

(Высшая школа общественного здравоохранения МЗ РК, г. Алматы)

ИЗУЧЕНИЕ ПРЕДИКТОРОВ ИНФИЦИРОВАНИЯ ВИЧ СРЕДИ НАСЕЛЕНИЯ КАЗАХСТАНА С ИСПОЛЬЗОВАНИЕМ МЕТОДИКИ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Аннотация

В работе с использованием метода бинарной логистической регрессии разработана математическая модель для изучения предикторов инфицирования ВИЧ в Казахстане. Установлено, что на регистрацию новых случаев ВИЧ-инфекции оказывают влияние следующие предикторы (переменные): область, место жительства, «объединенный код».

Ключевые слова: математическое моделирование, регрессия, ВИЧ-инфекция, предикторы.

Кілт сөздер: математикалық үлгілеу, регрессия, АИТВ індетін жұқтыру, предикторлар.

Keywords: mathematical modelling, regress, HIV-infection, predictors.

Большинство зарубежных и отечественных исследователей [1-6] рассматривают ВИЧ-инфекцию как сложное социально-биологическое явление. При этом в значительной степени авторы отдают приоритет социальным факторам, в частности, предикторам (эпидемиологическим переменным), связанным с ВИЧ-инфицированием. В доступной литературе мы нашли лишь единичные работы, посвященные в Казахстане исследованиям взаимосвязи социально-экономических, демо-графических, медицинских факторов на распространение ВИЧ-инфекции с использованием методов математики и статистики [7, 8].

В данной работе представлена математическая модель изучения влияния различных предикторов на проводимый в Казахстане скрининг населения на ВИЧ-инфекцию.

Дизайн исследования. Необходимыми материалами явились официальные результаты эпидемиологического скрининга на ВИЧ населения Казахстана по 16 регионам республики за 10-летний период (2001–2010 гг). В качестве предикторов использовались следующие переменные: область, год, пол, место жительства (город/село), код по форме 4, возраст на момент тестирования, результат обследования на ВИЧ. На основе вышеуказанных предикторов была разработана математическая модель с использованием методики бинарной логистической регрессии [9]. Данный метод используется, когда

зависимая переменная является бинарной, т.е. принимает 2 значения. Метод позволяет выделить «статистически значимые» объяснительные переменные (предикторы), влияющие на зависимую переменную. В качестве предикторов могут быть переменные следующих типов: количественные, двоичные, номинативные. В случае использования номинативных переменных создаются несколько «фиктивных» переменных (двоичных), комбинация которых описывает полностью номинативную переменную. В программе SPSS существует опция, которая позволяет этот процесс производить автоматически. Метод бинарной логистической регрессии позволяет учитывать эффект «конфаундер» – когда объяснительная переменная «коррелирует» с зависимой переменной, но сама не является объяснительной переменной. Корреляция данной переменной с результирующей объясняется корреляцией с другими предикторами, т.е. отсутствует причинно-следственная связь между данной переменной и зависимой переменной. Уравнение бинарной логистической регрессии имеет следующий вид:

$$p(A) = \frac{1}{1 + e^{-z}},$$

(1)

где A – событие, которое нужно предсказать; $p(A)$ – вероятность события A ; e – экспонента, равная 2,71...; z – определяется из уравнения:

$$z = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n,$$

(2)

где $b_i, i = 0, 1, \dots, n$ – искомые коэффициенты; $x_i, i = 1, 2, \dots, n$ – значения соответствующих переменных. Статистически значимое отличие коэффициента $b_i, i = 0, \dots, n$ от нуля позволяет говорить, что соответствующая переменная x_i статистически значимо влияет на вероятность события.

Вычисления производились с использованием статистической программы SPSS версии 15.1. В качестве зависимой переменной бралась переменная – результат тестирования на ВИЧ (0 – «отрицательный»; 1 – «положительный»). В качестве возможных предикторов (объяснительных переменных) были взяты следующие переменные: а) область, б) год тестирования, в) пол (1 – мужской; 2 – женский), г) место жительства (1 – город; 2 – село), д) код по форме 4, е) возраст на момент тестирования. В качестве номинативных переменных использовались следующие переменные: область, код по форме.

Для улучшения статистической достоверности модели, код по форме был сгруппирован в следующие группы:

– 101, 106, 107 – лица, бывшие в контакте с ВИЧ, проведение эпид. расследования, обследованные при аварийных ситуациях;

– 101.1 – половой контакт с ВИЧ;

– 101.3, 115, 116, 121, 124 – прочие, без выраженного риска;

– 102, 101.2 – ПИН;

- 103 – МСМ;
- 104 – с симптомами ИППП;
- 105 – РС;
- 108 – доноры;
- 109 – беременные;
- 110 – реципиенты компонентов крови и др. биологических материалов;
- 111 – военнослужащие;
- 112, 118 – СИЗО, ИУ, ИВС, ЦВИАРН, приемники – распределители;
- 113 – по клиническим показаниям;
- 113.5 – по туберкулезу легочному и внелегочному;
- 113.9 – хронические воспалительные заболевания неясной этиологии;
- 114 – обследованные анонимно;
- 117, 122, 123 – дети;
- 120 – медицинские работники;
- 200 – иностранные граждане.

Результаты. Предложенные переменные статистически значимо улучшили модель прогноза регистрации случаев ВИЧ – инфекции ($\chi^2 = 505,437$; $df = 37$; $p < 0,001$). Следовательно, среди рассматриваемых переменных есть объяснительные переменные (предикторы), которые объясняют 23,8% изменений зависимой переменной (R^2 Нагелькерке = 0,238).

Статистически значимыми предикторами являются следующие переменные: область (критерий Вальдовского = 55,464; $df = 15$; $p = 0,000$); место жительства ($b = -0,65 \pm 0,201$; $p = 0,001$); объединенный код по форме 4 (критерий Вальдовского = 228,461; $df = 18$; $p = 0,000$). Из приведенных выше переменных – переменная пол является бинарной. Значение $b = -0,65 \pm 0,201$ говорит о том, что у жителя села вероятность ВИЧ-инфицирования меньше, чем у жителя города.

Не влияющими на результат тестирования на ВИЧ являются следующие переменные: год ($p = 0,167$); пол ($p = 0,467$); возраст на момент тестирования ($p = 0,25$).

Для дальнейшего статистического анализа проводилось сравнение всех значений переменных, статистически значимо влияющих на зависимую переменную в модели бинарной логистической регрессии с эталонным (сравниваемым) значением переменной. Результаты оказались следующие:

1. В переменной «область» эталонным значением был «город Астана». В сравнении с городом Астана – различия с другими областями не обнаружены (возможно, из-за не достаточного объема выборки по городу Астана).

2. В переменной «объединенный код» эталонным значением был код «200 – иностранные граждане». Ниже перечислены объединенные коды, у которых скорректированный натуральный логарифм отношения шансов риска (логит) статистически значимо выше нуля. Следовательно, данные объединенные коды являются предикторами, принадлежность к которым статистически значимо увеличивает вероятность ВИЧ-инфицирования по сравнению с эталоном.

2а) по объединенному коду 101, 106, 107 – «лица, бывшие в контакте с ВИЧ, выявленные при проведении эпид. расследования, обследованные при аварийных ситуациях» статистически значимо выше. Натуральный логарифм отношения шансов риска на статус «ВИЧ+» составляет $b = 2,137 \pm 0,828$; $p = 0,01$.

2б) по коду 101.1 – «половой контакт с ВИЧ» статистически значимо выше натуральный логарифм отношения шансов риска на статус «ВИЧ+» ($b = 2,74 \pm 0,398$; $p = 0,000$).

2в) по объединенному коду 102, 101.2 – «ПИН» статистически значимо выше натуральный логарифм отношения шансов риска на статус «ВИЧ+» ($b = 1,99 \pm 0,355$; $p = 0,000$).

2г) по объединенному коду 112, 118 – «СИЗО, ИУ, ИВС, ЦВИАРН, приемники – распределители» статистически значимо выше натуральный логарифм отношения шансов риска на статус «ВИЧ+» ($b = 0,901 \pm 0,379$; $p = 0,017$).

2д) по коду 113.5 – «по туберкулезу легочному и внелегочному» статистически значимо выше натуральный логарифм отношения шансов риска на статус «ВИЧ+» ($b = 1,069 \pm 0,533$; $p = 0,045$).

Ниже перечислены объединенные коды, у которых скорректированный натуральный логарифм отношения шансов риска (логит) статистически значимо выше нуля. Следовательно, данные объединенные коды являются предикторами, принадлежность к которым статистически значимо уменьшает вероятность ВИЧ-инфицирования по сравнению с эталоном.

2а) по коду 108 – «доноры» статистически значимо ниже натуральный логарифм отношения шансов риска на статус «ВИЧ+» ($b = -1,128 \pm 0,565$; $p = 0,046$).

2б) по коду 109 – «беременные» статистически значимо ниже натуральный логарифм отношения шансов риска на статус «ВИЧ+» ($b = -1,357 \pm 0,518$; $p = 0,009$).

По остальным объединенным кодам статистически значимых различий по сравнению с кодом «200 – иностранные граждане» наше исследование не выявило.

На основании вышеизложенного, можно сделать следующие выводы по бинарной логистической регрессии:

1. В предложенной модели бинарной логистической регрессии предикторами служат следующие переменные: «область», «место жительства», «объединенный код».

2. Так как переменная «год» не является статистически значимой, то временные тенденции по регистрации новых случаев ВИЧ-инфекции отсутствуют.

3. Для людей, проживающих в городе, риск регистрации случаев ВИЧ-инфекции по сравнению с селом выше.

4. По сравнению с кодом «иностранные граждане» достоверно выше риск регистрации случаев ВИЧ-инфекции имеют следующие объединенные коды: «лица, бывшие в контакте с ВИЧ, проведение эпид. расследования, обследованные при аварийных ситуациях», «половой контакт с ВИЧ», «ПИН», «СИЗО, ИУ, ИВС, ЦВИАРН, приемники – распределители», «по туберкулезу легочному и внелегочному».

5. По сравнению с кодом «иностранные граждане» достоверно ниже риск регистрации случаев ВИЧ-инфекции имеют следующие коды: «доноры», «беременные».

Так как результаты бинарной логистической регрессии не позволили выявить статистически значимые региональные особенности, то для дальнейшего анализа использовались статистические данные из формы за 2001–2010 годы.

С помощью метода нормированных остатков (в дальнейшем н.о.) в зависимости от доли регистрации новых случаев ВИЧ-инфекции регионы Казахстана были разделены на категории: регионы с высоким уровнем выявления и регионы с низким уровнем выявления ВИЧ-инфекции за 2001–2010 годы. К регионам с высоким уровнем выявления «ВИЧ+» относятся: Алматинская область (н.о. = 9,6); Восточно-Казахстанская область (н.о. = 7,8); Карагандинская область (н.о. = 21,7); Костанайская область (н.о. = 3,6); Павлодарская область (н.о. = 16,8); город Алматы (н.о. = 35,7). К регионам с низким уровнем выявления «ВИЧ+» относятся: Акмолинская область (н.о. = -18,6); Актюбинская область (н.о. = -13,6); Атырауская область (н.о. = -15,2); Жамбылская область (н.о. = -5,3); Западно-Казахстанская область (н.о. = -12,5); Кызылординская область (н.о. = -21,3); Мангистауская область (н.о. = -15,7); Северо-Казахстанская область (н.о. = -6,5); Южно-Казахстанская область (н.о. = -3,9); город Астана (н.о. = -20,3).

На рисунке 1 показано распределение количества лиц, обследованных на ВИЧ, и доли регионов с высоким уровнем ВИЧ (при скрининге). Как следует из рисунка, происходит увеличение количества лиц, протестированных на ВИЧ (с 1 068 365 людей, протестированных на ВИЧ в 2001 году до 2 085 344 людей, протестированных на ВИЧ в 2010 году). С другой стороны, доля регионов с высоким уровнем выявления новых случаев ВИЧ-инфекции в общем объеме тестирования снижается с 57,1% в 2001 году до 41,96% в 2010 году. Следовательно, происходит увеличение объемов тестирования на ВИЧ главным образом за счет увеличения объемов в регионах с низким уровнем выявляемости «ВИЧ+» при скрининге.

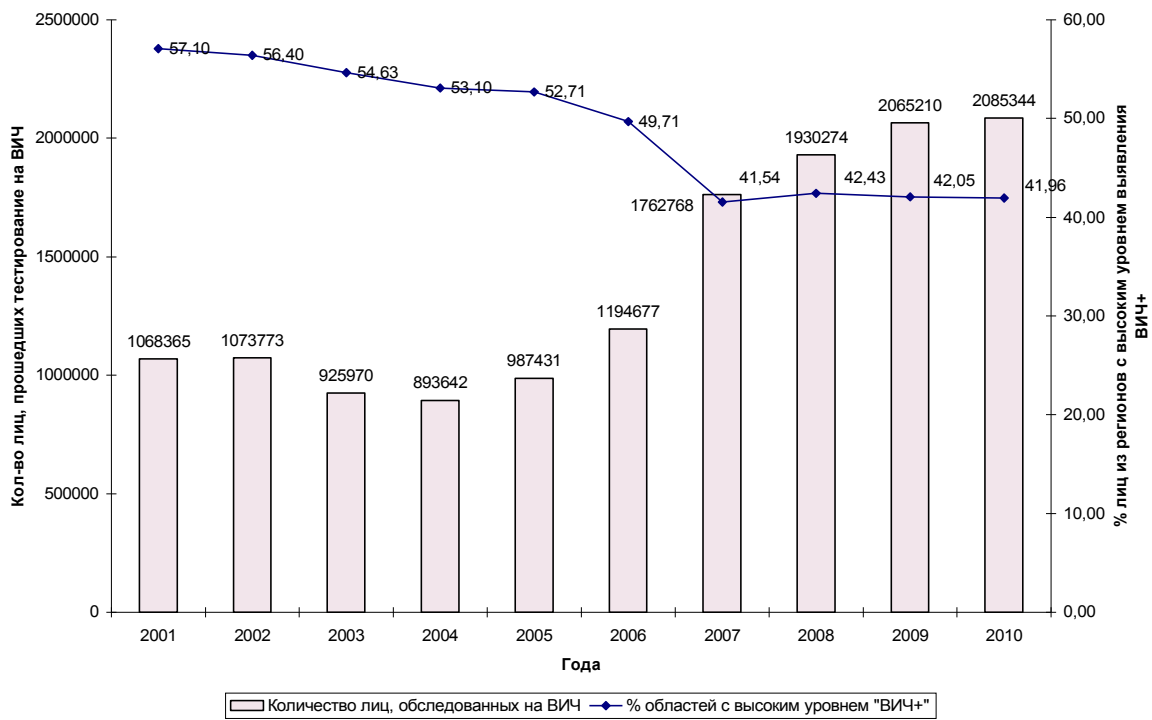


Рисунок 1 – Распределение количества лиц, обследованных на ВИЧ и доли регионов с высоким уровнем ВИЧ (при скрининге)

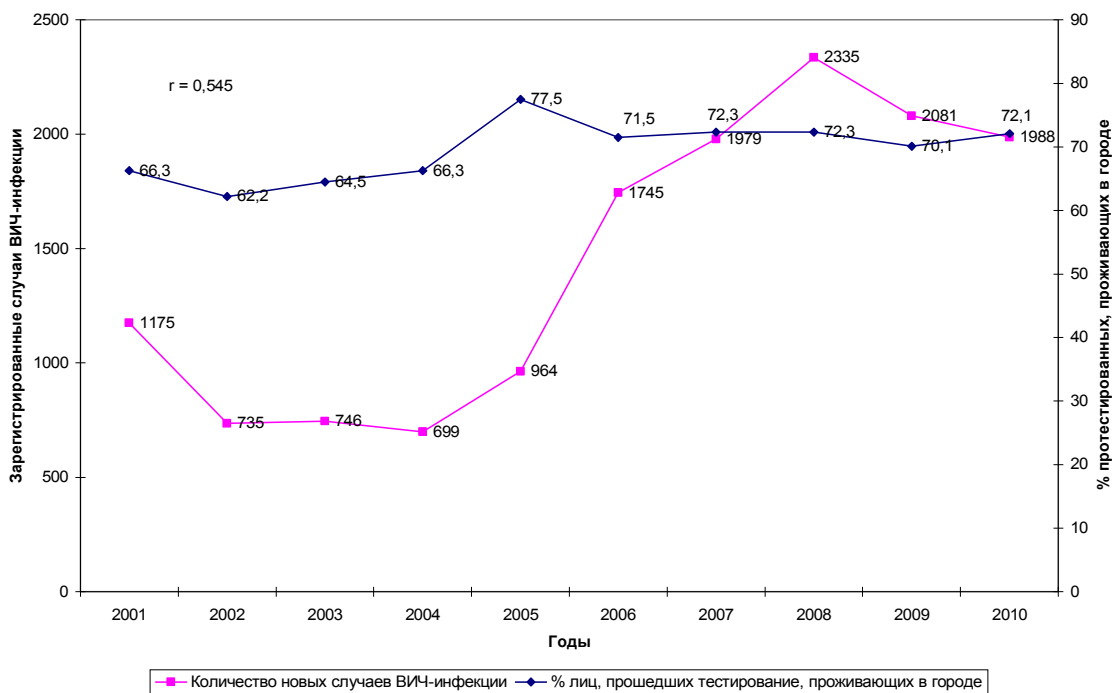


Рисунок 2 – Зависимость между новыми зарегистрированными случаями ВИЧ-инфекции

и % протестированных лиц, проживающих в городе

На рисунке 2 показана зависимость между количеством зарегистрированных случаев ВИЧ-инфекции (согласно отчетным данным формы 4) и процентом протестированных городских жителей по данным данного исследования). Между данными показателями существует прямая тесная корреляционная зависимость ($r=0,545$). В период 2008–2010 годы (когда выборка становится близка к планируемой выборке, т.е. репрезентативной) процент протестированных лиц, проживающих в городе, снижается, что соответствует динамике новых случаев ВИЧ-инфекции.

В процессе исследования были выявлены прямые корреляционные зависимости разной силы между объединенными кодами, являющимися предикторами и регистрацией новых случаев ВИЧ-инфекции.

Таким образом, математическое моделирование с применением методики бинарной логисти-ческой регрессии установило достоверное влияние на регистрацию в Казахстане новых случаев ВИЧ-инфекции следующих предикторов: область, место жительства, «объединенный код». Выявленные количественные связи могут использоваться при эпидемиологическом мониторинге и прогнозировании заболеваемости ВИЧ-инфекции.

ЛИТЕРАТУРА

- 1 Покровский В.В. Эпидемиология и профилактика ВИЧ-инфекции и СПИД. – М., 1996. – 246 с.
- 2 Адамян Р.Х. Анализ ситуации в мире и в Кыргызской Республике // Информационный бюллетень по проблеме СПИД и БППП. – ЮНЭЙДС, 1997. – 24 с.
- 3 Поддержка эпидемиологического надзора за ВИЧ-инфекцией в СНГ // ВОЗ. – 1995. – 32 с.
- 4 Шарман А. Синдром приобретенного иммунодефицита. – США., 2006. – 303 с.
- 5 Крюкова В.А., Ганина Л.Ю., Трумова Ж.З., Сапарбеков М.К. Основы эпидемиологического надзора за ВИЧ-инфекцией / Научный консультант проф. М. К. Сапарбеков. – Алматы, 2010. – 292 с.
- 6 Говард Либман, Харви Дж. Макадон. ВИЧ-инфекция. – М., 2012. – 560 с.
- 7 Есимова Т.М. Значение социальных факторов в развитии эпидпроцесса ВИЧ-инфекции на современном этапе: Автореф. канд. дис. – Алматы, 2007. – 25 с.
- 8 Хасанова М.А., Жусупов Б.С., Сапарбеков М.К., Асембеков Б.С. Изучение влияния социально-демографических и эпидемиологических факторов на распространение ВИЧ-инфекции // В сборнике материалов Центрально-Азиатской научно-практической конференции. – Иссык-Куль, 2009. – С. 58-61.

9 Мавров Г.И., Чинов Г.П. Математическое моделирование и прогноз распространения венерических болезней // Украинский журнал дерматологии, венерологии, косметологии. – 2006. – № 2. – С. 102-105.

REFERENCES

- 1 Pokrovskij V.V. Jepidemiologija i profilaktika VICH-infekcii i SPID. – M., 1996. – 246 s.
- 2 Adamjan R.H. Analiz situacii v mire i v Kyrgyzskoj Respublike // Informacionnyj bjulleten' po probleme SPID i BPPP. – JuNJeJDS, 1997. – 24 s.
- 3 Podderzhka jepidemiologicheskogo nadzora za VICH-infekciej v SNG // VOZ. – 1995. – 32 s.
- 4 Sharman A. Sindrom priobretennogo immunodeficita. – SShA., 2006. – 303 s.
- 5 Krjukova V.A., Ganina L.Ju., Trumova Zh.Z., Saparbekov M.K. Osnovy jepidemiologicheskogo nadzora za VICH-infekciej / Nauchnyj konsul'tant prof. M. K. Saparbekov. – Almaty, 2010. – 292 s.
- 6 Govard Libman, Harvi Dzh. Makadon. VICH-infekcija. – M., 2012. – 560 s.
- 7 Esimova T.M. Znachenie social'nyh faktorov v razvitii jepidprocessa VICH-infekcii na sovremennom jetape: Avtoref. kand. dis. – Almaty, 2007. – 25 s.
- 8 Hasanova M.A., Zhusupov B.S., Saparbekov M.K., Asembekov B.S. Izuchenie vlijanija social'no-demograficheskikh i jepidemiologicheskikh faktorov na rasprostranenie VICH-infekcii // V sbornike materialov Central'no-Aziatskoj nauchno-prakticheskoy konferencii. – Issyk-Kul', 2009. – S. 58-61.
- 9 Mavrov G.I., Chinov G.P. Matematicheskoe modelirovanie i prognoz rasprostraneniija venericheskikh boleznjej // Ukrainskij zhurnal dermatologii, venerologii, kosmetologii. – 2006. – № 2. – S. 102-105.

Резюме

Т. Ф. Балабаев, М. К. Сапарбеков

(ҚР ДМ Қоғамдық денсаулық сақтаудың жоғары мектебі, Алматы қ.)

ҚАЗАҚСТАН ХАЛҚЫНЫҢ АРАСЫНДА АИТВ ІНДЕТІН ЖҰҚТЫРУДЫҢ ПРЕДИКТОРЛАРЫН МАТЕМАТИКАЛЫҚ ҮЛГІЛЕУ ӘДІСІМЕН ЗЕРТТЕУ

Жұмыста бинарлық логистикалық регрессия әдісі қолданып, Қазақстандағы АИТВ індетін жұқтырудың предикторларын зерттеу үшін математикалық үлгісі өңделген. АИТВ-індетінің тіркелуіне келесі предиктор-лар (айнымалылар) әсер ететіні анықталған: облыс, тұрған жері, «біріктірілген код».

Кілт сөздер: математикалық үлгілеу, регрессия, АИТВ індетін жұқтыру, предикторлар.

Summary

T. F. Balabayev, M. K. Saparbekov

(High school of public healthcare of the MH RK, Almaty)

MATHEMATICAL MODELING OF HIV INFECTION PREDICTORS AMONG POPULATION IN KAZAKHSTAN

Main objective of the research is to make mathematical model for defining HIV infecting predictors in Kazakh-stan. There are following predictors that influence to the rate of the new cases of HIV infection: region, residency, and unified code.

Keywords: mathematical modelling, regress, HIV-infection, predictors.

Поступила 27.03.2013г