором ковариат $x = (x_1, x_2, \dots, x_m)^T$.

Прогнозирование выживаемости (отказов) позволяет определить вероятность $S(t)$ пережить момент времени t с начала наблюдения [1, 2, 3]:

$$S(t) = P(T \geq t) = 1 - F(t) = \int_t^{\infty} f(x) dx = \\ = \exp\left(-\int_0^t h(x) dx\right) = \exp(-H(t)).$$

Функции выживаемости $S(t)$, кумулятивного распределения отказов $F(t)$, плотности распределения $f(t)$, мгновенной интенсивности $h(t)$ и кумулятивной интенсивности отказов $H(t)$ взаимосвязаны:

$$S(t) = 1 - F(t), \quad H(t) = -\log S(t),$$

$$f(t) = -d[S(t)]/dt, \quad h(t) = -d[\log S(t)]/dt.$$

Особенностью анализа времени до события является проблема выбытия. На практике анализ завершается до наступления исхода у всех объектов, принимающих участие в исследовании. Наряду с полными наблюдения-

ми существуют неполные, изъятые или цензурированные случаи, когда исследователь теряет объект из вида и располагает лишь частью информации о времени его жизни. Эти сведения необходимо использовать. Цензурированные справа наблюдения имеют вид [1, 2, 3]:

$$(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n),$$

где n – объем выборки ($i = 1, \dots, n$); t_i – время до системного события T_i или цензурирования C_i ; $t_i = \min(T_i, C_i)$; δ_i – индикатор цензурирования, $\delta_i = 1$, если i -ое событие – исход и $\delta_i = 0$, если i -ое событие цензурировано.

Изучение и анализ источников позволил выполнить сопоставление целей и методов исследования с данными аналогичной литературы. Выбор параметризации базовой функции распределения для случая ускоренных испытаний при помощи компьютерного моделирования рассматривается в работе [1]. Разработчики автоматизированных информационных систем (АИС), формируя концепцию системы анализа эффективности лечения, ограничиваются непараметрическим подходом при выборе метода анализа выживаемости [4]. Метод Каплана–Мейера предлагается авторами в силу широкой распространенности. Дополнительным аргументом при его выборе, по-видимому, является относительная простота программной реализации с использованием рассматриваемых в этой работе интерфейса анализа данных и технологий Windows Presentation Foundation и ActiveX Data Objects .NET Entity Framework. Разработчики не уделили достаточного внимания возможностям использования альтернативных методов событийного анализа, что создает предпосылки проблем, связанных с хранением данных в АИС. Программная технология экспорта данных в статистические пакеты при помощи файлов формата `comma-separated values` позволяет расширить возможности АИС и частично преодолеть естественные ограничения разработки. В этом случае резко возрастают требования к врачу-исследователю при работе со статистическим инструментарием. По сведениям Минздрава РФ в автоматизации лечебно-диагностического процесса задействовано не более 20% вычис-

лительных мощностей медицинских учреждений. Оценки параметров теоретического распределения в зависимости от его вида представляют собой 1–2 числа. Их хранение и использование в рамках модели сущность–атрибут–значение (`entity–attribute–value`) является предпочтительным по сравнению с массивом значений, определяемых объемом выборки, в методе Каплана–Мейера и влияет на размер базы данных и быстродействие СУБД. Применение параметрического подхода способствует решению проблемы хранения разреженных данных, что может быть отнесено к дальнейшим направлениям работы исследователей и программистов АИС.

В статье [5] сопоставляются ретроспективный событийный анализ и проспективный подход к сбору и анализу данных. В целях предупреждения манипуляций с данными для доказательных клинических исследований рекомендуется применять проспективный метод. Делается вывод о возможности использования ретроспективного событийного анализа лишь в задачах поиска факторов риска и описательной оценки выживаемости по собранным данным. Для имитации ошибок в ретроспективном событийном анализе было использовано компьютерное моделирование с помощью макроязыка пакета статистического анализа SAS 9.1.3. Однако симулирование было ограничено базовым непараметрическим методом Каплана–Мейера. Возможности параметрического подхода в модельной симуляции сбора и обработки данных, выявления типовых методологических ошибок в анализе выживаемости не рассматривались. Использование теоретических распределений и расчет оценок параметров позволит автоматизировать выявление аномалий, заблуждений и фальсификаций при выходе значений оценок за допустимые пределы.

Вместе с тем, зарубежные исследователи в течение длительного времени широко применяют параметрический метод. Автор статьи [6], используя два базовых распределения, зависящих от параметров, и их инверсии, выполнил моделирование закона смертности на основе информационных данных

американских таблиц продолжительности жизни людей в возрасте до 90 лет. В работах [7–9] представлен математический аппарат и приложения параметрического подхода к задачам надежности, анализа выживаемости, качества жизни, генетики, медицины и финансов.

В связи с этим актуальной является задача изучения возможностей применения параметрического метода в анализе времени до события. Важность и практическая значимость работы заключается в том, что полученные оценки параметров позволили выявить гетерогенность групп пенсионной когорты при реализации модели единой системы медико-социальной поддержки граждан пожилого возраста в Вологодской области и муниципальной концепции активного долголетия. Цель работы состоит в исследовании алгоритмов и методов анализа выживаемости и обосновании их использования в сложном прикладном объекте исследования, включая вопросы аналитической поддержки принятия управленческих решений. Научная новизна работы заключается в сопоставлении непараметрического и параметрического подходов в анализе времени до события. Распределение наилучшего выбора позволяет эффективно хранить данные типа времени жизни в АИС, сокращать размер баз данных и повышать быстродействие СУБД. Использование параметрического метода распределений времен жизни дает возможность динамически оценивать эффективность мероприятий концепции активного долголетия и выявлять аномалии в демографических показателях.

МАТЕРИАЛЫ И МЕТОДЫ

В случае цензурированных данных типа времени жизни для оценки функции выживаемости $S(t)$ используются специальные статистические методы анализа выживаемости, отличающиеся от обычных дескриптивных статистик [10, 11]. Выделяют три группы подобных методов: непараметрический, полупараметрический и параметрический [12–14].

Непараметрический метод не требует знания априорной информации о регрессионной

функции и законе распределения результатов измерений. В большинстве случаев непараметрического оценивания $S(t)$ используют построение таблиц жизни и метод множительных оценок Каплана–Мейера. Оценка Каплана–Мейера:

$$\hat{S}(t) = \prod_{j:t_j \leq t} \frac{n-j}{(n-j+1)^{\delta_j}},$$

где n – общее число событий; j – его номер; $\delta_j = 1$, если j -ое событие – исход; $\delta_j = 0$, если j -ое событие цензурировано. Полупараметрический метод позволяет выполнить построение кривой дожития для сочетания прогностических факторов на основе регрессии пропорциональных рисков Кокса.

Параметрический метод анализа выживаемости исходит из предположения об определенной форме кривой $S(t)$. В ходе процедуры подгонки теоретического распределения (fitting – подбор распределения) к данным непараметрического исследования продолжительности жизни используются следующие базовые виды распределений, зависящие от параметров (представлены функции выживаемости $S(t)$, плотности распределения $f(t)$ и интенсивности отказов $h(t)$):

экспоненциальное (exponential)

$$S(t) = e^{-\lambda t}, \quad h(t) = \lambda,$$

Гомпертца (Gompertz)

$$S(t) = \exp(-\lambda \gamma^{-1} (e^{\gamma t} - 1)), \quad h(t) = \lambda e^{\gamma t},$$

лог-логистик (loglogistic)

$$S(t) = (1 + (\lambda t)^{1/\gamma})^{-1}, \quad f(t) = \frac{\lambda^{1/\gamma} t^{1/\gamma-1}}{\gamma (1 + (\lambda t)^{1/\gamma})^2},$$

Вейбулла (Weibull)

$$S(t) = \exp(-\lambda t^p), \quad h(t) = p \lambda t^{p-1},$$

лог-нормальное (lognormal)

$$S(t) = 1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right),$$

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2\sigma^2}(\log(t) - \mu)^2\right),$$

гамма (generalized gamma)

$$S(t) = \begin{cases} 1 - I(\gamma, u), & k > 0 \\ 1 - \Phi(z), & k = 0, \\ I(\gamma, u), & k < 0 \end{cases}$$

$$f(t) = \begin{cases} \frac{\gamma^\gamma}{\sigma t \sqrt{\gamma} \Gamma(\gamma)} \exp(z \gamma^{1/2} - u), & k \neq 0 \\ \frac{1}{\sigma t \sqrt{2\pi}} \exp(-z^2/2), & k = 0 \end{cases},$$

где $\Phi(z)$ – интегральная функция стандартного нормального распределения:

$$\Phi(z) = P[Z \leq z] = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.$$

Для обобщённого гамма-распределения:

$$\gamma = |k|^{-2}, \quad u = \gamma \cdot \exp(|k|z),$$

$$z = \text{sign}(k) \cdot (\log(t) - \mu) / \sigma.$$

$I(a, x)$ – неполная гамма-функция:

$$I(a, x) = \frac{1}{\Gamma(a)} \int_0^x e^{-t} t^{a-1} dt, \quad \Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt.$$

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ

Для создания условий, обеспечивающих повышение качества жизни и долголетие населения, необходимо формировать единое здоровьесберегающее пространство. В целях эффективного решения социально-экономических проблем следует учитывать и предупреждать процессы демографического старения. Рост доли пожилых людей в общей численности населения ведет к повышению иждивенческой нагрузки на его занятую часть. На муниципальном уровне в регионе вопросам долголетия и демографии посвящен ряд нормативно-правовых актов [15].

Методы анализа данных показали высокую эффективность при информационно-аналитической поддержке управленческих решений в задачах здравоохранения [16–19]. Авторами статьи выполнено мониторинговое медико-социологическое исследование. В целях удобства анализа все лица старшего возраста начали наблюдаться одновременно. Данные выборки цензурированы справа, ее объем обеспечивает ошибку выборки, которая не превышает 5 % с доверительной вероятностью $\alpha = 0.95$. Собранные сведения удовлетворяют базовым требованиям исследования выживаемости: для всех наблюдений известно время начала и окончания; тип на-

блюдения – полное или изъятое; наблюдения являются независимыми; выбор объектов случаен; для предотвращения систематической ошибки в виде смещения результата когорты объектов синхронизирована; отсутствие популяционных катаклизмов обеспечивает естественную равномерность выбытия [20]. Статистическая обработка данных выполнена при помощи пакета программ Stata.

Итоги применения параметрического метода исследования выживаемости в группах для мужской и женской части пенсионной когорты представлены в графическом виде на рис. 1. Выбор между различными распределениями осуществляется на основе критерия отношения правдоподобия (log likelihood (LL)) и информационного критерия Акаике (Akaike's information criteria – AIC).

Логарифмическая функция правдоподобия для параметрического метода имеет вид:

$$\log(L(x; \beta, \theta)) = \sum_{i=1}^n \left[\delta_i \left(\log r(x^i; \beta) + \log h_0(t_i; \theta) - r(x^i; \beta) H_0(t_i; \theta) \right) \right],$$

где x – вектор ковариат $x = (x_1, x_2, \dots, x_m)^T$; β – вектор параметров регрессии; θ – вектор параметров $\theta = (\theta_1, \theta_2, \dots, \theta_m)^T$; n – объем выборки ($i = 1, \dots, n$); $\delta_j = 1$, если j -ое событие – исход и $\delta_j = 0$, если j -ое событие цензурировано; $r(x; \beta)$ – неотрицательная функция от ковариат; $h_0(t; \theta)$ – базовая функция риска; $H_0(t; \theta)$ – базовая кумулятивная функция риска; t_i – время до системного события или наступления цензурирования [2].

Информационный критерий Акаике определяется как:

$$AIC = -2 \cdot (\log \text{likelihood}) + 2 \cdot (c + p + 1),$$

где c – количество ковариат, p – число вспомогательных параметров, определяемых видом используемого базового распределения [14].

Использование AIC обычно дает более точные результаты. Информационный характер критерия базируется на концепции Кульбака–Лейблера. AIC позволяет объединить требования качества подгонки распределения и минимизации числа параметров и выражается на основании меры ошибки модели

и штрафа, определяемого числом ее параметров. Итоги расчета индексов качества подбора распределений для мужской и женской групп приведены в табл. 1 и табл. 2 соответственно. Выбирается распределение, для которого значение логарифмического правдоподобия LL максимально (с учетом знака), а критерий Акаике AIC принимает минимальное значение [12, 13, 14].

Для мужской группы наибольшее значение логарифмического правдоподобия было обнаружено в случае гамма-распределения, однако лог-нормальное является предпочтительным с точки зрения критерия Акаике, который принимает в ее случае минимальное значение, впрочем, слабо отличающееся от

AIC для распределения гамма. На основании значений критериев подбора распределений (табл. 2) для женской группы выбрано параметрическое распределение выживаемости гамма.

При этом лог-нормальное распределение уступает гамма и находится на втором месте по величине значений LL и AIC. Обратим внимание, что значение критерия Акаике для экспоненциального распределения времен жизни является достаточно близким к AIC для лог-нормального распределения.

Особенностью экспоненциального распределения является то, что коэффициент смертности для него не зависит от времени наблюдения. Это свидетельствует о слабой

Таблица 1

Критерии подбора распределений для мужской группы

Распределение	Экспоненциальное	Вейбулла	Гомпертца	Лог-нормальное	Лог-логистик	Гамма
Log likelihood	-343,7	-340,4	-338	-336,9	-339	-336,4
AIC	689,4	684,7	680	677,8	682	678,8

Таблица 2

Критерии подбора распределений для женской группы

Распределение	Экспоненциальное	Вейбулла	Гомпертца	Лог-нормальное	Лог-логистик	Гамма
Log likelihood	-359,8	-359,8	-359,6	-358,1	-359,7	-356,9
AIC	721,7	723,6	723,1	720,1	723,3	719,7

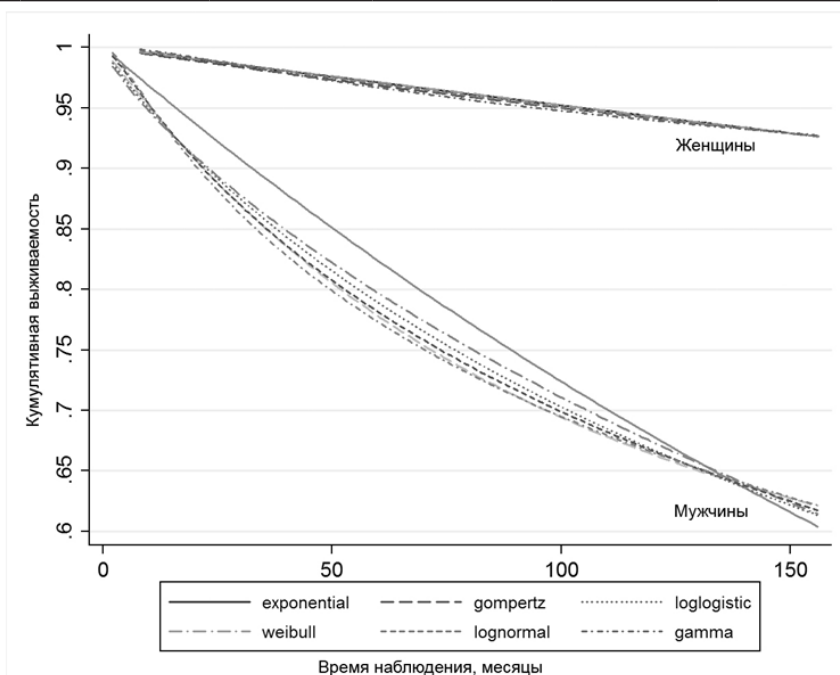


Рис. 1. Параметрические распределения выживаемости в группах

вариабельности (изменчивости) летальности в женской группе на достаточно продолжительном горизонте оценки.

Для проверки адекватности распределений также применяют метод анализа распределения остатков. На основе пропорциональных интенсивностей Кокса вычисляют остатки Кокса–Снелла (CS) [2, 21]:

$$CS_i = H_0(t_i) \exp(\beta x_i).$$

В том случае, если распределение удовлет-

ворительно описывает данные, остатки принадлежат базовому закону распределения, стандартизованному по единичному параметру масштаба. При компьютерном моделировании, используя CS как переменную времени, рассчитывают оценку функции выживания на основе метода Каплана–Мейера и логарифмируют результат, меняя знак выражения. Для расчета стандартной ошибки выживаемости используют формулу Гринвуда [14]:

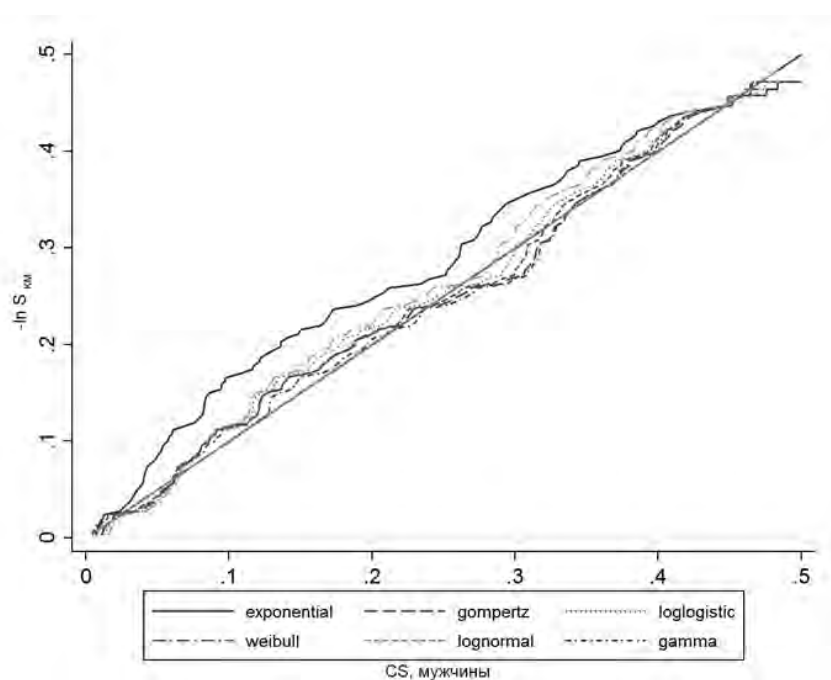


Рис. 2. Подбор распределения для мужской группы

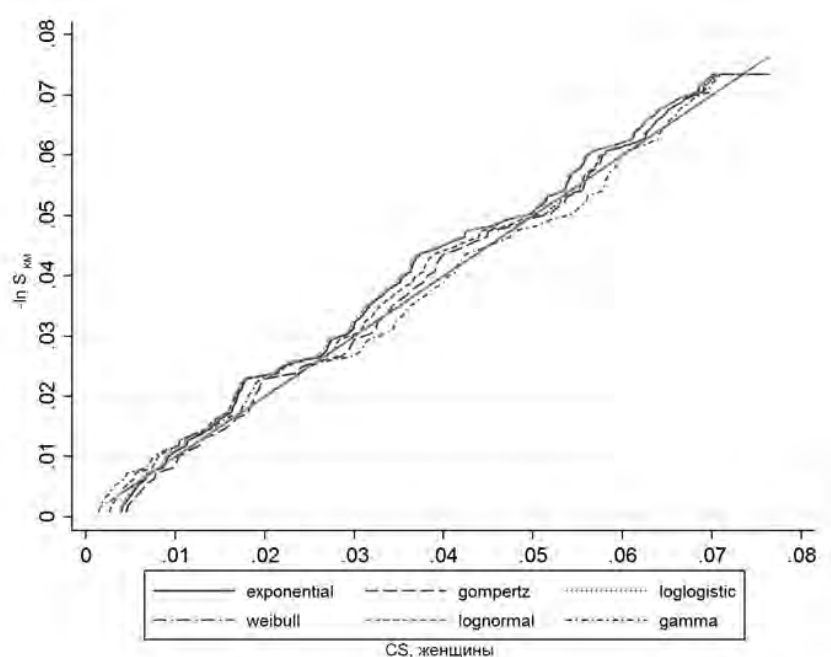


Рис. 3. Подбор распределения для женской группы

$$\widehat{Var}\left(\widehat{S}(t)\right) = \widehat{S}^2(t) \sum_{j:t_j \leq t} \frac{d_j}{n_j(n_j - d_j)},$$

где j – номер системного события, наступившего в момент времени t_j ; d_j – число исходов в t_j ; n_j – число объектов под наблюдением к t_j .

Для распределения, которое успешно согласуется с данными мониторинга, график функции совокупного риска $H(CS) = -\ln(S_{km})$ должен аппроксимировать прямую линию с единичным наклоном. Визуальная оценка от-

клонений шести параметрических распределений для мужской (рис. 2) и женской части (рис. 3) пенсионной когорты от линии, идущей под углом 45° , согласуется с выводами, полученными при расчете значений логарифмического правдоподобия и критерия Акаике.

Результаты расчета параметров отобранных распределений для мужской и женской части пенсионной когорты приведены в табл. 3. Динамическая оценка параметров функций выживаемости в группах позволит выполнить мониторинг эффективности

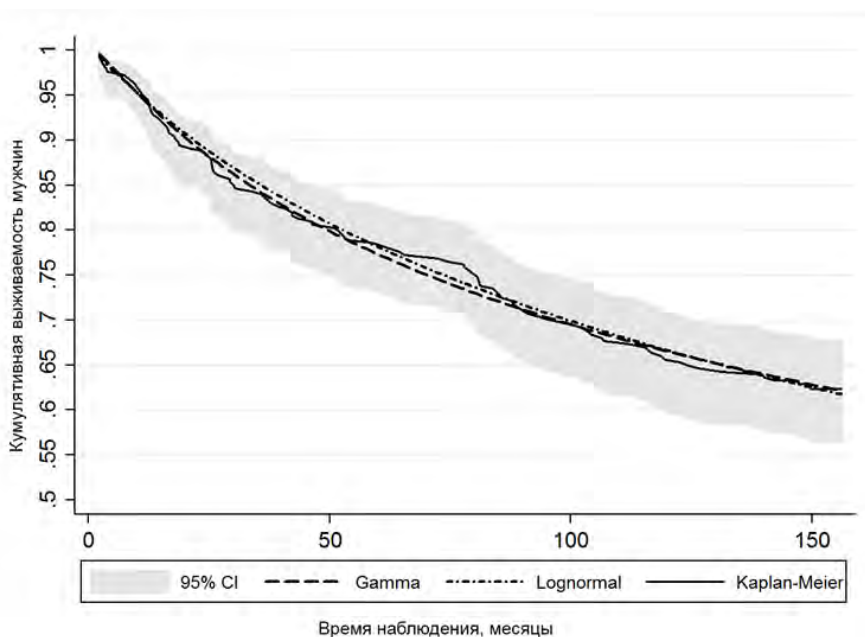


Рис.4. Функции выживаемости для мужской группы

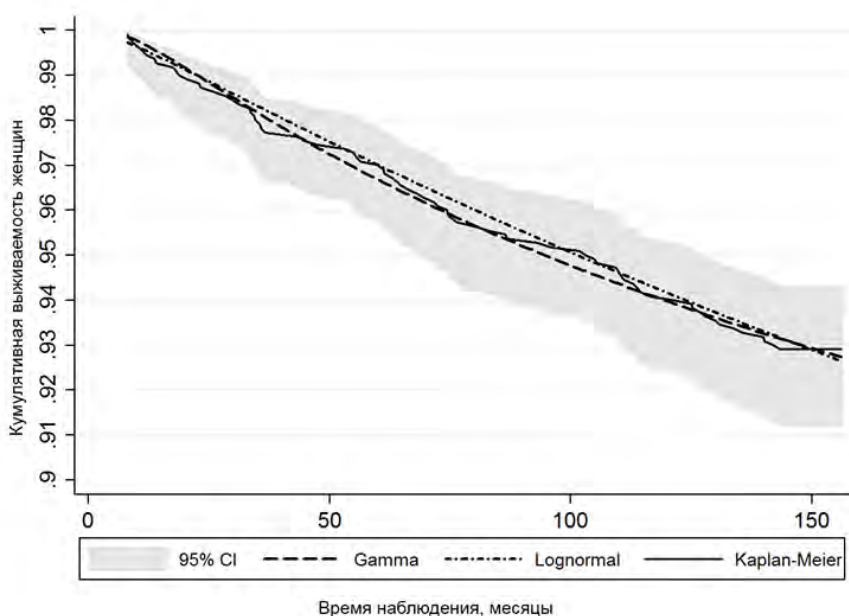


Рис.5. Функции выживаемости для женской группы

Таблица 3

Оценки параметров распределений

Параметры	Лог-нормальное		Гамма	
	μ , 95 % ДИ	σ , 95 % ДИ	μ , 95 % ДИ	σ , 95 % ДИ
Мужчины	5,7 (5,3; 5,9)	1,9 (1,7; 2,3)	5,3 (4,5; 6,2)	2,3 (1,8; 2,9)
Женщины	8,3 (7,5; 8,9)	2,2 (1,8; 2,7)	7,6 (5,2; 9,9)	4,3 (3; 6,1)

Таблица 4

Процентили выживаемости пенсионной когорты

Время выживания	5 % проц., 95 % ДИ	7 % проц., 95 % ДИ	10 % проц., 95 % ДИ	13 % проц., 95 % ДИ
Когорта, месяц	35,8 (26; 50,7)	61,2 (42,2; 77,8)	95,2 (78; 115,9)	140,1 (111,7; NA)

Таблица 5

Процентили выживаемости мужской и женской части пенсионной когорты

Время выживания	5% проц., 95% ДИ	7% проц., 95% ДИ	10% проц., 95% ДИ	37% проц., 95% ДИ
Мужчины, месяц	11,6 (7,3; 15,1)	13,2 (9,9; 18,6)	18,6 (12,9; 26)	147,2 (104,2; NA)
Женщины, месяц	105,1 (69,6; 127,9)	143,4 (110,9; NA)	NA	NA

адресных мероприятий в рамках концепции активного долголетия на протяжении всего периода ее реализации – по итогам каждого отчетного года и в целом.

В результате применения метода Каплана–Мейера к данным мониторинга были получены кривые дожития для мужской и женской части пенсионной когорты (рис. 4, 5). Кривые монотонны и не пересекаются. Выполнена проверка нулевой гипотезы H_0 об отсутствии статистически значимых различий выживаемости в группах на основе критерия Мантела-Кокса. По результатам проверки нулевая гипотеза была отвергнута на уровне значимости $p < 0,001$.

Для сравнения выживаемости в группах во времени наблюдения рассчитаны p -процентили, определяющие время, в течение которого исход наступит у p % участников исследования, и их 95 % доверительные интервалы (ДИ, confidence interval – CI) (табл. 4, 5). Множительная оценка Каплана–Мейера, в отличие от метода таблиц жизни,

не зависит от группировки времен жизни. Доверительные интервалы на графиках (рис. 4, 5) расширяются по мере наблюдения, а сама доверительная область образует «рукав». Это обусловлено выбытием наблюдаемых лиц и соответствующим ростом ошибки оценки выживаемости.

Процентили выживаемости пенсионной когорты, полученные на основе оценки функции выживания, свидетельствуют, что 5 % пенсионеров скончаются в течение 35,8 месяца, при 95 % доверительном интервале от 26 до 50,7 (табл. 4). С учетом конечной длительности горизонта оценки 10 % и 37 % процентили выживаемости не представлены (NA) для женской группы (табл. 5), как и часть верхних границ для 95 % ДИ (табл. 4, 5). Сопоставление процентилей выживаемости в группах демонстрирует, что 5 % мужчин проживут после выхода на пенсию менее одного (0,96) года, в то время как аналогичный показатель смертности для женской группы наблюдается по достижении почти 9 (8,75) лет пенсион-

ного стажа. В окрестности верхней границы горизонта оценки: 11,95 года – для женщин и 12,27 лет – для мужчин смертность составляет 7 % и 37 % соответственно. В пенсионной когорте риск смерти для мужской группы существенно и последовательно выше, чем для женской.

Процентили выживаемости пенсионной когорты, ее мужской и женской групп рекомендуется включить в число целевых показателей муниципальной концепции активного долголетия, подлежащих ежегодному мониторингу и позволяющих проводить динамический анализ эффективности ее реализации.

ЗАКЛЮЧЕНИЕ

Выполнено исследование алгоритмов и методов анализа выживаемости. Обосновано их использования в сложном прикладном объекте исследования – единой системе медико-социальной поддержки граждан пожилого возраста в Вологодской области и муниципальной концепции активного долголетия г. Вологды. На основе сравнительного анализа непараметрического и параметрического методов в анализе времени до события проведено исследование продолжительности жизни лиц старше трудоспособного возраста по половому признаку.

Для шести видов распределений: экспоненциального, Вейбулла, Гомпертца, лог-нормального, лог-логистик и гамма выполнен подбор параметров в двух группах пенсионной когорты. Подгонка параметрических распределений осуществлялась с использованием критериев максимального правдоподобия и Акаике, обеспечивающего компромисс между сложностью и точностью оценки выживаемости. Для мужской части пенсионной когорты к распределениям наилучшего выбора отнесены гамма и лог-нормальное распределения. Меры качества подбора распределений для женской группы позволяют сделать согласованный выбор в пользу гамма-распределения.

Выбор в пользу конкретного распределения в группах был подтвержден при помощи метода анализа распределения остатков.

Остатки Кокса–Снелла вычислены для шести параметрических распределений в двух группах. Для обеих групп рассчитаны оценки параметров и их доверительные интервалы.

Получена множительная оценка Каплана–Мейера и построены кривые дожития для мужской и женской части пенсионной когорты. На основе критерия Мантела–Кокса подтверждены различия выживаемости в группах ($p < 0,001$), что свидетельствует о скрытой гетерогенности. Для сравнения выживаемости рассчитаны процентили и их доверительные интервалы. В пенсионной когорте 5 % лиц скончаются через 35,8 месяца после выхода на пенсию. 5 % мужчин проживут на пенсии не более 0,96 года, 5 % женщин – не более 8,75 лет. После 12 лет жизни на пенсии выживаемость женщин составляет 93 %, мужчин – 63 %.

Использование оценок, полученных в результате применения параметрического метода в анализе времени до события, является предпочтительным по сравнению с массивом значений метода Каплана–Мейера. Выигрыш в хранении данных типа времени жизни пропорционален объему выборки и составляет от одного (для случая малых выборок) до двух-трех порядков при проведении мониторинговых медико-социологических исследований. По сравнению с непараметрическим методом, целевые числовые оценки параметров распределений времен жизни позволяют оперативно выявлять аномалии в демографических показателях и контролировать выживаемость при проведении демографических исследований в концепции активного долголетия.

СПИСОК ЛИТЕРАТУРЫ

1. Галанова Н. С. Выбор параметризации базовой функции распределения для АГТ-модели с помощью методов компьютерного моделирования / Н. С. Галанова, Е. В. Чимитова // Актуальные проблемы электронного приборостроения АПЭП–2010: материалы X международной конференции. – 2010. – С. 31–35.

2. Чимитова Е. В. Непараметрические критерии согласия в задачах проверки адекватности моделей надежности по цензурированным данным / Е. В. Чимитова, М. А. Ведерникова, Н. С. Галанова // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2013. – № 4 (25). – С. 115–124.
3. Pease K. Design and Analysis of Clinical Trials with Time-to-Event Endpoints / K. Pease. – Chapman & Hall/CRC, 2009. – 590 p.
4. Мельников М. П. Автоматизированная система анализа эффективности лечения в онкологии / М. П. Мельников, П. Н. Воробкалов, К. Д. Капланов // Современные проблемы науки и образования. – 2012. – № 3. – С. 158–166.
5. Куликов С. М. Анализ выживаемости или событийный анализ: типовые ошибки ретроспективного метода / С. М. Куликов, Е. Н. Паровичникова, В. Г. Савченко // Клиническая онкогематология. Фундаментальные исследования и клиническая практика. – 2010. – № 2. – Том 3. – С. 176–183.
6. Carriere J. F. Parametric models for life tables / J. F. Carriere // Transactions Society of Actuaries. – 1992. – Vol. XLIV. – pp. 77–99.
7. Parametric and Semiparametric Models with Applications to Reliability, Survival Analysis, and Quality of Life / M. S. Nikulin, N. Balakrishnan, M. Mesbah, N. Limnios. – Springer Science + Business Media, LLC, 2004. – 555 p.
8. Chen J. Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance / J. Chen, A. K. Gupta. – 2nd ed. – Springer Science + Business Media, LLC, 2012. – 273 p.
9. Marshall A. W. Life Distributions: Structure of Nonparametric, Semiparametric, and Parametric Families / A. W. Marshall, I. Olkin. – Springer Science + Business Media, LLC, 2007. – 782 p.
10. Флетчер Р. Клиническая эпидемиология. Основы доказательной медицины / Р. Флетчер, С. Флетчер, Э. Вагнер. – Москва: Медиа Сфера, 1998. – 352 с.
11. Реброва О. Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA / О. Ю. Реброва. – Москва: Медиа Сфера, 2006. – 312 с.
12. Liu Xian Survival Analysis: Models and Applications / Xian Liu. – Higher Education Press, 2012. – 446 p.
13. An Introduction to Survival Analysis Using Stata / M. Cleves, W. Gould, R. Gutierrez, Y. Marchenko. – 2nd ed. – Stata Press, College Station, TX, 2008. – 372 p.
14. Stata Survival Analysis and Epidemiological Tables Reference Manual Release 13. – Stata Press, 2013. – 553 p.
15. Вологда – город долгожителей: концепция активного долголетия на территории муниципального образования «Город Вологда» на период до 2035 года: решение Вологодской городской Думы от 29 декабря 2014 г. № 129 // КонсультантПлюс: справ. – правовая система / Компания «КонсультантПлюс».
16. Рапаков Г. Г. Интеллектуальный анализ медико-социологических данных с использованием метода Microsoft Decision Trees / Г. Г. Рапаков, В. А. Горбунов // Вестник Воронеж. гос. ун-та. Сер. Системный анализ и информационные технологии. – 2015. – № 2. – С. 130–137.
17. Рапаков Г. Г. Интеллектуальный анализ данных в здравоохранении региона (на материалах Вологодской области): монография / Г. Г. Рапаков, Г. Т. Банщиков. – Вологда: ВоГУ, 2014. – 79 с.
18. Рапаков Г. Г. Методы и алгоритмы машинного обучения при принятии управленческих решений в региональной системе медицинской профилактики (опыт Вологодской области): монография / Г. Г. Рапаков, Р. А. Касимов. – Вологда: ВоГУ, 2014. – 143 с.
19. Туровский Я. А. Разработка новых информационных медицинских технологий и их использование в учебном процессе и научных исследованиях вуза / Я. А. Туровский, С. А. Запрягаев, С. Д. Кургалин // Врач и информационные технологии. – 2008. – № 3. – С. 70–71.
20. Гланц С. Медико-биологическая статистика / С. Гланц. – Москва: Практика, 1999. – 459 с.

21. *Machin D.* Survival Analysis. A Practical Approach / D. Machin, Y. B. Cheung, M. K. B. Parmar. – 2 nd ed. – John Wiley & Sons, 2006. – 266 p.

Рапаков Георгий Германович – кандидат технических наук, доцент, доцент кафедры информационных систем и технологий, Вологодский государственный университет.
Тел.: +7(8172)729571
E-mail: grapakov@yandex.ru

Горбунов Вячеслав Алексеевич – доктор физико-математических наук, профессор, заведующий кафедрой информационных систем и технологий, Вологодский государственный университет.
Тел.: +7(8172)729571
E-mail: gorbunov1945@inbox.ru

Rapakov G. G. – PhD in Technical Science, Associate Professor, Information Systems and Technologies Department, Vologda State University.

Gorbunov V. A. – Doctor of Physico-Mathematical Science, Professor, Information Systems and Technologies Department, Head of Department, Vologda State University.