

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
КАФЕДРА МАТЕМАТИЧЕСКОЙ ТЕОРИИ ИГР И СТАТИСТИЧЕСКИХ  
РЕШЕНИЙ

**Зуенок Кристина Владимировна**

**Выпускная квалификационная работа бакалавра**

**Об исследовании доли ВИЧ-инфицированных среди  
некоторых групп населения**

Направление 02.03.02

Фундаментальная информатика и информационные технологии

Научный руководитель,  
доктор физ.-мат. наук,  
доцент  
Громова Е. В.

Санкт-Петербург  
2018

# Содержание

<b>Введение.</b>	<b>3</b>
<b>Постановка задачи.</b>	<b>5</b>
<b>Обзор литературы</b>	<b>6</b>
<b>Глава 1. Математическая модель</b>	<b>9</b>
1.1 Уравнение демографического баланса	9
1.2 Простейшая модель распространения инфекции	9
1.3 Модель распространения ВИЧ с оценкой поведения	11
<b>Глава 2. Выборка и оценка состояний</b>	<b>15</b>
2.1 Вычисление вероятностей для каждой из групп	15
2.2 Вычисление долей в общей популяции	16
<b>Глава 3. Программная реализация</b>	<b>20</b>
3.1 Выбор коэффициентов модели	20
3.2 Численный эксперимент.	21
3.3 Результаты	23
<b>Выводы</b>	<b>27</b>
<b>Заключение</b>	<b>28</b>
<b>Список литературы</b>	<b>29</b>
<b>Приложение.</b>	<b>32</b>

## Введение

Со времени начала глобальной пандемии ВИЧ/СПИДа в начале 1980-х годов около 70 миллионов человек были инфицированы и 35 миллионов умерли от болезней, связанных со СПИДом [1]. В связи с этим, распространение ВИЧ и СПИД на протяжении многих лет является предметом различных теоретических и медицинских исследований [2].

Актуальными проблемами общественного здравоохранения является сбор информации, ее анализ для оценки текущей ситуации и оптимальное распределение ограниченных ресурсов между различными программами лечения и профилактики. Чем больше информации становится известно о ВИЧ, тем сложнее становятся математические модели, разрабатываемые для контроля распространения заболевания.

Последние достижения в области профилактики ВИЧ демонстрируют новые формы высокоактивной антиретровирусной терапии (ВААРТ), ставшие настолько эффективными, что вирусная нагрузка у инфицированных пациентов становится необнаруживаемой, и пациенты по существу не способны заразить других людей. Альтернативой является использование новых методов предэкспозиционной или доконтактной профилактики ВИЧ, нацеленной на неинфицированных людей, снижающие вероятность заражения вирусом даже при частых контактах.

Однако, даже имея в распоряжении методы лечения, эффективные для блокировки передачи вируса от инфицированных лиц, а также для защиты неинфицированных лиц от заражения, все еще сложно однозначно ответить, как именно данные инновации должны быть применены.

Основным препятствием является то, что, хотя эффективность этих вмешательств на индивидуальном уровне может быть определена с помощью контролируемых испытаний, их эффективность для общественного здравоохранения не может быть установлена однозначно. Это связано с тем, что популяции отличаются друг от друга как динамикой передачи ВИЧ, так и доступностью ресурсов для профилактики.

Вместе с такой постановкой проблемы возникает другая - сложность оценки соотношения людей в популяции, которым необходима та или

иная терапия. Для отражения реальной ситуации, рассматриваемой в исследовании, выборка должна быть репрезентативной, то есть представлять генеральную совокупность - все население - наиболее полно, отражая настоящее соотношение входящих в нее групп. Если выборочная совокупность составлена правильно, то результаты, характеризующие всю генеральную совокупность, считаются надежными, иначе, при недостаточном объеме выборки или существенном смещении, ценность проведенного исследования значительно снижается.

Вне всяких сомнений, решение вышеупомянутых проблем должно основываться на модели основополагающих медицинских, биологических и социальных процессов, отражающих соответствующие особенности населения.

Данная работа рассматривает динамическую эпидемиологическую модель. Предполагается, что целевая популяция состоит из популяции МСМ ограниченной географической единицы (т.е. крупного города, региона или страны), демонстрирующей два разных уровня рискованного поведения. Кроме того, рассматриваются две разные стадии заболевания, острая и хроническая, так как острая стадия связана со значительно более высокой инфекционностью. Данные, доступные для исследования, формируются на основе опросов и тестов, полученных в некоторой высокорисковой среде, и представляют собой нерепрезентативную выборку.

## Постановка задачи

В работе рассматривается эпидемиологическая модель, описывающая население, которое может быть поделено на несколько различных между собой групп с соответствующими состояниями, основываясь на разнохарактерном поведении индивидов и их ВИЧ-статусе. При этом между состояниями осуществляются переходы в связи с инфицированием людей, сменой поведения и иными причинами, рассмотренными далее.

Формальная постановка задачи может быть представлена следующим образом: пусть существует некоторая популяция  $P$ , имеющая размер  $N = |P|$ , из которой производят выборку  $S$  при дополнительных условиях отбора, нарушающих ее репрезентативность, объемом  $N_S = |P|$ ,  $S \subset P$ .

Необходимо разделить рассматриваемое население  $P$  на возможные категории:  $S_{(.)}$ ,  $I_{(.)}$ .

Сформулировать и проанализировать математическую модель, учесть возможное взаимодействие между группами индивидов, развитие, приток новых исследуемых  $\alpha_{(.)}$  и их отток  $\mu$  из системы.

Оценить доли здоровых  $s_{(.)}$  и ВИЧ-инфицированных  $i_{(.)}$  людей из различных категорий во всей популяции  $P$ , на основании данных о заболеваемости  $\hat{s}_{(.)}$  и  $\hat{i}_{(.)}$  в некоторой выборке  $S$  (в частности, данных, полученных при тестировании отдельных индивидов в рамках исследования) в условиях нерепрезентативности выборки.

Оценить количество людей в различных состояниях  $S_{(.)}$ ,  $I_{(.)}$  для всей популяции  $P$ . Осуществить программную реализацию модели и анализ полученных результатов.

## Обзор литературы

Одной из первых работ, посвященных моделированию процессов, связанных с ВИЧ, является статья [3], представляющая численное моделирование нескольких процессов, которые учитывают такие факторы, как переменная инфекционность индивидов и переменные инкубационные периоды в гомосексуальной среде. Оценка поведения индивидов была учтена в работе [4], где Андерсон представил модель распространения заболевания среди гомосексуального населения Великобритании, используя реальные данные, и рассмотрел последствия изменений в сексуальном поведении. В последующих работах [5, 6] Андерсон и его коллеги в продолжение исследований проанализировали модель распространения ВИЧ в гетеросексуальных сообществах в развивающихся странах и изучили влияние стратегий контроля.

Хотя большинство случаев ВИЧ-заболеваний в развитых странах происходят в гомосексуальном обществе, болезнь распространяется на гетеросексуальное общество. Хеткотт [7] представил общую модель, которая объясняет взаимодействие между индивидами с различной сексуальной ориентацией и потребителями инъекционных наркотиков.

Несмотря на снижение во всем мире ежегодного числа новых случаев заболеваемости ВИЧ, среди мужчин, имеющих половые контакты с мужчинами (МСМ) в развитых странах заболеваемость увеличивается [8, 9]. Термин МСМ был впервые употреблен в 1990-х годах и используется в эпидемиологии и социальных исследованиях для обозначения мужчин, постоянно или периодически практикующих сексуальные контакты с другими мужчинами. При этом сексуальная ориентация и сексуальная идентичность не принимаются во внимание [10]. Значительную часть МСМ составляют бисексуальные мужчины и даже состоящие в гетеросексуальном браке, также категория включает в себя лиц мужского пола, идентифицирующих себя как женщин [11]. По данным UNAIDS [12], МСМ находятся на пятом месте по распространенности ВИЧ (после потребителей инъекционных наркотиков, их партнеров, работников сферы секс-услуг и их партнеров) в мире среди людей с рискованным поведением. Однако в конкретных странах ситуация

совершенно иная. Например, в Северной Америке, Австралии, Новой Зеландии и большинстве стран Западной Европы почти 70% всех новых случаев заражения ВИЧ приходится на МСМ [13]. В Канаде - 40%. В США в 2013 году 65% новых случаев заражения ВИЧ также связано с МСМ [14].

Согласно исследованиям, менее 5% всех мужчин, относящихся к категории МСМ, имеют доступ к услугам по профилактике и уходу в связи с ВИЧ, в которых они нуждаются. Этому способствуют различные факторы, в частности, низкий уровень информированности о ВИЧ и о рискованном сексуальном поведении, гомофобия, социальная стигматизация и дискриминация МСМ, в том числе и со стороны медицинских работников [15].

В статье [16] приводится анализ роста ВИЧ в упрощенной версии модели, представленной Хайманом и Стэнли [3] для МСМ-сообщества, в которой учитывается непропорциональное смешивание различных групп риска. Из-за упрощающих допущений, которые используют авторы, модель не может дать точные прогнозы распространения ВИЧ-инфекции. Однако простая форма модели отражает большую часть качественного поведения более сложных моделей и позволяет количественно анализировать распространение инфекции через различные группы риска и исследовать динамику распространения из одной группы риска в другую.

Одной из важнейших работ, рассматривающих моделирование ВИЧ-инфекции является статья [17], опубликованная в 1989 году. Особенностью работы стало использование байесовского подхода для включения в модели неопределенности. Автором были рассмотрены статистические модели, описывающие эпидемии ВИЧ-инфекции в США, которые объясняют наблюдаемую частоту случаев СПИДа в 1978-1987 годах. По оценкам лучшей модели, в апреле 1987 года в США было инфицировано приблизительно от 563 000 до 1 110 000 человек. Однако, данные оценки не учитывали незарегистрированные случаи СПИДа, также слишком большая неопределенность в данных и моделях, исключала возможность высокоточных прогнозов количества людей, которые в то время были инфицированы в США. Стоит отметить, что результаты лучших моделей свидетельствуют о том, что в 1990 году существовало по меньшей мере от 1 до 1.5 миллионов инфицированных людей, по оценкам Центров по

контролю за заболеваниями.

В статье [18] группа ученых обобщила используемые методологии, разработки ключевых моделей и рекомендованные стратегии борьбы с ВИЧ среди MSM. Была произведена оценка ряда исследований, вследствие которых появился ряд общих выводов и рекомендаций. Например, о том, что для эффективного контроля над ВИЧ среди MSM потребуются совмещений нескольких стратегий.

В качестве продолжения предыдущих исследований в работе [19] приводится анализ влияния различных вмешательств на MSM-население, осуществляемый индивидуально и в сочетании с профилактикой ВИЧ. Наиболее эффективным оказывается комбинация тестирования на ВИЧ два раза в год и доконтактной профилактики.

Следующей значимой работой в области исследования влияния поведения индивидов на распространение ВИЧ стала статья [20], принимающая во внимание и корректирующая нереалистичное предположение, присущее большинству моделей передачи ВИЧ, что поведение индивидов постоянно со временем. Для лучшего объяснения наблюдаемых эффектов, была сформулирована простая модель, в которой у людей предполагалось наличие эпизодического высокорискового и низкорискового поведения, связанного с периодом высокой трансмиссивности во время острой ВИЧ-инфекции, что приводило к резкому увеличению распространенности инфекции.

Отдельно стоит отметить работы, развивающие подход предыдущих исследований: [21] и [22], которые посвящены количественному контролю распространения ВИЧ. Авторами предлагается модель распространения ВИЧ, описывается численная схема для вышеупомянутой модели, решается проблема оптимального управления распределением ресурсов для лечения и профилактики ВИЧ.



# Глава 1

## Математическая модель

### 1.1 Уравнение демографического баланса

Общая структура моделирования для описания распространения инфекционных заболеваний - это баланс населения. Это означает, что все население делится на несколько однородных групп (ячеек). Иными словами, все индивиды в группе считаются одинаковыми в своей эволюции. Переменные состояния модели соответствуют количеству людей в каждой группе, а динамика  $i$ -го состояния описывается следующим дифференциальным уравнением:

$$\dot{X}_i = \sum_{i \neq j} (\alpha_{ij}(X) - \alpha_{ji}(X)) - \alpha_{ii}(X) + \omega_i, \quad (1)$$

где

$X_i$  - количество человек в  $i$ -ой группе,

$\alpha_{ij}(X)$  - скорость перехода из группы  $i$  в группу  $j$ ,

$\alpha_{ii}(X)$  - отток из  $i$ -й группы,  $\omega_i$  - приток в  $i$ -ю группу, который обычно считается постоянным.

Предполагается, что выполнены следующие условия регулярности:

A1.  $\alpha_{ij}(X) \geq 0$  и  $\omega_i(X) \geq 0$  для всех  $i, j$ , и  $X \in \mathbb{R}_{\geq 0}^n$ .

A2. Из  $X_i = 0$  следует  $\alpha_{ij}(X) = 0$  и  $\alpha_{ii}(X) = 0$ , то есть нет выходного потока из пустой группы.

Известно, что система (1), удовлетворяющая вышеупомянутым условиям, неотрицательна. Это означает, что для любого начального условия  $x_0 \in \mathbb{R}_{\geq 0}^n$ , решение  $x(t)$  системы (1) принадлежит  $\mathbb{R}_{\geq 0}^n$  для всех  $t \in [0, \infty)$ .

### 1.2 Простейшая модель распространения инфекции

Вся популяция может быть разделена на три категории:

$S$  - здоровые индивиды,

$I_A$  - инфицированные в острой стадии,

$I_C$  - инфицированные в хронической стадии.

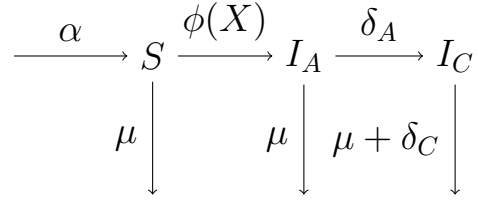


Рис. 1. Простейшая модель распространения инфекции

где

$\alpha$  - входной поток (рождаемость),

$\mu$  - выходной поток (смертность),

$\delta_A$  - продолжительность острой фазы,

$\phi(X)$  - уровень перехода от острой фазы к хронической (коэффициент заболеваемости).

Динамика, представленная на рисунке 1, может быть описана следующей системой ОДУ:

$$\begin{aligned}
 \dot{S} &= \alpha - (\phi(X) + \mu)S, \\
 \dot{I}_A &= \phi(X)S - (\mu + \delta_A)I_A, \\
 \dot{I}_C &= \delta_A I_A - (\mu + \delta_C)I_C,
 \end{aligned} \tag{2}$$

где

$$\phi(X) = \frac{\beta_A I_A + \beta_C I_C}{N},$$

$$X = [S \ I_A \ I_C],$$

$$N = S + I_A + I_C,$$

$\beta_A, \beta_C$  - вероятность заражения за одно взаимодействие для остро и хронически инфицированного соответственно.

### 1.3 Модель распространения ВИЧ с оценкой поведения

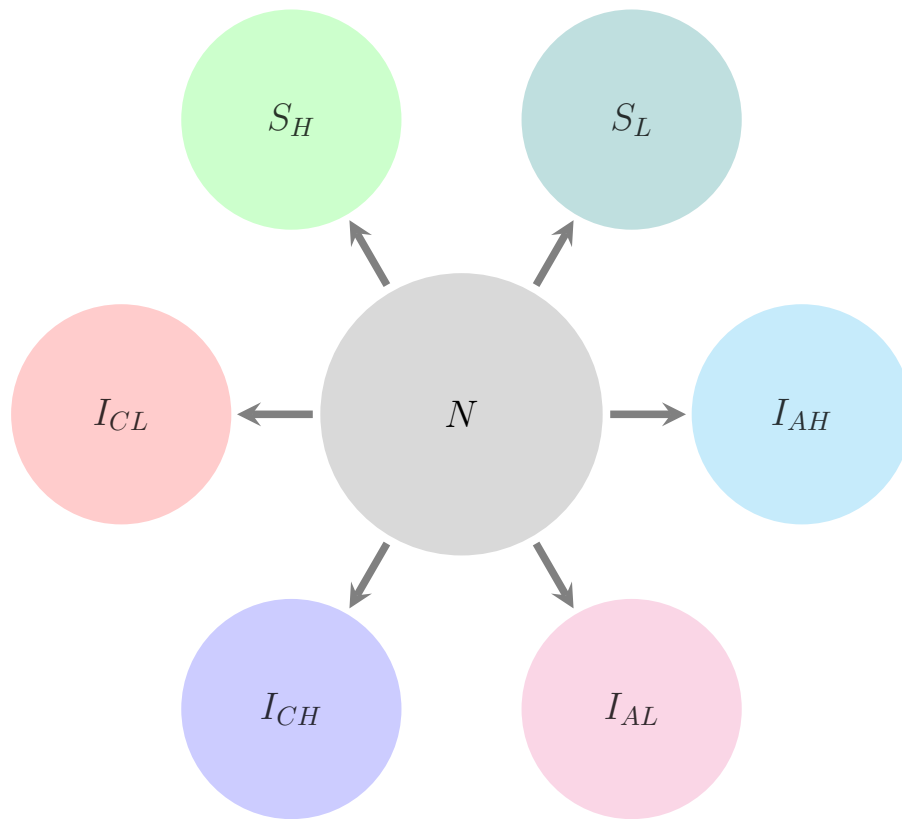


Рис. 2. Деление всего населения на группы относительно заболеваемости и уровня рискованности поведения

Рассмотрим модель распространения ВИЧ среди общего населения MSM в некотором крупном городе США. Подробный вывод модели представлен в [20].

Предположим, что существует некоторая среда высокого риска (HRE - high risk environment), где преобладают люди с высокорисковым поведением, например, бары или секс-клубы. Относительно такой среды можно оценить поведение каждого индивида. Тогда все население может быть поделено на 6 категорий:

$S_H$  - здоровые индивиды с высокорисковым поведением,

$S_L$  - здоровые индивиды с низкорисковым поведением,

$I_{AH}$  - инфицированные в острой стадии с высокорисковым поведением,

$I_{AL}$  - инфицированные в острой стадии с низкорисковым поведением,

$I_{CH}$  - инфицированные в хронической стадии с высокорисковым поведением,

$I_{CL}$  - инфицированные в хронической стадии с низкорисковым поведением.

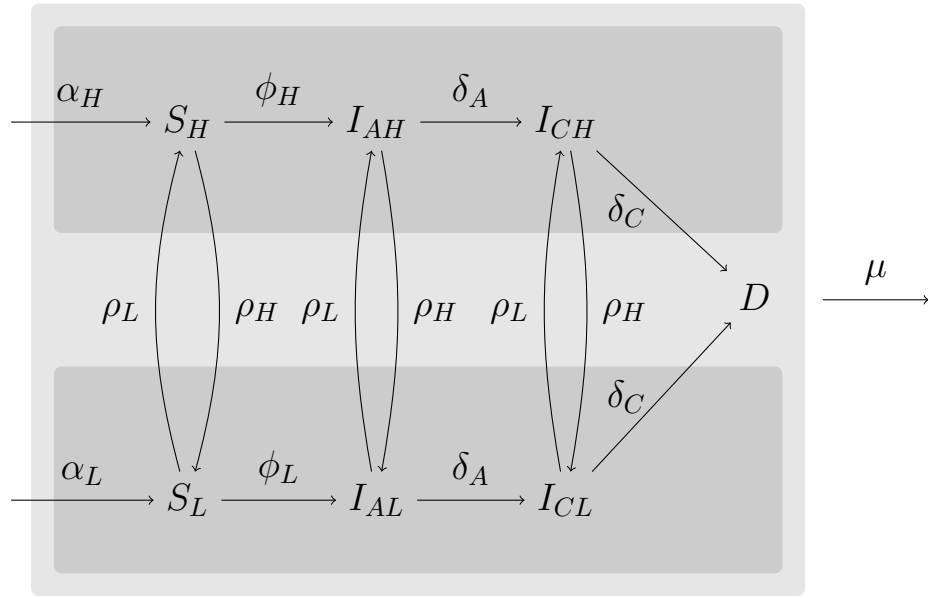


Рис. 3. Динамика системы. В верхнем темно-сером блоке показаны все высокорисковые состояния, в нижнем - все состояния с низким уровнем риска. Переменные на стрелках являются относительными скоростями перехода между ними. Существуют переходы между высоким и низким рисками в обоих направлениях для здоровых и инфицированных людей на любой стадии болезни. Отток  $\mu$  одинаково относится ко всем группам.

Переходы между состояниями происходят из-за того, что индивиды становятся инфицированными, прогрессируют от острой к хронической стадии и умирают от СПИДа ( $S \rightarrow I_A \rightarrow I_C \rightarrow D$ ). Более того, люди меняют свое поведение ( $(.)_L \leftrightarrow (.)_H$ ), переходя из категории высокого риска в категорию низкого риска и обратно. Стоит обратить внимание, что существует поток поступающий в систему в связи с достижением людьми возраста сексуальной активности и отток из системы из-за смерти, не связанной с ВИЧ, или из-за лиц, становящихся сексуально неактивными или оседающих в длительных моногамных отношениях.

Изучаемая динамика описывается следующей системой ОДУ:

$$\begin{aligned}
\dot{S}_H &= \alpha_H - (\phi_H(X) + \rho_H + \mu)S_H + \rho_L S_L, \\
\dot{S}_L &= \alpha_L - (\phi_L(X) + \rho_L + \mu)S_L + \rho_H S_H, \\
\dot{I}_{CH} &= \delta_A I_{AH} - (\rho_H + \mu + \delta_C)I_{CH} + \rho_L I_{CL}, \\
\dot{I}_{CL} &= \delta_A I_{AL} - (\rho_L + \mu + \delta_C)I_{CL} + \rho_H I_{CH}, \\
\dot{I}_{AH} &= \phi_H(X)S_H - (\rho_H + \mu + \delta_A)I_{AH} + \rho_L I_{AL}, \\
\dot{I}_{AL} &= \phi_L(X)S_L - (\rho_L + \mu + \delta_A)I_{AL} + \rho_H I_{AH},
\end{aligned} \tag{3}$$

где  $X = [I_{CH} \ I_{CL} \ I_{AH} \ I_{AL} \ S_H \ S_L]'$  - вектор переменных состояния;  
 $N$  - сумма всех состояний,  $N = \langle \mathbf{1}, X \rangle$ ,  $\mathbf{1}$  - столбец единиц;  $\phi_H(X)$   
и  $\phi_L(X)$  - нелинейные (рациональные) функции от  $X$ , принимающие  
неотрицательные значения для любого  $X \in \mathbb{R}_{\geq 0}^n$ . Все остальные термы  
являются неотрицательными константами (см. таблицу 1).

Переменные состояний	
$S(t)$	неинфицированные индивиды
$I..(t)$	инфицированные индивиды
$D(t)$	умершие из-за СПИДа
$N(t)$	все индивиды
$N..(t)$	все индивиды с данным рисковым поведением
Индексы	
$A/C$	стадия заболевания (острая/хроническая)
$H/L$	рисковый статус (высокий/низкий)
Параметры инфицирования	
$\phi_H, \phi_L$	коэффициент заболеваемости для высокого и низкого риска соответственно
$\delta_A$	коэффициент перехода заболевания из острой формы в хроническую
$\delta_C$	коэффициент смертности хронически инфицированных людей из-за СПИДа
Остальные параметры	
$\alpha_H, \alpha_L$	коэффициент набора новых высокорисковых и низкорисковых индивидов
$\beta_A, \beta_C$	вероятность заражения за одно взаимодействие для остро и хронически инфицированного соответственно
$\mu$	коэффициент взрослых смертей по причинам, не связанным с ВИЧ, людей, достигающих возраста сексуальной пассивности или оседающих в пожизненных моногамных отношениях
$\rho_H$	коэффициент, с которым высокорисковые индивиды становятся низкорисковыми
$\rho_L$	коэффициент, с которым низкорисковые индивиды становятся высокорисковыми

Таблица 1. Параметры модели.

## Глава 2

### Выборка и оценка состояний

Для оценки соотношения людей, принадлежащих определенным группам, индивиды выбираются случайным образом в местах, где преобладают люди с высокорисковым поведением. Обозначим вероятность того, что случайный высокорисковый (низкорисковый соответственно) человек находится в HRE в случайный момент времени в течение периода отбора:

$$\begin{aligned} p_H &= P(HRE|R = H), \\ p_L &= P(HRE|R = L). \end{aligned}$$

#### 2.1 Вычисление вероятностей для каждой из групп

Вероятность случайного индивида  $R$ , находящегося в HRE в случайный момент времени, относиться к одному из состояний, соответствующих категории высокого риска, т.е. находиться в  $X \in Z_H = \{S_H, I_{AH}, I_{CH}\}$ , вычисляется с помощью формулы Байеса:

$$\begin{aligned} P(X|HRE) &= \frac{P(HRE|R = H)P(X)}{P(HRE|R = H)P(R = H) + P(HRE|R = L)P(R = L)} \\ &= \frac{p_H \frac{X}{N}}{p_H \frac{N_H}{N} + p_L \frac{N_L}{N}} = \frac{r_b X}{r_b N_H + N_L}, \end{aligned} \quad (4)$$

где  $N_H = S_H + I_{AH} + I_{CH}$ ,  $N_L = S_L + I_{AL} + I_{CL}$ ,  $N = N_H + N_L$  и  $r_b = p_H/p_L$ . То есть,  $r_b$  - это преобладающий шанс человека с высоким риском попасть в HRE по сравнению с человеком с низким уровнем риска, в данной модели предполагаем  $r_b = 0.8/0.2$ . Вероятность случайного индивида, встречающегося в HRE, относиться к одному из состояний с низким уровнем риска,  $X \in Z_L = \{S_L, I_{AL}, I_{CL}\}$  вычисляется аналогично, получим:

$$\begin{aligned}
P(X|HRE) &= \frac{P(HRE|R=L)P(X)}{P(HRE|R=H)P(R=H) + P(HRE|R=L)P(R=L)} \\
&= \frac{p_L \frac{X}{N}}{p_H \frac{N_H}{N} + p_L \frac{N_L}{N}} = \frac{X}{r_b N_H + N_L}.
\end{aligned} \tag{5}$$

При этом выполняется условие:

$$\sum_{X \in Z_H} P(X|HRE) + \sum_{X \in Z_L} P(X|HRE) = 1.$$

Стоит отметить, что общее число лиц, посещающих HRE в единичный интервал времени, равно  $N$  или по другому, ожидаемое количество посещений HRE для индивида равно 1.

## 2.2 Вычисление долей в общей популяции

Важным моментом является то, что процедуру отбора можно использовать для оценки количества индивидов в различных состояниях. Пусть  $N_S$  - количество индивидов, которые были отобраны во время посещения HRE в единицу времени:

$$N_S = \hat{S}_H + \hat{S}_L + \hat{I}_{AH} + \hat{I}_{AL} + \hat{I}_{CH} + \hat{I}_{CL}.$$

Эти люди опрошены, протестированы (в случае необходимости) и поделены на соответствующие категории согласно результатам. Обозначим число лиц в соответствующих категориях через  $\hat{S}_{(\cdot)}$  и  $\hat{I}_{(\cdot)}$ ; кроме того, обозначим известные доли соответствующих групп внутри выборки  $\hat{s}_{(\cdot)}$  и  $\hat{i}_{(\cdot)}$ , где

$$\hat{s}_H = \frac{\hat{S}_H}{N_S}, \hat{s}_L = \frac{\hat{S}_L}{N_S}, \hat{i}_{AH} = \frac{\hat{I}_{AH}}{N_S}, \hat{i}_{AL} = \frac{\hat{I}_{AL}}{N_S}, \hat{i}_{CH} = \frac{\hat{I}_{CH}}{N_S}, \hat{i}_{CL} = \frac{\hat{I}_{CL}}{N_S}.$$

Таким образом доли соответствующих групп в общей популяции, обозначенные  $s_{(\cdot)}$  и  $i_{(\cdot)}$  могут быть вычислены из соответствующих формул Байеса (4), (5). При этом, чтобы система была хорошо обусловлена, накладывается дополнительное условие нормализации, в котором указано, что сумма всех дробей равна 1.

$$s_H + s_L + i_{AH} + i_{AL} + i_{CH} + i_{CL} = 1. \tag{6}$$



Принимая во внимание все вышеупомянутое, доли могут быть найдены как решение следующего набора линейных уравнений:

$$\left\{ \begin{array}{l} \hat{s}_H = \frac{S_H r_b}{r_b N_H + N_L}, \\ \hat{s}_L = \frac{S_L}{r_b N_H + N_L}, \\ \hat{i}_{AH} = \frac{I_{AH} r_b}{r_b N_H + N_L}, \\ \hat{i}_{AL} = \frac{I_{AL}}{r_b N_H + N_L}, \\ \hat{i}_{CH} = \frac{I_{CH} r_b}{r_b N_H + N_L}, \\ \hat{i}_{CL} = \frac{I_{CL}}{r_b N_H + N_L}, \\ s_H + s_L + i_{AH} + i_{AL} + i_{CH} + i_{CL} = 1. \end{array} \right.$$

Или эквивалентно:

$$\left\{ \begin{array}{l} s_H r_b = \hat{s}_H r_b (s_H + i_{AH} + i_{CH}) + \hat{s}_H (s_L + i_{AL} + i_{CL}), \\ s_L = \hat{s}_L r_b (s_H + i_{AH} + i_{CH}) + \hat{s}_L (s_L + i_{AL} + i_{CL}), \\ i_{AH} r_b = \hat{i}_{AH} r_b (s_H + i_{AH} + i_{CH}) + \hat{i}_{AH} (s_L + i_{AL} + i_{CL}), \\ i_{AL} = \hat{i}_{AL} r_b (s_H + i_{AH} + i_{CH}) + \hat{i}_{AL} (s_L + i_{AL} + i_{CL}), \\ i_{CH} r_b = \hat{i}_{CH} r_b (s_H + i_{AH} + i_{CH}) + \hat{i}_{CH} (s_L + i_{AL} + i_{CL}), \\ i_{CL} = \hat{i}_{CL} r_b (s_H + i_{AH} + i_{CH}) + \hat{i}_{CL} (s_L + i_{AL} + i_{CL}), \\ s_H + s_L + i_{AH} + i_{AL} + i_{CH} + i_{CL} = 1. \end{array} \right.$$

На практике члены некоторых групп, не могут быть распознаны во время обследования из-за инкубационного периода болезни. Такой промежуток времени, когда вирус ВИЧ уже присутствует в крови, а антитела еще не появились, называется "серологическим окном". В период "окна", несмотря на высокое содержание ВИЧ в крови, тесты показывают ложноотрицательный результат. При этом инфицированный человек уже может передавать вирус другим людям.

Обычно заметное количество антител к ВИЧ появляется в крови через 2-10 недель после заражения, однако разброс во времени может быть весьма велик. Так, у 90-95 % зараженных они обнаруживаются в течение 3-х месяцев после заражения. Наиболее ранний срок обнаружения антител - 2 недели от момента заражения. Поэтому при наличии данных,

свидетельствующих о контакте пациента с ВИЧ-инфицированными, обычно проводят повторные исследования через 2 - 3 месяца.

В данной модели выборка индивидов происходит в короткий промежуток времени, не позволяющий сделать повторный анализ и удостовериться в результате. Поэтому в рассматриваемом случае группы  $I_{AH}$  и  $I_{AL}$  объединяются с группами  $S_H$  и  $S_L$  соответственно. Это приводит к определенному смещению в оценках, которые могут быть исправлены с помощью дополнительного анализа.

Тогда система имеет вид:

$$\begin{cases} s_H r_b = \hat{s}_H r_b (s_H + i_{CH}) + \hat{s}_H (s_L + i_{CL}), \\ s_L = \hat{s}_L r_b (s_H + i_{CH}) + \hat{s}_L (s_L + i_{CL}), \\ i_{CH} r_b = \hat{i}_{CH} r_b (s_H + i_{CH}) + \hat{i}_{CH} (s_L + i_{CL}), \\ i_{CL} = \hat{i}_{CL} r_b (s_H + i_{CH}) + \hat{i}_{CL} (s_L + i_{CL}), \\ s_H + s_L + i_{CH} + i_{CL} = 1. \end{cases}$$

Выразим, например неизвестную  $s_L$  и подставим в остальные равенства:

$$\begin{cases} r_b s_H = \hat{s}_H r_b (s_H + i_{CH}) + \hat{s}_H (1 - s_H - i_{CH}), \\ r_b i_{CH} = \hat{i}_{CH} r_b (s_H + i_{CH}) + \hat{i}_{CH} (1 - s_H - i_{CH}), \\ i_{CL} = \hat{i}_{CL} r_b (s_H + i_{CH}) + \hat{i}_{CL} (1 - s_H - i_{CH}), \\ 1 - s_H - i_{CH} - i_{CL} = \hat{s}_L r_b (s_H + i_{CH}) + \hat{s}_L (1 - s_H - i_{CH}), \\ s_H + s_L + i_{CH} + i_{CL} = 1. \end{cases}$$

Сгруппируем:

$$\begin{cases} r_b s_H + \hat{s}_H (1 - r_b) (s_H + i_{CH}) = \hat{s}_H, \\ r_b i_{CH} + \hat{i}_{CH} (1 - r_b) (s_H + i_{CH}) = \hat{i}_{CH}, \\ i_{CL} + \hat{i}_{CL} (1 - r_b) (s_H + i_{CH}) = \hat{i}_{CL}, \\ 1 - s_H - i_{CH} - i_{CL} + \hat{s}_L (1 - r_b) (s_H + i_{CH}) = \hat{s}_L, \\ s_H + s_L + i_{CH} + i_{CL} = 1. \end{cases}$$

Перейдем к матрицам:

$$\left[ \begin{pmatrix} r_b s_H & 0 & 0 \\ 0 & r_b i_{CH} & 0 \\ 0 & 0 & i_{CL} \end{pmatrix} + (1 - r_b) \begin{pmatrix} \hat{s}_H & 0 & 0 \\ 0 & \hat{i}_{CH} & 0 \\ 0 & 0 & \hat{i}_{CL} \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} s_H \\ i_{CH} \\ i_{CL} \end{pmatrix} \right] = \begin{pmatrix} \hat{s}_H \\ \hat{i}_{CH} \\ \hat{i}_{CL} \end{pmatrix},$$

В общем виде получим:

$$\left[ \begin{pmatrix} r_b I_{[H]} & 0 \\ 0 & I_{[L]} \end{pmatrix} + (1 - r_b) \begin{pmatrix} \text{diag}(\hat{x}_H) & 0 \\ 0 & \text{diag}(\hat{x}_L) \end{pmatrix} \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} \end{pmatrix} \right] \begin{pmatrix} x_H \\ x_L \end{pmatrix} = \begin{pmatrix} \hat{x}_H \\ \hat{x}_L \end{pmatrix}, \quad (7)$$

где элементами векторов  $x_H = (s_H \ i_{CH})^T$ ,  $x_L = (i_{CL})^T$  являются доли во всей популяции высокорисковых и низкорисковых групп соответственно; элементы векторов  $\hat{x}_H = (\hat{s}_H \ \hat{i}_{CH})^T$  и  $\hat{x}_L = (\hat{i}_{CL})^T$  - известные доли соответствующих групп внутри выборки;  $I_{[H]}$  и  $I_{[L]}$  - единичные матрицы, размерности которых равны длинам соответствующих векторов  $x_H$  и  $x_L$ ;  $\mathbf{1}$  и  $\mathbf{0}$  - матрицы, состоящие из единиц и нулей, где размерность верхней матрицы из единиц соответствует размерности матрицы  $I_{[H]}$ , а нижней матрицы из нулей - размерности  $I_{[L]}$ , причем верхняя матрицы из нулей и нижняя из единиц - все соответствующих размерностей - дополняют исходную матрицу до квадратной;  $\text{diag}(\hat{x}_H)$ , и  $\text{diag}(\hat{x}_L)$  - диагональные матрицы, элементы которых соответствуют элементам соответствующих векторов.

Заметим, что вектор  $x_L$  не содержит компоненту  $s_L$ , которая вычисляется из ограничения нормировки, т.е.  $s_L = 1 - s_H - i_{CH} - i_{CL}$ . Очевидно, что систему (7) можно переформулировать, чтобы исключить любую другую компоненту вместо  $s_L$ .

Доказательство обратимости матрицы

$$M = \left[ \begin{pmatrix} r_b I_{[H]} & 0 \\ 0 & I_{[L]} \end{pmatrix} + (1 - r_b) \begin{pmatrix} \text{diag}(\hat{x}_H) & 0 \\ 0 & \text{diag}(\hat{x}_L) \end{pmatrix} \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} \end{pmatrix} \right] \quad (8)$$

приведено в [22].

Из этого следует, что система (7) хорошо определена и в качестве решения можно получить вектора  $x_H$  и  $x_L$ . Если известна оценка размера всей популяции  $\hat{N}$ , можно получить оценки переменных состояния, например  $\hat{S}_H = s_H \cdot \hat{N}$  для числа здоровых высокорисковых индивидов и аналогично для других состояний.

## Глава 3

### Программная реализация

#### 3.1 Выбор коэффициентов модели

Коэффициенты, описывающие поведение индивидов, включая отношение их уровней активности, вероятность передачи инфекции при взаимодействии, вероятность смены поведения в ту или иную сторону, получены из анализа данных о длительных половых контактах МСМ-населения в трех крупных городах США [22, 23].

Итак, будем рассматривать сценарий вспышки в крупном городе США, размер популяции  $P$  которого составляет 100 000 МСМ-человек. Предполагается, что инфицированные индивиды в острой фазе в среднем в 15 раз более заразны, чем хронически инфицированные (согласно результатам исследования [24]). Кроме того, отметим, что 90% населения имеют поведение с низким риском, остальные 10% имеют поведение с высоким риском, то есть соотношение МСМ-населения, подверженного низкому риску, к МСМ-населению с высоким риском составляет 9:1. Следовательно, установим коэффициенты вхождения новых индивидов  $\alpha_L = 250$  и  $\alpha_H = 28$  для низкого и высокого риска соответственно, что гарантирует, что вновь входящий человек имеет высокорисковое поведение с вероятностью 0.1.

Коэффициент взрослой смертности полагаем равным  $\mu = \frac{1}{12 \cdot 30}$ , основываясь на предположении, что индивиды, чей отток из системы не связан с ВИЧ-инфекцией, сексуально активны, то есть находятся в системе, в среднем 30 лет. Предполагается, что здоровые восприимчивые люди заражаются на некоторой общей территории и что у индивидов из группы высокого риска в десять раз больше сексуальных контактов, чем у лиц из группы низкого риска, основываясь на анализе продольных половых поведенческих данных в статье [23].

На основании исследований [20] коэффициент перехода заболевания из острой формы в хроническую  $\delta_A$  полагаем равным  $\frac{1}{4}$ , коэффициент смертности хронически инфицированных людей из-за СПИДа  $\delta_C$  - равным

Параметры инфицирования	
$\delta_A$	$\frac{1}{4}$
$\delta_C$	$\frac{1}{116}$
$\beta_A$	0.015
$\beta_C$	0.001
Остальные параметры	
$\rho_H$	0.8
$\rho_L$	0.2
$r_b$	4
$\alpha_H$	28
$\alpha_L$	250
$\mu$	$\frac{1}{360}$
$N$	100 000

Таблица 2. Коэффициенты модели.

$\frac{1}{116}$ . Далее коэффициент перехода  $\rho_H$  из категории высокого риска с категорию низкого считаем равным 0.8, и наоборот - из низкого риска в высокий -  $\rho_L$  считаем равным 0.2. Вероятности заражения за одно взаимодействие для остро и хронически инфицированных индивидов  $\beta_A$  и  $\beta_C$  равны 0.015 и 0.001 соответственно. Все упомянутые коэффициенты, не изменяющиеся в рамках решения задачи, указаны в таблице 2.

## 3.2 Численный эксперимент

Для автоматизации вычислений и демонстрации применимости вышеизложенных методов и подходов, была осуществлена программная реализация модели на языке Python.

Для получения результатов использовались программные библиотеки *NumPy* и основанная на ней высокоуровневая *Pandas* для обработки и анализа данных, а также *SciPy* для выполнения научных и инженерных расчетов (в частности интегрирования) и *Matplotlib* для визуализации данных.

В Приложении А задаются основные коэффициенты-константы модели, затем в функции HIV рассчитываются коэффициенты для каждого

рассматриваемого момента времени (в данном случае раз в месяц), а затем решается система дифференциальных уравнений, результатом которой является нерепрезентативная выборка, смещенная в сторону высокорискового поведения индивидов.

В Приложении В происходит оценка долей рассматриваемых категорий с учетом нерепрезентативности данных.

Последним этапом реализации программы является оценка количества людей в различных состояниях с помощью решения системы уравнений в матричном виде, что описано в Приложении С.

Построение диаграмм представлено в Приложении D.

Программный код для самостоятельного ознакомления и возможных дальнейших исследований может быть найден по ссылке: [\*\*https://github.com/chris142/BachelorDiplomaHIV\*\*](https://github.com/chris142/BachelorDiplomaHIV)

### 3.3 Результаты

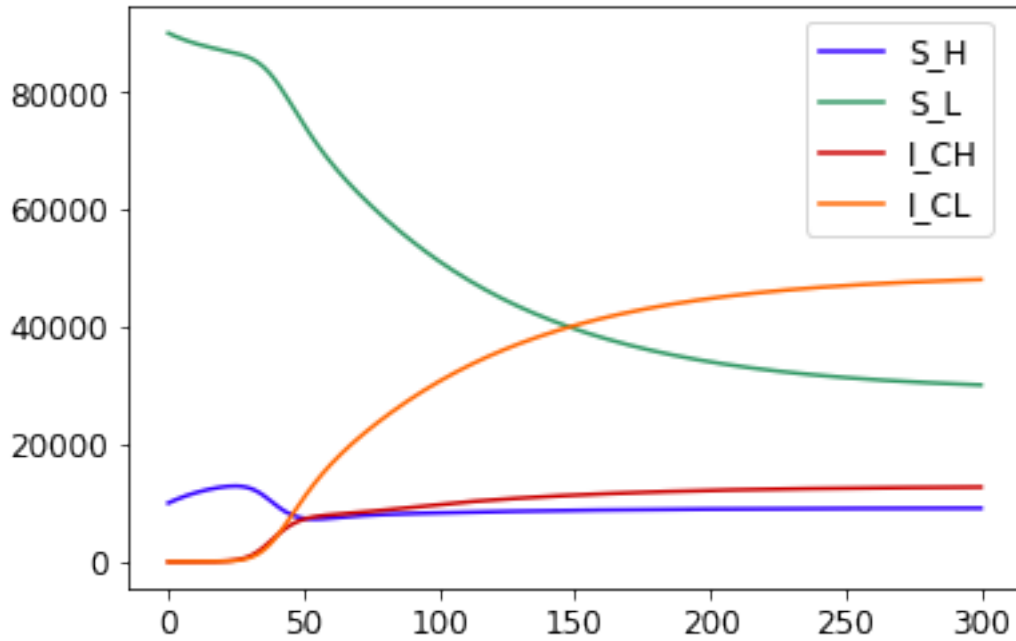


Рис. 4. Динамика популяции по нерепрезентативным данным.

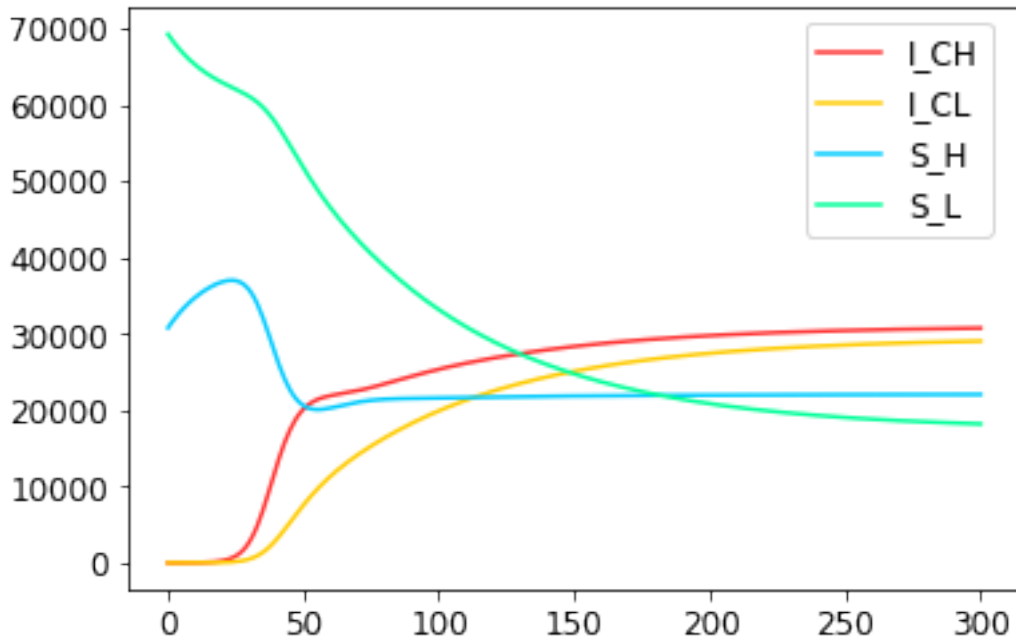


Рис. 5. Реальная динамика популяции.

Графики на рисунке 4 отражают динамику популяции за 300 месяцев на основе нерепрезентативных данных, полученных в результате выборки в высокорисковой среде. Зеленым и синим цветом выделены низкорисковые  $S_L$  и высокорисковые  $S_H$  здоровые индивиды соответственно, желтым

и красным - низкорисковые  $I_{CL}$  и высокорисковые  $I_{CH}$  хронически инфицированные индивиды соответственно.

На рисунке 5 показана искомая реальная динамика популяции. Зеленый, голубой, желтый и красный графики представляют  $S_L$ ,  $S_H$ ,  $I_{CL}$  и  $I_{CH}$  категории соответственно.

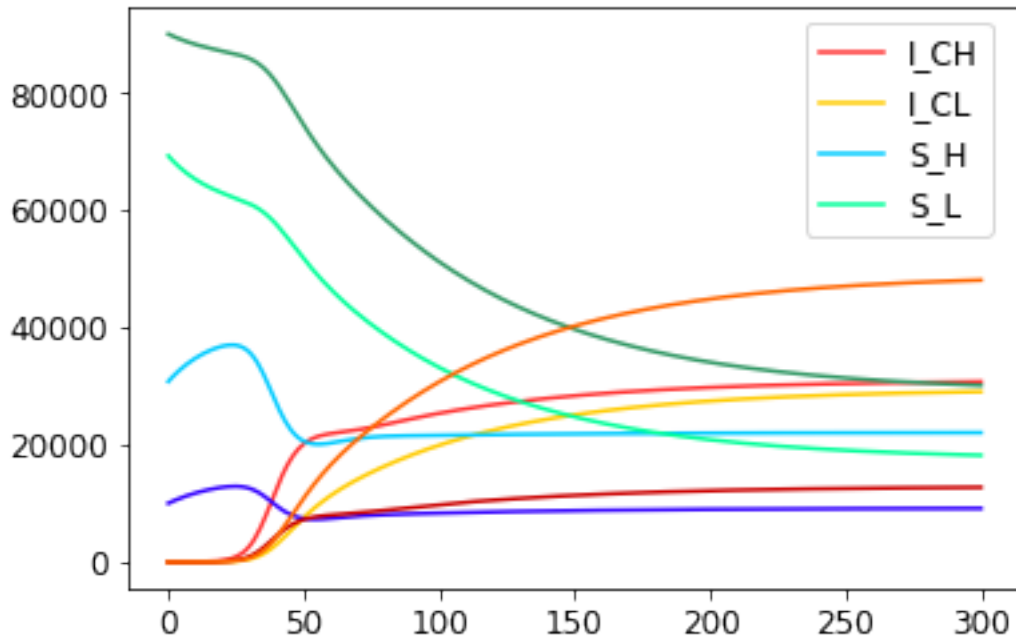


Рис. 6. Сравнение результатов.

Рисунок 6 демонстрирует существенную разницу в оценке популяции и ее развития до и после учета репрезентативности данных. Заметно снижение числа здоровых индивидов и рост числа инфицированных, придерживающихся высокорискового поведения. Стоит обратить внимание, что в общем число индивидов с высокорисковым поведением значительно больше, чем показывает тестирование и опросы. На основании чего можно сделать вывод о недостаточно высоком уровне жизни рассматриваемого населения.



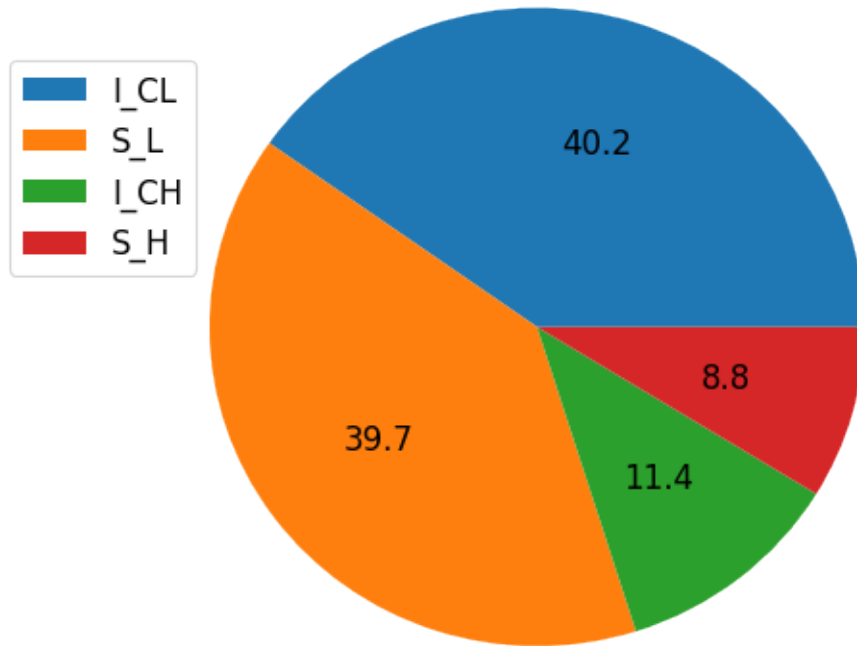


Рис. 7. Распределение населения по результатам тестирования,  $T = 150$ .

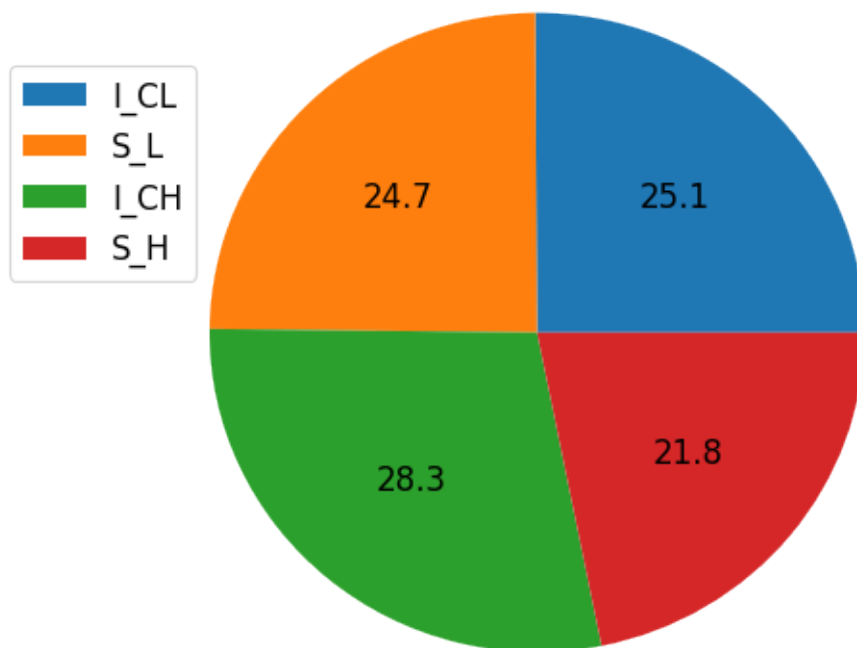


Рис. 8. Распределение населения после переоценки,  $T = 150$ .

Рассмотрим, например, отношение долей в популяции по результатам тестирования (рисунок 7) и после их переоценки (рисунок 8). Очевидно, преобладание индивидов с низким уровнем риска ложно, тогда можно

выдвинуть предположение о недостаточной пропаганде здорового образа жизни среди населения.

## Выводы

Для решения поставленной задачи была разработана математическая модель, ее программная реализация, и численный эксперимент, проведенный на основе имеющихся данных о результатах опроса и тестирования. Такой эксперимент позволяет сравнить видение ситуации непосредственно на основе опросов и после переоценки долей с использованием формулы Байеса. Зная число людей после опроса и тестирования можно легко рассчитать соотношение таких категорий во всей популяции, указав результаты тестирования в абсолютных величинах, размер выборки и размер населения.

Результаты эксперимента указывают на необходимость уделения повышенного внимания группе хронически инфицированных индивидов  $I_{CH}$ , рост которой был не так явно выражен до переоценки и коррекции данных.

Сложностью исследования является невозможность распознавания острой фазы инфицирования при анонимном единовременном тестировании, искажающая реальную картину течения заболевания. В будущем возможно более тщательное и подробное исследование, уделяющее особое внимание выявлению индивидов, находящихся в острой фазе инфицирования.

Интересным продолжением работы может стать использование временных рядов для прогнозирования динамики популяции.

## Заключение

В ходе данной работы были достигнуты следующие результаты:

Были проведен анализ различных существующих эпидемиологических моделей распространения инфекции.

На основе этого была составлена математическая модель, представляющая популяцию как совокупность шести различных групп, учитывающая два уровня рискованного поведения и различные стадии инфицирования рассматриваемого населения: острая и хроническая.

Приведен метод оценки соотношения категорий людей в выборке для нерепрезентативных данных, основываясь на их поведении, а затем вычисление долей в общей популяции для конкретного примера четырех состояний и для общего случая.

Проанализированы коэффициенты модели, осуществлена программная реализация и проведен численный эксперимент для конкретной выборки.

В дальнейшем приведенные методы и непосредственно программная реализация могут быть применены для составления плана лечения и профилактики населения. Это позволит прийти к решению, в какой мере инвестиции в предупреждающую и сдерживающую терапию должны быть увеличены в будущем для снижения уровня заболеваемости ВИЧ.

Несмотря на глобальное развитие общества, ВИЧ остается экономически и социально значимым инфекционным заболеванием и актуальной проблемой для здравоохранения. Отсутствие на сегодняшний день достаточного количества математических моделей, способных наиболее полно отразить особенности конкретного населения, подчеркивает актуальность проведенной работы.

## Список литературы

- [1] WHO (2017) Global health observatory (GHO) data: HIV/AIDS. <http://www.who.int/gho/hiv/en/> (last accessed 16 April 2018).
- [2] R. M. Anderson, R. M. May, and A. R. McLean, Possible demographic consequences of AIDS in developing countries, *Nature (Land.)* 332:228-234 (1988).
- [3] J. M. Hyman and E. A. Stanley, Using mathematical models to understand the AIDS epidemic, *Math. Biosci.* 90:415-474 (1988).
- [4] R. M. Anderson, S. P. Blythe, S. Gupta, and E. Konings. The transmission dynamics of the human immunodeficiency virus type 1 in the male homosexual community in the United Kingdom: the influence of changes in sexual behaviour, *Phil. Trans. Roy. Soc. B* 325:45-98 (1989).
- [5] R. M. Anderson, T. W. Ng, M. C. Boily, and R. M. May, The influence of different sexual contact patterns between age classes on the predicted demographic impacts of AIDS in developing countries, *Ann. N.Y. Acad. Sci.* 569:240-274 (1989).
- [6] R. M. May and R. M. Anderson, The transmission dynamics of human immunodeficiency virus (HIV), *Phil. Trans. Roy. Soc. B* 321:565-607 (1989).
- [7] H. W. Hethcote, A model for HIV transmission and AIDS, in *Mathematical Approaches to Problems in Resource Management and Epidemiology*, Springer-Verlag, Berlin, 1989, eds. C. Castillo-Chavez, S. A. Levin, C. A. Shoemaker, pp. 164-176.
- [8] Hall H. I., Song R., Rhodes P., Prejean J., An Q., Lee L. M., Karon J., Brookmeyer R., Kaplan E. H., McKenna M. T. & Janssen R. S. (2008) Estimation of HIV incidence in the United States. *JAMA* 300, 520–529.
- [9] Beyrer C., Baral S. D., Collins C., Richardson E. T., Sullivan P. S., Sanchez J., Trapence G., Katabira E., Kazatchkine M., Ryan O., Wirtz A. L. & Mayer K. H. (2016) The global response to HIV in men who have sex with men. *The Lancet* 388, 198–206.

- [10] Ю. Саранков. Медицинские потребности и проблемы мужчин, имеющих секс с мужчинами. // «СПИД Фонд Восток-Запад», 2006.
- [11] Paul Van de Ven, Pamela Rodden, June Crawford, Susan Kippax: A comparative demographic and sexual profile of older homosexually active men, *Journal of Sex Research*, 1997.
- [12] Данные «Объединённой программы ООН по ВИЧ/СПИДу». Декабрь 2006. <http://www.unaids.org/ru>.
- [13] СПИД и сексуальные отношения между мужчинами. Технический обзор, ЮНЭЙДС, 2000.
- [14] Diagnoses of HIV Infection in the United States and Dependent Areas. Surveillance Report, 2016.
- [15] С. Тюрин. Профилактика ВИЧ и ИППП среди МСМ. РОО «СПИД инфосвязь», 2007.
- [16] Gary De Young, Philip K. Main and Michael Nakamaye. Analysis of a Risk-Based Model for the Growth of AIDS Infection // Baltimore conference, August 1990
- [17] Jeremy M. G. Taylor. Models for the HIV infection and AIDS epidemic in the United States // *Statistics in Medicine*, 1989.
- [18] Punyacharoensin, N., Edmunds, W. J., De Angelis, D., White, R. G. Mathematical models for the study of HIV spread and control amongst men who have sex with men // *Eur J Epidemiol*, 2011.
- [19] Punyacharoensin, N., Edmunds, W. J., De Angelis, D., White, R. G. Effect of pre-exposure prophylaxis and combination HIV prevention for men who have sex with men in the UK: a mathematical modelling study // *Lancet HIV*, 2016.
- [20] Zhang X., Zhong L., Romero-Severson E., Alam S. J., Henry C. J., Volz E. M. & Koopman J. S. (2012) Episodic HIV risk behavior can greatly amplify HIV prevalence and the fraction of transmissions from acute HIV infection. *Stat. Commun. Infect. Dis.* 4.
- [21] D. Gromov, I. Bulla, E. O. Romero-Severson, and O. S. Serea. Numerical optimal control for hiv prevention with dynamic budget allocation // *Mathematical Medicine & Biology*, 2017.

- [22] D. Gromov, I. Bulla, E. O. Romero-Severson. Optimal Resource Allocation for HIV Prevention and Control // September 22, 2017
- [23] E. O. Romero-Severson, E. Volz, J. S. Koopman, T. Leitner, and E. L. Ionides. Dynamic variation in sexual contact rates in a cohort of HIV-negative gay men. *American Journal of Epidemiology*, 182(3):255–262, 2015.
- [24] S. E. Bellan, J. Dushoff, A. P. Galvani, and L. A. Meyers. Reassessment of HIV-1 acute phase infectivity: Accounting for heterogeneity and study design with simulated cohorts. *PLoS Med*, 12(3):e1001801, 2015.

## Приложение

### Приложение А. Описание модели и решение СОДУ.

---

```
# coding: utf-8
import pandas as pd
import numpy as np
from scipy import integrate
import matplotlib.pyplot as plt

# known coefficients of the model
mu = 1 / (12 * 30)
delta_A = 1 / 4
delta_C = 1 / 116
x = 1 / 12
y = 0
r_b = 0.8 / 0.2
lambda_H = 43.478;
lambda_L = 5.4896
beta_A = 0.015;
beta_C = 0.001
rho_H = 0.041667;
rho_L = 0.0046296
pi = 0.5

# HIV-model
def HIV(t, X):
    # a column vector
    dX = np.zeros((len(X), 1))

    # state variables
    # S_H = X(1)
```



```

# S_L = X(2)
# I_AH = X(3)
# I_AL = X(4)
# I_CH = X(5)
# I_CL = X(6)
S_H, S_L, I_AH, I_AL, I_CH, I_CL = X

# auxiliary variables
N_H = S_H + I_AH + I_CH
N_L = S_L + I_AL + I_CL
N = N_H + N_L

# calculated coefficients of the model
alpha = mu * N + delta_C * (I_CH + I_CL)
alpha_H = 0.9 * alpha
alpha_L = 0.1 * alpha

theta = lambda_H * N_H + lambda_L * N_L

sigma = (beta_A * (lambda_H * I_AH \
                  + lambda_L * I_AL) \
        + beta_C * (lambda_H * I_CH \
                  + lambda_L * I_CL)) / theta

psi_H = pi * lambda_H * (beta_A * I_AH / N_H \
                        + beta_C * I_CH / N_H)
psi_L = pi * lambda_L * (beta_A * I_AL / N_L \
                        + beta_C * I_CL / N_L)

tau_H = (1 - pi) * lambda_H * sigma
tau_L = (1 - pi) * lambda_L * sigma

phi_H = psi_H + tau_H
phi_L = psi_L + tau_L

```

```
# ODE
```

```
dS_H = alpha_H \
        - (phi_H + rho_H + mu) * S_H \
        + rho_L * S_L
```

```
dS_L = alpha_L \
        - (phi_L + rho_L + mu) * S_L \
        + rho_H * S_H
```

```
dI_AH = phi_H * S_H \
        - (rho_H + mu + delta_A) * I_AH \
        + rho_L * I_AL
```

```
dI_AL = phi_L * S_L \
        - (rho_L + mu + delta_A) * I_AL \
        + rho_H * I_AH
```

```
dI_CH = delta_A * I_AH \
        - (rho_H + mu + delta_C) * I_CH \
        + rho_L * I_CL
```

```
dI_CL = delta_A * I_AL \
        - (rho_L + mu + delta_C) * I_CL \
        + rho_H * I_CH
```

```
dX = np.array(
    [dS_H, dS_L, dI_AH, dI_AL, dI_CH, dI_CL]
)
return dX
```

```
# ODE solution
```

```
r = integrate.ode(HIV).set_integrator(
    "dopri5",
    atol=1e-8,
    rtol=1e-4,
```

```

)

# 300 months interval
X0 = np.array([10000, 89998, 1, 1, 0, 0])
T0 = 0
T1 = 300
dt = 1
X = []
# [0 300], 300 points
# for step = 1
t = np.array(list((i for i in range(T0, T1 + 1))))
r.set_initial_value(X0, T0)
for ti in t:
    X.append(r.integrate(ti))
T = np.array(t)
X = np.array(X)
data = pd.DataFrame(
X,
columns=['S_H', 'S_L', 'I_AH', 'I_AL', 'I_CH', 'I_CL'],
)
data['T'] = T

# the states are:
# S_H, S_L -- high- and low-risk susceptibles,
# I_AH, I_AL -- high- and low-risk infected in acute phase,
# we add S_H and I_AH (and the same for S_L and I_AL)
# as we cannot distinguish them,
# finally I_CH, I_CL -- high- and low-risk infected in
#                                     # chronical phase.

# ODE fractions plots
plt.plot(T, data.S_H.values + data.I_AH.values, \
          '#3300FF', label='S_H')
plt.plot(T, data.S_L.values + data.I_AL.values, \
          '#339966', label='S_L')

```

```
plt.plot(T, data.I_CH.values, \
         '#CC0000', label='I_CH')
plt.plot(T, data.I_CL.values, \
         '#FF6600', label='I_CL')
```

---

## Приложение В. Оценка долей в популяции.

---

```
# number of all high-risk people
N_H = data.S_H + data.I_AH + data.I_CH
# number of all low-risk people
N_L = data.S_L + data.I_AL + data.I_CL
N = N_H + N_L

# computing the corrected fractions
# high-risk susceptibles
F_SH = (r_b * data.S_H) \
        .divide(r_b * N_H + N_L) \
        .multiply(N) \
    + (r_b * data.I_AH) \
        .divide(r_b * N_H + N_L) \
        .multiply(N)
# low-risk susceptibles
F_SL = data.S_L \
        .divide(r_b * N_H + N_L) \
        .multiply(N) \
    + data.I_AL \
        .divide(r_b * N_H + N_L) \
        .multiply(N)
# high-risk infected
F_ICH = r_b * data.I_CH \
        .divide(r_b * N_H + N_L) \
        .multiply(N)
# low-risk infected
F_ICL = data.I_CL \
        .divide(r_b * N_H + N_L) \
        .multiply(N)

data2 = pd.DataFrame(
    np.array([F_SH, F_SL, F_ICH, F_ICL]).transpose(),
```

```

columns=['F_SH', 'F_SL', 'F_ICH', 'F_ICL']
)

# given fractions plots
plt.plot(T, data.S_H.values + data.I_AH.values, \
         '#3300FF', label='S_H')
plt.plot(T, data.S_L.values + data.I_AL.values, \
         '#339966', label='S_L')
plt.plot(T, data.I_CH.values, \
         '#CC0000', label='I_CH')
plt.plot(T, data.I_CL.values, \
         '#FF6600', label='I_CL')

# Bayes-corrected fractions plots
plt.plot(T, data2.F_ICH.values, \
         '#FF3333', label='F_ICH')
plt.plot(T, data2.F_ICL.values, \
         '#FFCC00', label='F_ICL')
plt.plot(T, data2.F_SH.values, \
         '#00CCFF', label='F_SH')
plt.plot(T, data2.F_SL.values, \
         '#00FF99', label='F_SL')

```

---

## Приложение С. Оценка количества людей в различных состояниях.

---

```
# matrix with r_b coefficients calculation
RB = pd.DataFrame(
    np.array([
        np.array([r_b, 0, 0]),
        np.array([0, r_b, 0]),
        np.array([0, 0, 1]),
    ]),
    index=['0', '1', '2'],
)

ANS = []
for i in range(0, 301):
    F = np.array([
        F_SH[i],
        F_ICH[i],
        F_ICL[i]
    ]) / 100000
    # diagonal matrix calculation
    DG = pd.DataFrame(np.array([
        np.array([F_SH[i], 0, 0]),
        np.array([0, F_ICH[i], 0]),
        np.array([0, 0, F_ICL[i]]),
    ])) / 100000
    # 1 and 0 matrix calculation
    SMTH = pd.DataFrame(
        np.array([
            np.array([1, 1, 0]),
            np.array([1, 1, 0]),
            np.array([1, 1, 0]),
        ])
    )
```

```

SMTH2 = ((1 - r_b) * DG).dot(SMTH)
COEF = RB + SMTH2.values
ans = list(np.linalg.pinv(COEF.values).dot(F))
# the last fraction calculation
# 1 - sum of others
ans.insert(1, 1 - sum(ans))
ANS.append(pd.Series(ans))
ANS = pd.DataFrame(
    ANS,
#     columns=['S_H', 'S_L', 'I_CH', 'I_CL'],
)
ANS = ANS.rename(
    columns={
        0: 'A',
        1: 'B',
        2: 'C',
        3: 'D',
    },
)
plt.plot(T, data.S_H.values + data.I_AH.values, \
         '#3300FF', label='S_H')
plt.plot(T, data.S_L.values + data.I_AL.values, \
         '#339966', label='S_L')
plt.plot(T, data.I_CH.values, \
         '#CC0000', label='I_CH')
plt.plot(T, data.I_CL.values, \
         '#FF6600', label='I_CL')

```

---



## Приложение D. Построение диаграммы.

---

```
import matplotlib as mpl
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
import datetime as dt
import csv

data_names = ['I_CL', 'S_L', 'I_CH', 'S_H']
data_values = [
    data2.F_ICL[150],
    data2.F_SL[150],
    data2.F_ICH[150],
    data2.F_SH[150],
]

dpi = 80
fig = plt.figure(
    dpi = dpi,
    figsize = (500 / dpi, 500 / dpi)
)
mpl.rcParams.update({'font.size': 9})

plt.title('Population distribution')

xs = range(len(data_names))

plt.pie(data_values, autopct='%.1f', radius = 1.1)
plt.legend(
    bbox_to_anchor = (-0.16, 0.45, 0.25, 0.25),
    loc = 'lower left',
    labels = data_names
)
```

---