

А.В. СУВОРОВА, Т.В. ТУЛУПЬЕВА, А.Л. ТУЛУПЬЕВ
**ОБОБЩЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИОННАЯ МОДЕЛЬ
ДЛЯ ПРОГНОЗА ВРЕМЕННОГО ИНТЕРВАЛА
МЕЖДУ ПОСЛЕДНИМ ЭПИЗОДОМ
РИСКОВАННОГО ПОВЕДЕНИЯ И МОМЕНТОМ ИНТЕРВЬЮ
НА ОСНОВЕ СОЦИАЛЬНО-ДЕМОГРАФИЧЕСКИХ
И ПСИХОЛОГИЧЕСКИХ ОСОБЕННОСТЕЙ**

Суворова А.В., Тулупьева Т.В., Тулупьев А.Л. **Обобщенная линейная регрессионная модель для прогноза временного интервала между последним эпизодом рискованного поведения и моментом интервью на основе социально-демографических и психологических особенностей.**

Аннотация. В данной работе исследуются взаимосвязи между последними эпизодами потребления алкоголя как показателя рискованности поведения и социально-демографическими и психологическими характеристиками респондента. Для этого строится регрессионная модель с интервалом между интервью и последним эпизодом поведения в качестве зависимой переменной. Рассмотрены критерии качества для регрессионных моделей.

Ключевые слова: регрессионный анализ, диагностики модели, рискованное поведение, последние эпизоды.

Suvorova A.V., Tulupyeva T.V., Tulupyev A.L. **Generalized linear regression model for prediction of interval between the interview moment and the last episode of risky behavior on the base of psychosocial and demographic peculiarities.**

Abstract. This paper studies associations between the last alcohol consumption (that can be used as an indicator of a risky behavior rate) and socio-demographical and psychological characteristics of respondents. We construct the appropriate linear regression model with the interval between the last alcohol consumption and an interview as a response variable. We examine the model using such diagnostics as residuals plots, test for outliers test for normality and others.

Keywords: regression analysis, diagnostics, risky behavior, last episodes.

1. Введение. Текущая ситуация с эпидемией ВИЧ в России вызывает серьезную озабоченность специалистов в области общественного здоровья [1, 2]. На данный момент в России официально зарегистрированы 636 979 ВИЧ-инфицированных [3]. Однако среди россиян в возрасте 15-49 лет распространенность ВИЧ составляет 1% [0.9%– 1.2%], и согласно оценкам около 980,000 людей живут с ВИЧ (оцениваемый интервал 840000–1200000) [4]. В Санкт-Петербурге, втором по величине городе России, показатель распространенности ВИЧ один из самых высоких в стране (1001.1 людей, живущих с ВИЧ на 100 тыс. населения) [3].

Таким образом, необходимо уделить особое внимание разработке действенных мер по сдерживанию распространения ВИЧ-инфекции. В подавляющем большинстве случаев заражение происходит при непо-

средственном участии ВИЧ-инфицированных людей. Для того чтобы лучше понять, какие подходы по планированию и проведению поведенческих интервенций среди ВИЧ-инфицированных должны формироваться, необходимо заниматься тем психологическим контекстом, в котором осуществляется их рискованное поведение [5].

Для изучения поведенческих особенностей ВИЧ-инфицированных в СПИИРАН в 2007–2008 гг. было организовано и проведено на базе ГУЗ «Центр по профилактике и борьбе со СПИД и инфекционными заболеваниями» в Санкт-Петербурге полевое исследование, фокусирующееся на их адаптивном потенциале с целью формирования моделей рискованного поведения и оценки его параметров в контексте адапционных стилей и социальных связей личности для последующего обеспечения разработки, рационального планирования и эффективного выполнения программ по поведенческой интервенции среди указанных лиц.

Основной задачей являлось определение наиболее ярко выраженных малоадаптивных способов поведения ВИЧ-инфицированных, которые могут представлять опасность с точки зрения вероятности распространения ВИЧ, что было осуществлено на основе полевого исследования среди ВИЧ-инфицированных [5].

При интервьюировании и тестировании пациентов СПИД-Центром использовался инструментарий [6], общий анализ полученных данных описан в [5, 7-9]. Целью же данной работы является построение регрессионной модели для прогноза временного интервала между последним эпизодом рискованного поведения и моментом интервью на основе социально-демографических и психологических особенностей респондента.

2. Используемые переменные и описание выборки. В данной работе исследуются взаимосвязи между последними эпизодами потребления алкоголя как показателя рискованности поведения и шестью социально-демографическими и психологическими характеристиками респондента. Для этого строится регрессионная модель с интервалом между интервью и последним эпизодом поведения в качестве зависимой переменной. Множество потенциальных предикторов включает (используемое далее в работе обозначение переменной указано в скобках):

- пол: 0 — женский, 1 — мужской (Sex);
- возраст (Age);
- потребность в поиске новых ощущений (New)
- склонность к риску (Risky).
- копинг-поведение, положительная переоценка (Coping). Копинг-поведение возникает, когда человек попадает в проблем-

ную, травмирующую ситуацию. «Копинг» — это индивидуальный способ взаимодействия с ситуацией в соответствии с ее собственной логикой, значимостью в жизни человека и его психологическими возможностями. Положительная переоценка включает в себя усилия по созданию положительного значения с фокусированием на росте собственной личности [5].

- психологическая защита, замещение (Defense). Замещение — это психологическая защита, осуществляющая перенос реакции с недоступного объекта на доступный или замену неприемлемого действия на приемлемое. Замещение выступает как разрядка эмоций на объекты, животных или людей, воспринимаемых индивидом как менее опасные, чем те, которые действительно вызывают эти эмоции [5].

Данные, анализируемые в данной работе, были собраны в 2007 году. Всего в исследовании приняли участие 301 ВИЧ-инфицированный пациент: 135 (44.9%) мужчин и 166 (55.1% женщин). Большая часть ВИЧ-инфицированных находилась в возрастном диапазоне до 35 лет (71.9% мужчин и 78.3% женщин). Каждая анкета с пропущенными данными исключалась из анализа. Таким образом, в данной работе анализировались ответы 252 респондентов.

3. Методы. В качестве зависимой переменной рассмотрим логарифм временного интервала между последним эпизодом и интервью ($\log(AIc)$). Такая трансформация обусловлена несколькими причинами [10]. Во-первых, при таком моделировании зависимая переменная принимает только неотрицательные значения, что обоснованно, так как исследуется временной промежуток, который не может быть меньше нуля. Во-вторых, это позволит приблизить к нормальному распределению этой величины (рис. 1) и распределение остатков модели (рис. 2). Невыполнение условия нормальности остатков линейной модели может привести к ряду проблем, например, оценка параметров методом наименьших квадратов может быть неоптимальной, хотя и останется лучшей среди всех несмещенных оценок; построенные доверительные интервалы также могут оказаться неверными [10].

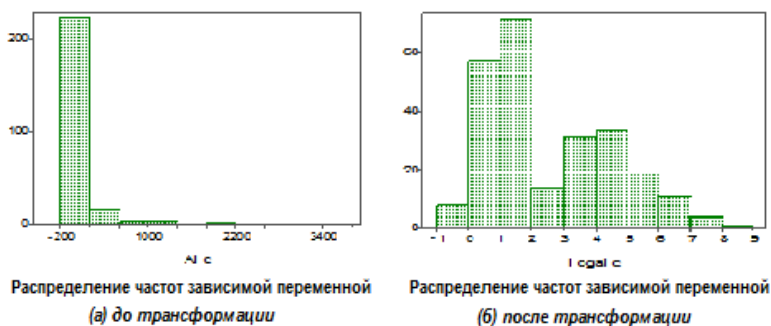


Рис.1. Трансформация зависимой переменной: распределение частот

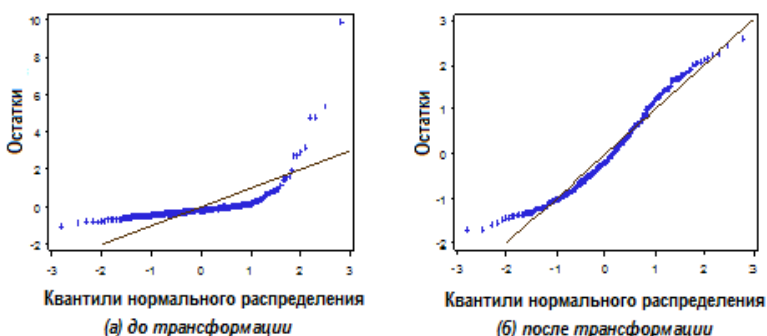


Рис.2. Трансформация зависимой переменной: нормальность остатков

Одной из важных проблем, возникающей при регрессионном анализе, является мультиколлинеарность, то есть наличие линейной зависимости между предикторами — «независимыми переменными модели». Такая зависимость приводит к неустойчивости оценок параметров этой модели, что выражается в увеличении статистической неопределенности — дисперсии оценок [10]. Это означает, что конкретные результаты оценки могут сильно различаться для разных выборок несмотря на то, что выборки однородны, и даже приводить к противоположным выводам при небольшом изменении выборки.

Для обнаружения мультиколлинеарности переменных была проанализирована их корреляционная матрица, что выявило потенциальную зависимость между переменными Risky и New (склонность к риску и потребность в поиске новых ощущений соответственно). Даль-

нейший анализ с помощью коэффициента VIF^1 ($VIF \ll 10$) позволил заключить, что в исследуемой выборке нет мультиколлинеарности.

Таким образом, множество потенциальных предикторов включает все шесть переменных (*Age*, *Sex*, *Risky*, *New*, *Coping*, *Defence*). Однако, лучшая, наиболее объясняющая зависимости модель не обязательно включает все перечисленные переменные. Существует несколько критериев [10], по которым определяется качество модели, например, коэффициент корреляции (R^2), критерий Акаике (AIC), Байесовский критерий (BIC), критерий Маллоуза (C_p). Согласно всем перечисленным выше критериям оптимальным набором предикторов является *Age*, *Coping*, *Risky* (возраст, копинг-поведение и склонность к рискованному поведению). Метод пошаговой регрессии (*stepwise regression*), представляющий собой итеративную процедуру добавления переменных, где на каждом шаге признаки проверяются на возможность добавления в модель или удаления из модели, на основе F-статистики приводит к тому же подмножеству переменных.

Отметим, что следующая по качеству модель включает переменную *New* в дополнение к трем, перечисленным выше. Кроме того, как было показано ранее, переменные *Risky* и *New* имеют достаточно высокий коэффициент парной корреляции. Для проверки влияния указанных факторов была рассмотрена еще одна модель, включающая в себя слагаемое, описывающее взаимодействие между этими переменными (*Risky*New*). Однако качество такой модели оказалось хуже по всем рассмотренным критериям. Формальный тест на проверку гипотезы о необходимости включения взаимодействия *Risky*New* также показал с 99% уровнем значимости² ($\alpha = 0,01$), что модель не включает указанный признак.

Таким образом, модель включает три предиктора *Risky*, *Age*, *Coping*. Оценки параметров модели приведены в табл. 1, т.е. модель имеет вид:

$$\log(Alc) = 0,16753 + 0,03144(Age) + 0,69563(Coping) - 0,01564(Risky)$$

¹ VIF — Variance Inflation Factor (коэффициент возрастания дисперсии)

² Часто уровнем значимости называется α , а не $1-\alpha$, однако это звучит неестественно: хотелось бы, чтобы для более «значимых» результатов уровень значимости был выше. В данной работе уровнем значимости будет называться $1-\alpha$ [11].

Таблица 1. Оценка параметров

| Variable | DF | Estimate | St. Error | t-value | Pr> t |
|-------------|----|----------|-----------|---------|--------|
| (Intercept) | 1 | 0,16753 | 0,62147 | 0,27 | 0,7877 |
| Age | 1 | 0,03144 | 0,01577 | 1,99 | 0,0472 |
| Coping | 1 | 0,69563 | 0,23027 | 3,02 | 0,0028 |
| Risky | 1 | -0,01564 | 0,00877 | -1,78 | 0,0758 |

4. Диагностики регрессионной модели. Построение регрессионных моделей — это многоступенчатый, итерационный процесс. Диагностика регрессионных моделей позволяет обнаружить несоответствие модели данным и наметить пути для дальнейшего улучшения построенной модели. Важный аспект диагностики регрессионной модели — выявление выбросов (outliers) и влияющих наблюдений (influential cases). Выбросом называется точка, значение которой выделяется из общей выборки; влияющим называется наблюдение, удаление которого из структуры данных влечет за собой существенные изменения в параметрах модели.

Согласно тесту Бонфферони с 90% уровнем значимости ($\alpha = 0,1$) в данной модели нет выбросов среди значений зависимой переменной. Однако есть ряд выбросов среди значений предикторов (рис. 3). В последнем случае выбросами считаются точки, для которых значение показателя разбалансировки (leverage) больше, чем удвоенное среднее значение показателя разбалансировки по всей выборке (горизонтальная линия на рис. 3) [10].

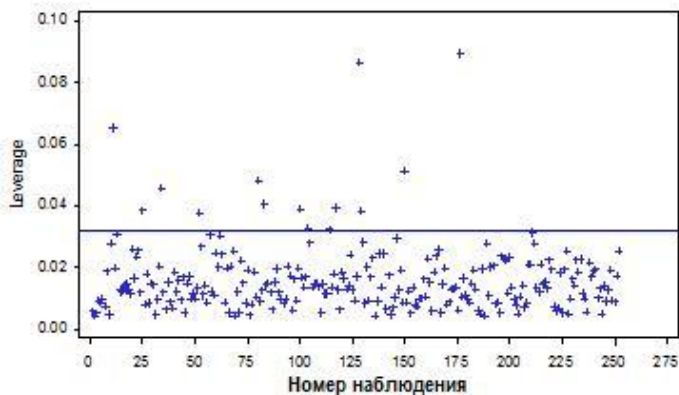


Рис. 3. Определение выбросов среди наблюдений.

Следует отметить, что выбросы не обязательно оказывают существенное влияние на модель в отличие от влияющих наблюдений. Для проверки, является ли выброс влияющим наблюдением, применяется ряд критериев — DFFITS, DFBETAS и расстояние Кука [10]. Согласно всем этим критериям в выборке всего три влияющих наблюдения: 11-е, 83-е и 114-е. Существуют различные методы, позволяющие снизить влияние этих наблюдений на модель, но в данной работе был применен один из самых простых способов — удаление указанных наблюдений из выборки. Такое удаление обоснованно, т.к. общий размер выборки ($n = 252$) достаточно велик, а кроме того, модель (2), построенная по данным без выделенных влияющих наблюдений, отличается от рассмотренной ранее модели (1) только небольшим изменением в значении параметров:

- полная выборка

$$\log(\text{Alc}) = 0,1675 + 0,0314(\text{Age}) + 0,6956(\text{Coping}) - 0,0156(\text{Risky}) \quad (1)$$

- без наблюдений 11, 83 и 114

$$\log(\text{Alc}) = 0,1517 + 0,0309(\text{Age}) + 0,6995(\text{Coping}) - 0,0151(\text{Risky}) \quad (2)$$

Далее исследуется именно модель (2).

Наиболее простым и в то же время эффективным средством диагностики регрессионной модели является построение графиков распределения остатков. Так, при анализе графиков (рис. 4), соответствующих удаленным из модели переменным *Defence* и *New* (психологическая защита и потребность в поиске новых ощущений), не было отмечено видимых зависимостей, то есть нет оснований считать, что эти предикторы оказывают влияние на зависимую переменную.

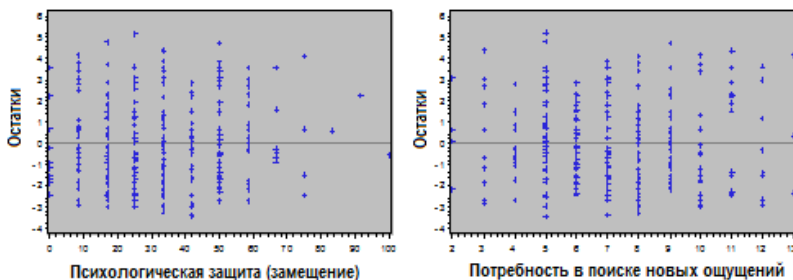


Рис. 4. Графики остатков для исключенных из модели переменных

Значимые зависимости не были выявлены и при анализе влияния взаимодействия переменных *Age*, *Risky*, *Coping*, а также при анализе возможной нелинейной зависимости для какой-либо из этих переменных.

ных. Таким образом, модель (2) не требует включения дополнительных слагаемых, описывающих нелинейность или взаимодействие переменных.

Как уже отмечалось ранее, для адекватности оценок параметров регрессионной модели, а как следствие, и выводов относительно этой модели должно быть выполнено условие нормальности распределения остатков. Для проверки на нормальность строится график Q-Q plot [10]. На графике (рис. 5) точки, соответствующие упорядоченным и нормированным остаткам модели, близки к прямой линии, что свидетельствует о нормальном распределении остатков.

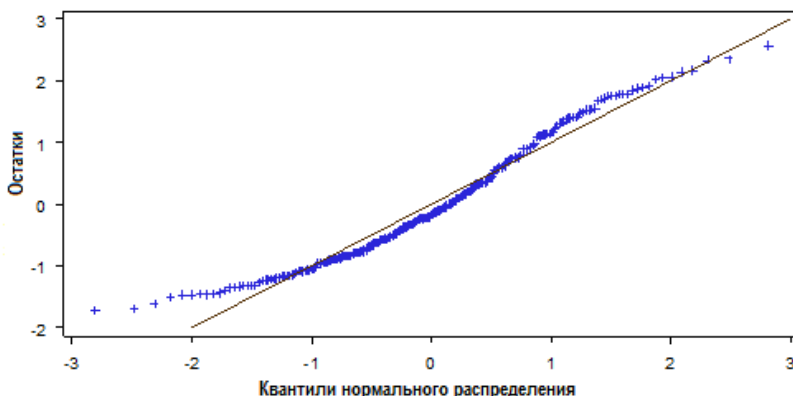


Рис. 5. Нормальность распределения остатков

Формальный корреляционный тест на нормальность также позволяет сделать вывод с 99% уровнем значимости ($n = 249, \alpha = 0,01$), что распределение остатков приближено к нормальному.

Отметим, что при построении линейной регрессионной модели мы делаем ряд предположений: о линейности модели, о нормальном распределении и независимости остатков и о том, что остатки имеют одинаковое распределение, в частности, одинаковую дисперсию [10]. Соблюдение последнего из условий можно проверить как по графику распределения остатков, так и по формальному критерию — тесту Брауна-Форсайта. В основе указанного критерия лежит идея о разделении данных на две группы согласно уровню одного из предикторов, входящих в модель, а затем сравнении дисперсий остатков в этих группах, что позволяет определить, изменяется ли дисперсия в зависимости от значения этого предиктора [10]. Все три теста (по уровню переменных Age, Risky и Coping соответственно) с 95% уровнем значимости показывают, что дисперсия остатков постоянна при изменении значений переменных-предикторов.

Таким образом, модель удовлетворяет всем условиям, необходимым при построении линейной регрессионной модели. Последний этап анализа модели включает тест на регрессию (Test of regression relation). Другими словами, это тест на проверку гипотезы о существовании регрессионного отношения (регрессионной связи) между зависимой переменной и предикторами Age, Coping, Risky. В рассматриваемой модели можно с 99% уровнем значимости утверждать, что как минимум один из коэффициентов β_1, β_2 и β_3 не равен нулю, то есть зависимость между давностью последнего потребления алкоголя и возрастом, склонностью к риску и положительной переоценкой существует.

Однако, коэффициент множественного определения $R^2 = \frac{SSR}{SSTO} = 0.0656$ показывает, что данная модель объясняет только небольшую долю (6,56%) поведения (вариации) зависимой переменной. Включение в модель всех шести социально-демографических и психологических показателей также не улучшает ситуацию: значение коэффициента R^2 увеличивается лишь на 0.003, в то время как нормированный по количеству переменных коэффициент R^2_{adj} уменьшается значительно.

5. Интервальные оценки параметров. При построении модели были получены точечные оценки параметров, в этом разделе будет уделено внимание построению интервальных оценок. Причем интервал будет строиться не отдельно для каждого параметра, а совместно для всех четырех параметров. Точечные оценки параметров модели приведены в табл. 2.

Таблица 2. Оценка параметров измененной модели

| Variable | DF | Estimate | St. Error | t-value | Pr> t |
|-------------|----|----------|-----------|---------|--------|
| (Intercept) | 1 | 0,15171 | 0,62030 | 0,24 | 0,8070 |
| Age | 1 | 0,03092 | 0,01611 | 1,92 | 0,0561 |
| Coping | 1 | 0,69952 | 0,22845 | 3,06 | 0,0024 |
| Risky | 1 | -0,01513 | 0,00874 | -1,73 | 0,0846 |

Для получения интервальных оценок сначала вычисляется совместное критическое значение Бонферрони с 90% уровнем значимости $B = t(1 - \alpha/2g, n - p) = t(1 - 0.1/8, 249 - 4) = t(0.9875, 245) = 2.255262$.

Тогда интервалы имеют следующий вид [10]:

$$\beta_0 : b_0 \pm Bs\{b_0\} = 0.15171 \pm 2.255262(0.62030) = [-1.24, 1.55],$$

$$\beta_1 : b_1 \pm Bs\{b_1\} = 0.03092 \pm 2.255262(0.01611) = [-0.005, 0.067],$$

$$\beta_2 : b_2 \pm Bs\{b_2\} = 0.69952 \pm 2.255262(0.22845) = [0.184, 1.215],$$

$$\beta_3 : b_3 \pm Bs\{b_3\} = -0.01513 \pm 2.255262(0.00874) = [-0.035, 0.0045].$$

Другими словами, с 90% уверенностью можно утверждать, что все значения параметров модели лежат в соответствующих интервалах одновременно. Отметим, что взаимосвязи очень слабые, т.к. почти все интервалы содержат ноль, наиболее сильное влияние оказывает переменная *Coping* (копинг-поведение, положительная переоценка).

6. Заключение. Построенная регрессионная модель имеет вид:

$$\log(Alc) = 0,1517 + 0,0309(Age) + 0,6995(Coping) - 0,0151(Risky)$$

или

$$Alc = \exp(0,1517 + 0,0309(Age) + 0,6995(Coping) - 0,0151(Risky)).$$

Согласно исследуемым данным время между интервью и последним эпизодом потребления алкоголя взаимосвязано с возрастом респондента, такой копинг-стратегией как положительная переоценка и склонностью к риску, причем эта связь позитивна для первых двух показателей и негативна для последнего. Отметим, что эти взаимосвязи очень слабые, почти все интервальные оценки содержат ноль; самая сильная по сравнению с остальными взаимосвязь выявлена между зависимой переменной и копинг-стратегией. Построенная модель является лучшей, *описывающей* взаимосвязи между давностью потребления алкоголя и шестью социально-демографическими и психологическими характеристиками респондента, но неудобна для прогнозирования: данная модель объясняет только небольшую долю (6,56%) поведения (вариации) зависимой переменной. Включение в модель всех шести социально-демографических и психологических показателей также не улучшает ситуацию. Для построения более подходящих моделей, описывающих зависимости между эпизодами поведения и социально-демографическими и психологическими характеристиками респондентов, необходимо расширить множество потенциальных предикторов.

Применяя ряд предположений о распределении длин временных интервалов между моментом интервью и последним эпизодом поведения, можно строить более точные регрессионные модели [12, 13]. Однако, при анализе той же выборки множество предикторов, включенных в модель, оказывается тем же самым (возраст, копинг-стратегия — положительная переоценка, склонность к риску) и направления связи также не меняются.

К возможным направлениям развития данного исследования можно также отнести анализ данных о других видах рискованного поведения: о потреблении наркотиков, отклонении от приема лекарственных препаратов, рискованном сексуальном поведении. В этом случае представляет интерес тот факт, является ли множество предикторов, включенных в модель, одинаковым для разных видов поведения. Кроме того, взаимосвязи между эпизодами рискованного поведения и психологическими характеристиками можно исследовать и на примерах, не связанных с эпидемиологией, например, используя данные, относящиеся к информационной безопасности [14, 15].

Поддержка исследований. Часть результатов, представленных в настоящей работе, была получена на основе результатов исследований, поддержанных грантом РГНФ № 07-06-00738а «Взаимосвязь адаптивных стилей ВИЧ-инфицированных и степени рискованности их поведения», госконтрактом № 2.442.11.7489, шифр 2006-РИ-19.0/001/209, на НИР «Психологическая защита и копинг-стратегии ВИЧ-инфицированных с точки зрения опасности для общественного здоровья» в рамках ФЦНТП «Исследования и разработки по приоритетным направлениям развития науки и техники на 2002–2006 годы», грантом СПбНЦ РАН № 2-199 на 2007 год «Моделирование и измерение количественных характеристик ВИЧ-рискованного поведения на основе обработки ответов респондентов». Руководитель проектов — Т. В. Тулупьева.

Литература

1. World Health Organization. Russian Federation: summary country profile for HIV/AIDS treatment scale-up, 2005. [Электронный ресурс] <http://www.who.int/hiv/HIVCP_RUS.pdf> (По состоянию на 28.02.2012)
2. Niccolai L. et al. Estimates of HIV incidence among drug users in St. Petersburg, Russia: continued growth of a rapidly expanding epidemic // *European Journal of Public Health*. 2011. Vol. 21. Issue 5. Pp. 613-619.
3. Federal Scientific Methodic Center for HIV Fight and Prevention. The amount of HIV-infected persons in Russia by October 31, 2011. [Электронный ресурс] <<http://www.hivrussia.org/stat/2011.shtml>>. (По состоянию на 28.02.2012)
4. Joint United Nations Programme on HIV/AIDS (UNAIDS). Russian Federation HIV and AIDS Estimates, 2009. [Электронный ресурс] <<http://www.unaids.org/en/regionscountries/countries/russianfederation/>> (По состоянию на 28.02.2012)
5. Тулупьева Т.В., Пащенко А.Е., Тулупьев А.Л., Красносельских Т.В., Казакова О.С. Модели ВИЧ-рискованного поведения в контексте психологической защиты и других адаптивных стилей. СПб.: Наука, 2008. 140 с.
6. Тулупьева Т. В., Тулупьев А. Л., Пащенко А. Е., Сироткин А. В., Столярова Е. В. Социо-психологические механизмы регуляции и адмиссивность поведения личности: Учеб.-метод. пособие. СПб.: Речь, 2007. 32 с.
7. Тулупьева Т.В., Тулупьев А.Л., Пащенко А.Е. Оценка интенсивности поведения респондента в условиях информационного дефицита // *Труды СПИИРАН*. Вып. 7. СПб.: Наука, 2008. С. 239–254.

8. Тулупьева Т.В., Пащенко А.Е., Тулупьев А.Л., Голянич В.М. Модели ВИЧ-рискованного поведения в контексте психологической защиты и адаптации // Вестник СПбГУ. 2010. Серия 12. Вып. 1. С. 95–104.
9. Пащенко А.Е., Тулупьев А.Л., Николенко С.И. Моделирование заражения ВИЧ-инфекцией на основе данных о последних эпизодах рискованного поведения // Изв. Высш. учеб. заведений: Приборостроение, 2006. №11. С. 33–34.
10. Kutner M., Neter J., Nachtsheim C., Li W. Applied Linear Statistical Models. Fifth edition. McGraw-Hill Inc., 2004. 1370 pp.
11. Боровков А.А. Математическая статистика. Новосибирск: Наука; Издательство Института математики, 1997. 772 с.
12. Зельтерман Д., Тулупьев А.Л., Суворова А.В., Пащенко А.Е., Мусина В.Ф., Тулупьева Т.В., Красносельских Т.В., Гро Л., Хаймер Р. Обработка систематической ошибки, связанной с длиной временных интервалов между интервью и последним эпизодом в гамма-пуассоновской модели поведения // Труды СПИИРАН. 2011. Вып. 16. С. 160–185.
13. Зельтерман Д., Суворова А.В., Пащенко А.Е., Мусина В.Ф., Тулупьев А.Л., Тулупьева Т.В., Гро Л.Е., Хаймер Р. Диагностика регрессионных уравнений в анализе интенсивности рискованного поведения по его последним эпизодам // Труды СПИИРАН. 2011. Вып. 17. С. 33–46.
14. Тулупьев А.Л., Азаров А.А., Тулупьева Т.В., Пащенко А.Е., Степанкин М.В. Социально-психологические факторы, влияющие на степень уязвимости пользователей автоматизированных информационных систем с точки зрения социоинженерных атак // Труды СПИИРАН. 2010. Вып. 1 (12). С. 200–214.
15. Тулупьева Т.В., Тулупьев А.Л., Азаров А.А., Пащенко А.Е. Психологическая защита как фактор уязвимости пользователя в контексте социоинженерных атак // Труды СПИИРАН. 2011. Вып. 18. С. 74–92.

Суворова Алена Владимировна — младший научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики СПИИРАН, аспирант математико-механического факультета Санкт-Петербургского государственного университета (СПбГУ). Область научных интересов: математическая статистика, теория вероятности, применение методов математического моделирования в эпидемиологии. Число научных публикаций — 21. SuvorovaAV@iias.spb.su, www.tulupyev.spb.ru; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450. Научный руководитель — А.Л. Тулупьев.

Suvorova Alena Vladimirovna — junior researcher, Laboratory of Theoretical and Interdisciplinary Computer Science, SPIIRAS, PhD student, Faculty of Mathematics and Mechanics of St. Petersburg State University (SPbSU). Research interests: mathematical statistics, probability theory, application of mathematical modeling in epidemiology. The number of publications — 21. SuvorovaAV@iias.spb.su, www.tulupyev.spb.ru; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450. Scientific advisor — A.L. Tulupiev.

Тулупьева Татьяна Валентиновна — канд. психол. наук, доцент; старший научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики СПИИРАН, доцент кафедры информатики математико-механического факультета С.-Петербургского государственного университета (СПбГУ), доцент кафедры психологии

управления и педагогики Северо-Западной академии государственной службы (СЗАГС). Область научных интересов: применение методов математики и информатики в гуманитарных исследованиях, информатизация организации и проведения психологических исследований, применение методов биostatистики в эпидемиологии, психология личности, психология управления. Число научных публикаций — 70. TVT@iias.spb.su, www.tulupyev.spb.ru; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; п.т. +7(812)328-3337, факс +7(812)328-4450.

Tulupyeva Tatiana Valentinovna — PhD in Psychology, associate professor; senior researcher, Laboratory of Theoretical and Interdisciplinary Computer Science, SPIIRAS, associate professor, Computer Science Department, Faculty of Mathematics and Mechanics, St. Petersburg State University (SPbSU), associate professor, Management Psychology and Pedagogic Department, North-West Academy of Public Administration (NWAPA). Research interests: application of mathematics and computer science in humanities, informatization of psychological studies, application of biostatistics in epidemiology, psychology of personality, management psychology. Number of publications — 70. TVT@iias.spb.su, www.tulupyev.spb.ru; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450.

Тулупьев Александр Львович — д-р физ.-мат. наук, доцент; заведующий лабораторией теоретических и междисциплинарных проблем информатики СПИИРАН, профессор кафедры информатики математико-механического факультета С.-Петербургского государственного университета (СПбГУ). Область научных интересов: представление и обработка данных и знаний с неопределенностью, применение методов математики и информатики в социокультурных исследованиях, применение методов биostatистики и математического моделирования в эпидемиологии, технология разработки программных комплексов с СУБД. Число научных публикаций — 210. ALT@iias.spb.su, www.tulupyev.spb.ru; СПИИРАН, 14-я линия В.О., д. 39, Санкт-Петербург, 199178, РФ; п.т. +7(812)328-3337, факс +7(812)328-4450.

Tulupyev Alexander Lvovich — PhD in Appl. Math. and CS, Dr. Sci. in CS, associate professor; head of laboratory, Theoretical and Interdisciplinary Computer Science Laboratory, SPIIRAS, professor, Computer Science Department, Faculty of Mathematics and Mechanics, St. Petersburg State University (SPbSU). Research interests: uncertain knowledge and data representation and processing, application of mathematics and computer science in sociocultural studies, applications of biostatistics and mathematical modeling in modern epidemiology, software technologies and development of information systems with databases. The number of publications — 210. ALT@iias.spb.su, www.tulupyev.spb.ru; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)3284450.

Рекомендовано ТИМПИ СПИИРАН, зав. лаб. д-р физ.-мат. наук, доцент А.Л. Тулупьев. Статья поступила в редакцию 23.03.2012.

РЕФЕРАТ

Суворова А.В., Тулупьева Т.В., Тулупьев А.Л. Обобщенная линейная регрессионная модель для прогноза временного интервала между последним эпизодом рискованного поведения и моментом интервью на основе социально-демографических и психологических особенностей.

В подавляющем большинстве случаев заражение ВИЧ-инфекцией происходит при непосредственном участии ВИЧ-инфицированных людей. Для того чтобы лучше понять, какие подходы по планированию и проведению поведенческих интервенций среди ВИЧ-инфицированных должны формироваться, необходимо заниматься тем психологическим контекстом, в котором осуществляется их рискованное поведение.

Для изучения поведенческих особенностей ВИЧ-инфицированных в СПИ-ИРАН в 2007–2008 гг. было организовано и проведено на базе ГУЗ «Центр по профилактике и борьбе со СПИД и инфекционными заболеваниями» в Санкт-Петербурге полевое исследование, фокусировавшееся на их адаптивном потенциале, с целью формирования моделей рискованного поведения и оценки его параметров в контексте адаптационных стилей и социальных связей личности для последующего обеспечения разработки, рационального планирования и эффективного выполнения программ по поведенческой интервенции среди указанных лиц.

В данной работе исследуются взаимосвязи между последними эпизодами потребления алкоголя как показателя рискованности поведения и шестью социально-демографическими и психологическими характеристиками респондента, такими как пол респондента, его возраст, потребность в поиске новых ощущений, склонность к риску, копинг-стратегия (положительная переоценка), психологическая защита (замещение). Для этого строится регрессионная модель с интервалом между интервью и последним эпизодом поведения в качестве зависимой переменной, а также проводится анализ этой модели.

Согласно исследуемым данным время между интервью и последним эпизодом потребления алкоголя взаимосвязано с возрастом респондента, такой копинг-стратегией как положительная переоценка и склонностью к риску, причем эта связь позитивна для первых двух показателей и негативна для последнего. Построенная модель является лучшей, описывающей взаимосвязи между давностью потребления алкоголя и шестью социально-демографическими и психологическими характеристиками респондента, но неудобна для прогнозирования: данная модель объясняет только небольшую долю поведения зависимой переменной. Включение в модель всех шести социально-демографических и психологических показателей также не улучшает ситуацию. Для построения более подходящих моделей, описывающих зависимости между эпизодами поведения и социально-демографическими и психологическими характеристиками респондентов, необходимо расширить множество потенциальных предикторов и рассмотреть другие виды рискованного поведения.

SUMMARY

Suvorova A.V., Tulupyeva T.V., Tulupyev A.L. **Generalized linear regression model for prediction of interval between the interview moment and the last episode of risky behavior on the base of psychosocial and demographic traits.**

In most cases, people acquire HIV-infection when they have risky contacts with HIV-positive people. As a result, to understand what special prevention approach should be implemented or developed, we have to study risky behavior of HIV-infected people and their psychological traits that lead to such behavior.

In 2007–2008, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS) in collaboration with St. Petersburg Municipal AIDS Center conducted three closely-related pilot studies on the interdependencies among HIV-positive people risky behavior, their adaptive styles, other psychological traits, and socio-demographical characteristics. The major goal of the studies was to identify the most expressed maladaptive styles of HIV-infected people behavior that can lead to higher rates of HIV transmission. In order to reach the goal, HIV-infected people psychological defense mechanisms and coping strategies expression levels were measured and their relationships with risky behaviors manifestations and predictors such as inclination to risk, sensation seeking, drug and alcohol abuse, deviations from HAART regiment were discovered.

This paper studies association between time till last alcohol consumption (that can be used as an indicator of risky behavior rate) and socio-demographical and psychological characteristics of respondents. We construct the appropriate linear regression model with time till last alcohol consumption as response variable. We examine the model using such diagnostics as residuals plots, test for outliers test for normality and others.

According to the data the time till last alcohol consumption is associated with age of the respondent, his Coping Strategy score and Inclination to Risk Behavior measure. Response variable is negatively associated with Inclination to Risk Behavior measure, positively associated with age and Coping Strategy score. This model is the best that *describes* association between time till last alcohol consumption and 6 socio-demographical and psychological characteristics, but it is not very good for making predictions. The variation in response variable is reduced by 6,56 percent. This is quite small value, but using the whole set of predictor variables does not increase this value.

To obtain better models, we should try to find some new predictor variables and get a new dataset with, perhaps, other psychological characteristics.