

## Оценка распространенности употребления наркотиков\*

Управление Организации Объединенных Наций по наркотикам и преступности

Уважаемые коллеги, предлагаемый Вашему вниманию документ представляет собой Модуль 2 Инструментария Глобальной программы по оценке масштабов злоупотребления наркотиками (ГПО), разработанный при поддержке Управления Организации Объединенных Наций по наркотикам и преступности (UNODC) в рамках деятельности, осуществляемой в контексте ГПО. Модуль был подготовлен технической группой экспертов. Особая признательность выражается Колину Тэйлору, который осуществлял координацию проекта и редакцию настоящего издания, второму редактору Мэтью Хикману и Ребекке Маккетин, которая координировала и консультировала эту работу на заключительном этапе. В состав Группы технического консультирования входили следующие консультанты: Колин Тэйлор, Национальный наркологический центр, Лондон; Мэтью Хикман, Королевский колледж, Лондон; Майкл Лински, Национальный научно-исследовательский наркологический центр, штат Новый Южный Уэльс, Австралия; Лукас Виссинг, Европейский центр мониторинга наркотиков и наркомании; Пол Гриффитс, Ребекка Маккетин и Камран Ниаз, UNODC; Анинда Чаттерджи, Объединенная программа Организации Объединенных Наций по ВИЧ/СПИД, Таиланд. Помощь в этой работе оказывал Мэтью Уорнер-Смит, UNODC, ЮАР. UNODC хотело бы выразить признательность за оказанную поддержку многим национальным партнерам, которые опробовали настоящий инструментарий и прислали свои отклики, и институтам и частным лицам, которые представили образцы различных бланков для сбора данных и описание механизмов, а также другие соответствующие материалы. В частности, выражается благодарность Рабочей группе по эпидемиологии Сообщества, Группе Помпиду Совета Европы, Европейскому центру мониторинга наркотиков и наркомании, Эпидемиологической сети Сообщества по вопросам развития Юга Африки по проблеме употребления наркотиков, Восточноафриканской информационной системе по наркотикам и Сети информации о наркотиках Карибского бассейна.

### Использование метода двойного охвата

**Ч**то такое оценка по методу двойного охвата? Она заключается в том, что берутся обычные источники данных, идентифицирующие лиц, для которых потребление наркотиков носит проблемный характер, и с помощью этих источников осуществляется оценка распространенности, т.е. определяется частота случаев проблемного потребления наркотиков. В эпидемиологии или в социальных исследованиях источники данных представляют собой, как правило, списки лиц, в нашем случае лиц, для которых потребление наркотиков носит проблемный характер. Метод двойного охвата — это косвенный метод оценки распространенности, который предусматривает использование информации о наложениях между неполными списками или источниками данных. Списки считаются неполными, так как ни один источник не может перечислить всех лиц, для которых потребление наркотиков носит проблемный характер, — в противном случае оценка распространенности не была бы столь трудной, — так как редко когда система отчетности является совершенно безукоризненной и обеспечивает представление полной информации по наблюдаемым лицам. Наложения между списками — это лица, указанные более чем в одном списке.

Цель заключается в том, чтобы определить, сколько лиц, для которых потребление наркотиков носит проблемный характер, должно быть включено в список, чтобы он был полным списком интересующей нас целевой популяции, например всех лиц, для которых потребление наркотиков носит проблемный характер, в нашем городе в данном году. Отсюда Хук и Ригал (1995 г.) говорят об оценке по методу двойного охвата как о способе оценки с учетом «неполного установления». Поэтому такая оценка имеет явные преимущества при проведении эпидемиологических исследований в целях оценки распространенности того или иного заболевания на основе обычных источников данных.

### Обоснование

В технико-статистическом плане эта методика заключается в сведении данных в неполную множественную таблицу взаимной сопряженности признаков и в их анализе по стандартному статистическому методу логарифмически-линейного моделирования (или регрессии Пуассона). Затем определяется, сколько еще данных по данной модели необходимо внести в список, чтобы он был полным. Ниже объясняется, что такое таблица сопряженности и как проводится этот анализ.

\* Продолжение. Начало см. Наркология. — 2009. — №5.

Метод двойного охвата был разработан специалистами по экологии животных для оценки популяций оленей, рыбы или других животных в данном месте. Источниками данных являются выборка животных, пойманных, помеченных и затем отпущенных, и вторая выборка, показывающая, сколько животных было помечено в ходе первой выборки (наложение). Отсюда и название — двойной охват.

В области эпидемиологии человека партии отловленных и помеченных животных заменяются списками лиц, «охваченных» в обычном источнике данных. В настоящее время метод двойного охвата играет важную роль в эпидемиологии наркотизма, его можно приспособлять и применять в большинстве ситуаций на местах.

#### Метод двойного охвата с использованием двух источников данных

#### Тематическое исследование 5. Практический пример на основе исследования, проведенного Mastro и его коллегами в Бангкоке в 1991 г.

Были использованы два источника данных:

1) списки употреблявших опиаты лиц, включенных в программы лечения метадонем в наркологических клиниках Бангкока. Эти списки были составлены на основе регулярно собиравшихся сведений о лечении пациентов в период с апреля по май 1991 г.;

2) список лиц, арестованных бангкокской полицией в июне—сентябре 1991 г., проверка которых на опиаты дала положительные результаты. Группы исследователей 891 раз посетили 72 полицейских участка для выявления лиц, употребляющих опиаты, и сбора мочи и сведений о задержанных: лица с положительными результатами анализа на опиаты составили второй источник данных.

На основе статистических данных о лечении метадонем, представленных специализированными клиниками, было выявлено 4064 чел., потреблявших опиаты; из 8212 лиц, опрошенных в полицейских участках, было выявлено 1540 чел., употреблявших опиаты. Эти два источника позволяют составить неполные списки лиц, потреблявших опиаты в Бангкоке в данный период времени, — неполные по той причине, что каждый из списков представляет собой неполный перечень всех потреблявших опиаты лиц, которые входят в обследуемую популяцию.

Для выявления наложений в двух источниках лиц, потреблявших опиаты, в каждом из них полностью указывались имена, пол и даты рождения, на основе которых и производилось сопоставление данных. В результате выяснилось, что в обоих источниках значилось 171 лицо как потреблявшие опиаты, ибо они находились на лечении метадонем, были арестованы и анализ их мочи дал положительную реакцию на опиаты.

В табл. 5 приводятся сведения из обоих источников и указываются случаи их наложения. Данные сведены в таблицу сопряженности, на основе которой рассчитывается показатель распространенности. По обоим источникам данных (лиц, проходивших лечение метадонем, задержанных и давших положительную реакцию мочи на опиаты) мы допускаем, что из общего числа лиц, потреблявших опиаты, доля лиц, зарегистрированных в наркологических клиниках, аналогична доли лиц, выявленных среди арестованных, а также среди неарестованных. Заполнив таблицу исходя из этого допущения, можно получить оценку, согласно которой в Бангкоке в 1991 г. доля лиц, потреблявших наркотики путем инъекций, составляла 36 600 чел. (0,5% от всего населения).

Таблица 5

Оценка числа лиц, употреблявших опиаты, Бангкок, 1991 г.

Обнаружены в И1 <sup>a</sup>	Обнаружены в И2 <sup>b</sup>	Число выявленных лиц	Ключевой параметр
Да	Да	171	Совпадают в И1 и И2 (m)
Да	Нет	3893	Находятся только в И1 (c)
Нет	Да	1369	Находятся только в И2 (b)
Нет	Нет	—	"Скрытая численность" (x)
Всего в И1	Всего в И2	Всего	Число лиц, употреблявших опиаты, в популяции (N)
4064 (m + c)	1540 (m + b)	—	

Примечание. <sup>a</sup>И1 = пациенты клиник, проходившие курс поддерживающего лечения метадонем; <sup>b</sup>И2 = арестованные с позитивным анализом мочи

Таким образом, число лиц, употреблявших опиаты в Бангкоке в 1991 г., подсчитано следующим образом:

$$\text{наблюдаемое число} = m + b + c = 171 + 3893 + 1369 = 5433$$

$$\text{скрытое число (x)} = 1369 \times 3893 / 171 = 31\,166$$

$$\text{оценка популяции} = 5\,433 + 31\,166 = 36\,599$$

округленная оценка числа лиц, употреблявших наркотики путем инъекций, в Бангкоке в 1991 г. = 36 600 (32 000 — 40 800) (95%-ный интервал достоверности)

В этом примере, как и в экологическом исследовании, были взяты два различных периода времени (апрель—май и июнь—сентябрь), показывающие, что второй охват происходит после первого. Для эпидемиологических исследований это не важно. Для них можно было бы взять оба источника данных за один и тот же период времени (см. «Допущения», ниже). Кроме того, была подсчитана общая численность популяции за весь 1991 г. Строго говоря, в Бангкоке было рассчитано общее число лиц, употреблявших опиаты путем инъекций, лишь за указанный период времени, т.е. с апреля по сентябрь, однако вполне можно допустить, что эти данные соответствуют численности за год. Другими словами, можно допустить, что потенциальное число употреблявших опиаты лиц, которые не были охвачены оценкой, является ничтожным, т.е. речь в данном случае идет о лицах, только приступивших к употреблению опиатов, либо о лицах, прекративших их употребление, умерших или выехавших из данного района.

Если обратиться к первоначальным определениям, то станет ясно, что табл. 5 является примером таблицы сопряженности — таблицы сопряженности 2 x 2 с четырьмя ячейками. Таблица неполная, так как неизвестно число лиц, не включенных в оба источника данных, т.е. «скрытых» или ненаблюдаемых лиц, употребляющих наркотики путем инъекций. Другими словами, оба источника данных не «полностью охватывают» целевую популяцию, т.е. общее число лиц, употребляющих наркотики путем инъекций.

В рамках методов двойного охвата эта таблица используется для подсчета числа лиц, необходимого для того, чтобы оба списка стали полными. Допущения по этому методу приводятся ниже, но одно следует все-таки назвать сейчас. В случае применения метода двойного охвата с двумя выборками следует исходить из того, что включение тех или иных данных в один источник не зависит от того, включены ли эти данные в другой. Это допущение означает, что существует одинаковая вероятность того, что за лечением в клинику обратятся лица, потреблявшие наркотики, как из числа арестованных, так и из числа тех, кто не был арестован; и наоборот, если кто-либо, для кого потребление наркотиков носит проблемный характер,

прошел курс лечения метадонном, то это не означает, что его обязательно должны арестовать.

В ходе исследования были проанализированы данные о 5433 потреблявших опиаты лицах, указанных в сводных данных из обоих источников, и было подсчитано, что их общее число составляет 36 600 чел., другими словами, исследованием охватывалось одно из семи (5433/36 600) лиц, потреблявших опиаты. Это относительно высокий показатель, а его точность зиждется на допущении, что оба источника данных не зависят один от другого. Если бы они не были взаимно не зависимыми, а этого, исходя из результатов исследования по двум выборкам, утверждать нельзя, то чем больше соотношение между наблюдаемой и общей популяцией, тем больше вероятность ошибки.

Это важно иметь в виду при планировании исследования. В идеальном случае можно собрать данные о значительном числе людей, с тем чтобы сократить соотношение между наблюдаемыми и ненаблюдаемыми лицами, однако в исследованиях, связанных с проблемным потреблением наркотиков, это не всегда возможно. С другой стороны, можно провести исследование по методу двойного охвата с использованием множественных источников данных, о чем говорится в следующем разделе. Конечно, можно также провести исследование по двум выборкам, однако для этого рекомендуется уточнить, можно ли рассматривать оба источника данных как независимые либо установить шкалу зависимости, с тем чтобы рассматривать результаты оценки как минимальные, если между источниками данных существует положительная зависимость, и как максимальные, если между источниками данных существует отрицательная зависимость.

#### Оценка по методу двойного охвата данных из нескольких источников

Под термином «несколько источников данных» понимаются три или более списка, например лица, для которых потребление наркотиков носит проблемный характер, которые были арестованы, которые находятся на лечении, которые ночуют в ночлежных домах либо которые были доставлены в травмопункты или отделения неотложной помощи. Затем эти несколько источников данных сопоставляются для определения частичного совпадения по каждой комбинации из двух или более источников данных.

Полученные таким образом данные анализируются на основе регрессии Пуассона или программы логарифмически-линейного моделирования. Примером простейшего логарифмически-линейного моделирования является приводимая выше таблица сопряженности. Такое моделирование настолько просто, что для этого не нужна ни компьютерная программа, ни даже компьютер. Можно обойтись без описания расчетов с

использованием трех или более источников данных для оценки числа лиц, для которых потребление наркотиков носит проблемный характер. Для этого вполне можно использовать пакеты статистических программ, которые в письменном виде выглядят чрезвычайно сложными. Допущение независимости как в предыдущем примере, столь важное для расчетов, может быть сделано в нескольких различных формах, которые рассматриваются в следующем разделе.

Читателям, интересующимся уравнениями расчета, следует обратиться к Бишопу и другим (1975 г.), а также к Хуку и Ригалу (1995 г.). В целом, для расчетов по методу двойного охвата с использованием нескольких источников данных рекомендуется пользоваться статистической поддержкой на местах для разъяснения или применения регрессии Пуассона (или программы логарифмически-линейного моделирования), необходимых для анализа данных. В инструкции ЕЦМНН подробно разбираются примеры использования пакета программ для общего линейного интерактивного моделирования (ОЛИМ) (программа интерактивного статистического анализа) и пакета статистических программ для общественных наук (СПОН) в ходе двойного охвата.

#### **Зависимости между источниками данных**

Зависимость между любыми двумя источниками (И1 и И2) считается положительной, если лицо, указанное в И1, будет скорее значиться в И2, чем лицо, не указанное в И1, или отрицательной, если лицо, не указанное в И1, вряд ли будет значиться в И2, чем лицо, не указанное в И1. Эти зависимости можно проверить в качестве «условий взаимодействия» в рамках анализа модели оценки на основе регрессии Пуассона, и их необходимо учитывать и проверять в ходе подгонки модели к данным. Можно также проверить более сложные виды взаимодействия или зависимости, например, включающие в себя комбинацию из трех источников данных. К данным можно подогнать несколько моделей регрессии Пуассона, в соответствии с которыми зависимости включаются или исключаются. Лучшая модель обычно определяется путем сопоставления, какая из них даст наилучший прогноз данных, измеренных с помощью стандартных статистических коэффициентов «неадекватности».

#### **Подгонка модели**

Подобрать модель к данным из нескольких источников можно практически следующим образом:

1) подготавливаются три или более источника данных по лицам, для которых потребление наркотиков носит проблемный характер;

2) эти источники данных сопоставляются и выявляются лица, включенные более чем в один источник данных, а также эти источники данных;

3) подготавливается таблица с несколькими источниками данных для анализа в файле данных;

4) проводится анализ таблицы с использованием регрессии Пуассона;

5) осуществляется подгонка модели с различными видами взаимодействия между источниками данных к возможным видам зависимости;

6) выбирается оптимальная модель по стандартным замерам неадекватности (в связи с этим следует проконсультироваться со статистиком);

7) с помощью модели оценивается число лиц, составляющих ненаблюдаемую часть популяции, и рассчитываются доверительные интервалы;

8) весь анализ повторяется по возможности по различным подгруппам (мужчины, женщины, различные возрастные группы и т.д.).

Ниже приводится пример анализа метода двойного охвата с использованием нескольких источников данных в рамках исследования, проведенного в Глазго в 1993 г. Фришером и др.

#### **Тематическое исследование 6.**

##### **Проведенное в Глазго исследование лиц, употреблявших наркотики путем инъекций (Исследование по методу двойного охвата с использованием нескольких источников данных со стратификацией)**

Frischer и его коллеги (1993 г.) под целевой популяцией имеют в виду лиц, употребляющих путем инъекций различные наркотики, а не только героин, так как в Глазго обычно путем инъекций потребляется значительное число разных видов наркотиков, хотя, строго говоря, в рамках исследования были собраны данные о потреблении самых различных видов наркотиков (героина, других опиатов, амфетамина, кокаина и бензодиазепинов).

Исследователями были собраны из четырех источников данные за год о лицах:

1) с позитивной и негативной реакцией на ВИЧ, для которых потребление наркотиков путем инъекций имело опасный характер;

2) обратившихся в специализированные наркологические учреждения за лечением наркомании (в том числе от злоупотребления героином, другими опиатами, кокаином, амфетамином и бензодиазепинами);

3) арестованных за преступления, связанные с наркотиками (кроме каннабиса);

4) зарегистрированных в пунктах обмена игл.

В рамках такого исследования по методу двойного охвата с использованием нескольких источников данных лица, употребляющие наркотики путем инъек-

## МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ НАРКОЛОГИИ

ций, сопоставляются по всем четырем источникам с тем, чтобы можно было подсчитать число лиц, выявленных по каждой комбинации точек выборки. В общей сложности были собраны сведения о 3444 лицах: сведения о 508 лицах были получены из полиции (И1), о 1179 — из пунктов обмена игл (И2), о 507 — из лабораторий по проверке на ВИЧ (И3) и о 1250 — из специализированных наркологических учреждений (И4). После сопоставления на предмет совпадения было выявлено 578 лиц, которые были обнаружены более чем в одном источнике данных (например, 4 — во всех четырех источниках, 41 — в И2, И3 и И4 и 147 — в И2 и И4). Таким образом, общая численность лиц, употреблявших наркотики путем инъекций и охваченных исследованием, составила 2866 чел.

Для исследования в Глазго было подобрано несколько моделей, начиная от модели с допущением о том, что все источники являются независимыми — показатели взаимодействия не устанавливались, — до модели, предусматривавшей сложную взаимозависимость, в рамках которой устанавливалось взаимодействие между всеми тремя источниками данных.

Полностью независимая модель совершенно не подошла.

Оптимальная модель для данных в табл. 5 включала в себя взаимодействия между тремя источниками данных: данными пунктов обмена игл (И3), данными результатов анализа на ВИЧ (И2) и данными о лечении наркомании (И4). В отличие от этих данных, полицейские протоколы об арестах за хранение (И1) в рамках этой оптимальной модели не использовались для определения показателя взаимодействия, что свидетельствует о независимости этого источника от всех остальных источников данных.

Взаимодействие между источниками было обычно положительным. Это означало, что лица, потреблявшие наркотики путем инъекций и включенные в один из списков, скорее всего будут также значиться в каком-то другом списке. В этом случае, если источники данных взаимозависимы, подгонка модели, допускающей полную независимость источников данных, может привести к занижению оценки ненаблюдаемой части совокупности. В данном исследовании результаты оценки на основе полностью независимой модели оказались фактически ниже результатов оценки по конечной модели.

Таблица 6

Данные исследования по методу двойного охвата с использованием нескольких источников данных для исследования популяции лиц, потреблявших наркотики путем инъекций, проведенного в Глазго, Шотландия, в 1990 г.

Обнаружены в И1 <sup>a</sup>	Обнаружены в И2 <sup>b</sup>	Обнаружены в И3 <sup>c</sup>	Обнаружены в И4 <sup>d</sup>	Число выявленных лиц
Да	Да	Да	Да	4
Да	Да	Да	Нет	2
Да	Да	Нет	Да	13
Да	Да	Нет	Нет	56
Да	Нет	Да	Да	8
Да	Нет	Да	Нет	17
Да	Нет	Нет	Да	50
Да	Нет	Нет	Нет	358
Нет	Да	Да	Да	41
Нет	Да	Да	Нет	52
Нет	Да	Нет	Да	147
Нет	Да	Нет	Нет	864
Нет	Нет	Да	Да	116
Нет	Нет	Да	Нет	267
Нет	Нет	Нет	Да	871
Нет	Нет	Нет	Нет	—
Всего в И1 508	Всего в И2 1791	Всего в И3 507	Всего в И4 1250	Общая популяция —

Примечание. <sup>a</sup>И1 = арестованные за хранение незаконных наркотиков (за исключением каннабиса); <sup>b</sup>И2 = зарегистрированные по программе обмена игл; <sup>c</sup>И3 = проверявшиеся на ВИЧ как наркоманы на игле; <sup>d</sup>И4 = находившиеся на лечении в специализированных наркологических клиниках

По оценке исследователей, наркотики путем инъекций потребляли в целом 8500 лиц, или 1,35% взрослого населения Глазго в возрасте от 15 до 54 лет. В рамках исследования было собрано достаточно данных для дальнейшего анализа (т.е. для дальнейшего моделирования) данных по полу и возрастным группам. Эти данные представлены в табл. 6 и являются примером стратификации, разрешающей проблему неоднородности.

Наиболее подходящей является модель, которая включала в себя взаимодействия между источниками И2 и И3, И2 и И4, И3 и И4, а также трехстороннее взаимодействие между И2, И3, И4; источник И1 не зависел от всех остальных. Эта модель дала результаты, представленные в табл. 7.

Одна особенность исследования в Глазго заключалась в том, что источник данных полиции был совершенно независимым от других источников. Три остальных источника данных (И2, И3 и И4) были зависимыми. Их можно было бы объединить в единый список данных под названием «Другие источники», а по данным полиции провести простое исследование на основе двух источников. Есть смысл также вернуться к тематическому исследованию 1, выше, и рассчитать показатель распространенности на основе расчетов данных из двух показанных источников. В этом случае были бы получены следующие данные:

1) объединенный список лиц, обратившихся за лечением, регистр ВИЧ, список лиц, пользовавшихся услугами пунктов обмена игл — 2508;

2) список арестованных полицией — 508;

3) число совпадений между списком «полиции» и «объединенным» списком — 150.

Эти расчеты дают такой же ответ, как и вышеупомянутая регрессия Пуассона с использованием множественных источников данных, потому что особые обстоятельства оценки множественных источников данных свидетельствуют о том, что такая оценка ана-

логична оценке с использованием двух независимых источников данных. И это вполне понятно: если кто-то внесен в список лиц, обратившихся за лечением, то можно, очевидно, предположить, что это лицо, весьма вероятно, зарегистрировано также на пункте обмена игл, и тем более, если оно внесено в регистр ВИЧ — здесь прослеживается взаимозависимость между тремя источниками. С другой стороны, по крайней мере, очевидно, что лица, задержанные полицией, были арестованы случайно во время облав на наркоманов независимо от того, обращались ли они за лечением, были ли зарегистрированы на пунктах обмена игл или включены в регистр ВИЧ.

В данном случае мы пытаемся показать, что исследователь не мог знать, что источник данных полиции является независимым, а все другие источники данных — зависимыми, без исследования на основе множественных источников данных и анализа данных с использованием моделей регрессии Пуассона. На этот вопрос можно было бы ответить с помощью исследования по двум источникам, однако проверить даже самые очевидные допущения, положенные в основу данного исследования, совершенно невозможно.

Этот пример свидетельствует также о том, что лишь в особых обстоятельствах, как, например, в данном исследовании, следует объединять различные источники данных в один источник. Хотя и считается нормальным, что шансы на «охват» лиц, для которых потребление наркотиков носит проблемный характер, являются различными в зависимости от источника данных, важно подчеркнуть, что у каждого члена совокупности шанс должен быть таким же, как и у любого другого наблюдаемого лица в любом данном списке. Это, вероятно, будет невозможно, если объединить списки из совершенно различных источников данных, например, регистр ВИЧ и список арестованных полицией, и это будет противоречить допущению об однородности.

Таблица 7

Оценка числа лиц, употреблявших наркотики путем инъекций, в Глазго — число и показатель распространенности

Наименование	Замеченные/известные	Оценка скрытой совокупности	Итоговое число	Расчетный показатель распространенности (%)
Все	2866	5628	8494	1,4
Мужчины	1977	3567	5544	1,8
Женщины	889	2349	3238	1,0
15—19 лет	264	640	904	1,0
20—24 лет	1137	1613	2750	2,6
25—29 лет	878	1724	2602	2,7
30—34 лет	342	796	1138	1,4
> 35 лет	245	1273	1518	0,6

### Источники данных и выявление случаев совпадения между ними

#### Критерии использования источников данных

В рамках исследований методом двойного охвата оптимальные источники данных в идеальном случае позволяют следующее:

- а) четко идентифицировать целевую популяцию, подлежащую оценке (например, лиц, употребляющих героин путем инъекций, или лиц, для которых употребление наркотиков носит проблемный характер);
- б) собирать огромное множество данных, которые можно использовать для выявления случаев совпадения;
- с) собирать возможные стратификационные переменные;
- д) предоставлять данные в электронной форме, с тем чтобы не сличать их вручную.

К сожалению, наблюдается это редко. Поэтому необходимо обеспечить достаточно надежные источники данных, в частности относительно нижеследующих вопросов:

- 1) позволяют ли они идентифицировать целевую популяцию, которую мы хотим оценить?
- 2) какие идентификаторы собираются и какие данные будут предоставлены?
- 3) как будут собираться или предоставляться данные?
- 4) сколько случаев будет охвачено?
- 5) как эти источники данных будут согласовываться с другими источниками данных?

Последний вопрос относится к одному из допущений по методу двойного охвата (см. «Допущения», ниже): являются ли источники данных представительными с точки зрения целевой популяции. Следует воспользоваться источниками данных системы уголовного правосудия и системы здравоохранения, если есть такая возможность.

Можно воспользоваться обычными источниками данных, с этой целью лучше всего произвести инвентаризацию в изучаемой области и определить, сколько источников данных имеется на местах и существуют ли какие-либо другие потенциальные источники. В суммарной форме (будучи сгруппированными) их можно использовать в качестве опорных данных для оценки на основе множительных методов, а в дезагрегированной форме (несгруппированными в одну линию данных по каждому сообщению) — для исследований по методу двойного охвата. Следует помнить, что источники данных для проведения оценки по методу двойного охвата необязательно должны быть полными (т.е. включать каждый возможный случай), однако они должны быть точными и надежными (т.е. содержать правильные идентификаторы и точную информацию о наркотиках).

Источники данных с небольшими числами можно в какой-то мере объединять с другими источниками

данных, хотя здесь есть свои сложности. Исследования по методу двойного охвата на основе более пяти источников данных сложны для анализа. Это объясняется потенциально большим числом моделей, которые придется подгонять в случае использования большего числа источников данных.

#### Выделение совпадений между источниками данных

Если только данных не тысячи, то их лучше всего сопоставлять вручную. Выявлять совпадения можно также с помощью электронных таблиц или баз данных, которые позволяют по-разному сгруппировать источники данных, с тем чтобы выявить лиц, включенных в обе базы данных. Например, можно взять два списка в разбивке по полу и датам рождения и определить, имеются ли еще лица с таким же именем или инициалами; либо взять списки в разбивке по полу и фамилиям и определить, имеются ли в них одинаковые даты рождения.

Следует также определить, что является совпадением, поскольку различия между источниками данных могут объясняться ошибками при нажатии клавиши или незначительными вариациями имен собственных. Например, в одной базе данных может значиться Уильям Шекспир, а в другой — Билл Шекспир, в одной стоять дата рождения 26 апреля 1563 г., а в другой — 20 апреля 1563 г. Чем больше будет данных по одному и тому же лицу, тем легче будет их сравнивать, однако придется мириться с возможными ошибками в источниках. В целом, если дата рождения и пол одни и те же и одно из имен собственных является одним и тем же, то в большинстве исследований это будет рассматриваться как совпадение.

Важно, чтобы выбираемые критерии были четкими и конкретными. Для этого, например, можно принять строгое определение, провести анализ, а затем взять более свободное определение и сравнить показатели распространенности.

#### Исследования без использования стандартных источников данных

Что делать, если нет местных источников данных или ни по одному из них нельзя получить какую-либо идентифицирующую информацию или это невозможно из-за требований конфиденциальности? Отказаться или попробовать другой метод? Может быть, произвести расчеты по методу двойного охвата на основе обследования лиц, для которых потребление наркотиков носит проблемный характер? Такая методика очень подходит, если с географической точки зрения данная совокупность является дискретной и имеет нечто общее, что можно было бы использовать для идентификации ее членов: например, лица, употребляющие наркотики путем инъекций на улице, или уличные проститутки, употребляющие наркотики путем инъекций. Здесь приведены два примера такой методики (тематические исследования 7 и 8).

**Тематическое исследование 7.****Исследование, проведенное в Дакке, Бангладеш, для оценки распространенности проституции (двойной охват без использования стандартных баз данных)**

В настоящем примере эта методика использовалась для оценки числа уличных проституток в Дакке. Ввиду отсутствия каких-либо дополнительных данных о популяции проституток в городе и обычных источников данных, определяющих эту популяцию, было проведено полное этнографическое обследование. Для этого было опрошено несколько категорий ключевых респондентов, в том числе проститутки, сутенеры, рикши и водители такси, полицейские и местные мастаны («крутые»). Выяснилось, что в городе проституция носит практически полностью уличный характер. Проститутки «работали» на определенных улицах, углах и в парках города, особенно поздно вечером. Весь город был поделен на районы, в которых проститутки предлагали свои услуги.

С помощью самих проституток во всех известных заведениях и кварталах города были распространены цветные карточки (сначала красные). Карточки раздавались с позднего вечера до полуночи. По цветной карточке можно было пройти бесплатное медицинское обследование в одной из общих клиник города. Поскольку карточки были номерные, можно было легко составить выборку распространенных карточек, т.е. выборку охвата. Эта процедура была повторена через 7 дней. На этот раз карточка, которую также можно было использовать для бесплатного лечения, была зеленого цвета. Второй раз проституткам задавался только один вопрос, получили ли они до этого такую же красную карточку. На основе этих данных путем простого расчета в тематическом исследовании 5 было выведено предположительное число уличных проституток (приблизительно 5 тыс. чел.).

**Тематическое исследование 8.****Обследование уличных проституток в Глазго (двойной охват без использования стандартных источников данных)**

В Глазго был обследован основной район красных фонарей. Было установлено, сколько уличных проституток работает в определенный период времени, сколько работает каждую ночь и учитывались ли они в ходе обследований, проводившихся до этого. В общей сложности более чем за 7 мес. было установлено 1145 контактов с 206 женщинами (147, или 71% из них потребляли наркотики путем инъекций). По каждой женщине, с которой был установлен контакт, были собраны идентификаторы, состоявшие из начальных букв имени и фамилии и года рождения, что позволило составить

«историю охвата» за обследуемый период. Анализ историй охвата женщин с использованием «открытых методов двойного охвата» позволяет предположить, что их популяция каждую ночь неизменно насчитывала приблизительно 200 чел., но в неделю эта численность колебалась приблизительно на 8% в ту или иную сторону. В итоге можно подсчитать, что за год на улицы выходили около 1150 проституток.

Примененная в Бангладеш методика представляет собой простое исследование по методу двойного охвата с использованием двух источников данных.

Каких-либо идентифицирующих сведений о лицах не собиралось, однако были идентифицированы случаи «совпадения», поэтому для оценки общей популяции можно было использовать указанные выше расчеты. Эта методика считается исследованием по методу двойного охвата в связи с тем, что совпадения из источника данных 2 были фактически идентифицированы как лица, указанные в источнике данных 1.

Методика, использовавшаяся для проведения исследования в Глазго, предусматривала сбор некоторых идентификаторов для выявления совпадений между членами популяции, работавших ночью. Для подсчета размеров популяции был задействован другой набор статистических методов. Вопрос об использовании такой методики рекомендуется рассматривать только после консультаций со статистиком, имеющим опыт работы по открытым моделям двойного охвата и проведению оценки.

**Достоверность оценки**

Откуда известно, что результаты обследования достоверны? Ответ простой — ниоткуда. Результаты могут быть неверными. Однако, чтобы судить, насколько результаты достоверны, можно воспользоваться информацией самих исследователей на местах, т.е. поступить так же, как поступают специалисты по экологии животных, чтобы снять неопределенность в методе. В таком случае будут поставлены следующие вопросы:

- 1) отвечают ли результаты оценки нашим ожиданиям?
- 2) соответствуют ли они доказательственной базе, т.е. оценкам, полученным с использованием других методов, или оценкам за другие годы?
- 3) являются ли они до смешного низкими или позволяют считать, что каждый пятый человек потребляет наркотики путем инъекций? Если это так, то, очевидно, результаты неверны.
- 4) является ли доверительный интервал ниже 95% негативным? Если да, то можно использовать другие методы, но необходимо будет проверить точность источников данных и выявить случаи совпадения.
- 5) в рамках метода двойного охвата с использованием множественных баз данных являются ли показатели зависимости между базами данных достоверными?



**Тематическое исследование 9.**

**Исследование, проведенное на острове Джерси (лица, употреблявшие героин, называли вымышленные имена)**

Преимущество метода двойного охвата с использованием множественных источников данных заключается в том, что если возникают проблемы, то они могут быть связаны только с одним источником данных и в этом случае от использования этого источника можно отказаться, а результаты оценки пересчитать. Например, в ходе экспресс-оценки на Джерси была проведена оценка распространенности. Джерси — это относительно богатый остров, расположенный в проливе между Англией и Францией, с населением менее 100 тыс. чел. Было задействовано 5 источников данных: общинное обследование лиц, употреблявших наркотики путем инъекций; список находившихся на наркологическом лечении лиц, употреблявших героин; регистр наркоманов, составленный на основе информации, которую врачам предписывалось представлять о лицах, злоупотребляющих опиатами и кокаином; список пострадавших от передозировки героином лиц, доставленных в травмпункты и отделения неотложной помощи, и список лиц, арестованных полицией за правонарушения, связанные с наркотиками, в частности с героином. По первой модели с использованием всех пяти источников данных были получены результаты, свидетельствующие о том, что контингент лиц, употреблявших наркотики путем инъекций, составляет свыше 2000 чел. и что соответственно коэффициент распространенности на Джерси очень высокий (в целом свыше 2% и более 4% для возрастной группы от 15 до 54 лет).

Авторы догадывались, что в ходе общинного обследования опрашиваемые называли вымышленные имена. Это подтверждается сопоставлением уровня совпадений между данными общинного обследования и наркологической службы (10% совпадений) с долей респондентов в общинном обследовании, сообщивших о том, что в предшествующем году они лечились от наркомании (свыше 30%). Очевидно, что-то было не так, поскольку ожидалось, что совпадения данных между обоими источниками (данных лечения и общинного обследования) составят около 30%. Без общинного обследования был получен более приемлемый и достоверный результат (около 750 чел., или в целом 0,8%).

В данном случае это пример «классификационной ошибки», который также говорит о важности опыта и знаний, приобретенных на местах, так как таким образом можно выявить результаты оценки, не внушающие доверия, а также возможные смещения.

**Допущения**

В заключение будут рассмотрены допущения, лежащие в основе методики двойного охвата. Эти принципы носят в основном не теоретический, а практический характер, поэтому они будут излагаться вкратце. Однако важно отдавать себе отчет в том, какие допущения делаются, с тем, чтобы можно было правильно истолковать результаты двойного охвата. Последствия каждого допущения будут рассматриваться с точки зрения практики использования метода двойного охвата.

**Закрытая популяция — в ходе исследования не учитываются ни смертные случаи, ни новые лица, ни случаи включения или исключения из исследования**

Очевидно, что такое условие гарантировать невозможно. Поэтому имеется в виду, что продолжительность исследования должна быть небольшой по сравнению с продолжительностью периода потребления наркотиков лицами, для которых потребление наркотиков носит проблемный характер. Например, обычно продолжительность исследования составляет один год, в этом случае данное допущение не слишком нарушается. Для сравнения можно взять исследование, в рамках которого используются данные, собранные за 5 лет или более, в частности значительное число лиц, для которых потребление наркотиков носит проблемный характер и которые не потребляли наркотики в течение всего периода исследования либо по той причине, что они перестали употреблять наркотики; они умерли или выехали из данного района; или они лишь недавно стали употреблять наркотики; или недавно приехали в данный район. Кроме того, это допущение предполагает, что при использовании метода двойного охвата в период особой распространенности наркомании, т.е. быстрого роста числа лиц, которые только начинают употреблять наркотики путем инъекций или для которых употребление наркотиков только приобретает проблемный характер, для исследования необходимо брать как можно более короткий период времени либо откладывать проведение исследования до тех пор, пока рост не стабилизируется.

Возможно использование открытых моделей двойного охвата, которые позволяют оценить динамику изменения популяции, например миграцию, или проследить случаи смерти и рождения. Такие модели обычно применяются в области экологии животных. Однако для них требуются другие уравнения и статистические знания. Нам известно только одно исследование, в котором эти модели использовались для оценки наркомании (см. тематическое исследование 8).

**Без ошибок в классификации — индивидуумы правильно сопоставляются по всем источникам данных**

В области социальных исследований это означает, что лица, включенные более чем в один список, выявляются правильно. Необходимо сделать все возможное, чтобы обеспечить сбор точных и надежных данных, и следует надеяться, что какие-либо вымышленные имена, даваемые обследуемыми лицами, для которых употребление наркотиков носит проблемный характер, будут выявлены или будут прослеживаться по всем источникам данных. Вполне возможно, что требования конфиденциальности не позволяют собирать и использовать имена и фамилии лиц полностью. Вместо этого необходимо собирать сведения, обеспечивающие совпадение данных по всем источникам без раскрытия личности человека.

В данных примерах Мастро и другие использовали полные имена для выявления случаев совпадения; Фришер и другие, как и во многих других исследованиях проблемного потребления наркотиков методом повторного охвата для выявления случаев совпадения использовали сведения, касавшиеся даты рождения, пола и инициалов (первая буква имени и первая буква фамилии). В примере с Бангладеш исследование, проводившееся без указания источников данных, строилось на сообщениях уличных проституток.

В исследовании на острове Джерси классификационная ошибка привела к завышенной оценке масштабов злоупотребления героином, и это было неизбежно. Поскольку ряд лиц, опрошенных в ходе общинного обследования, назвали не свои имена, не соответствующие именам, указанным в других источниках данных, то в результате некоторые реальные случаи совпадения не были выявлены.

**Исключение неоднородности — все лица, для которых потребление наркотиков носит проблемный характер, имеют равный шанс быть включенными в источник данных**

В социальных исследованиях неоднородность неизбежна. Некоторые лица, для которых потребление наркотиков носит проблемный характер, по причине своего пола, возраста, этнической принадлежности или некоторых других особенностей чаще, чем другие лица, включаются в тот или иной источник данных. Например, в некоторых странах темнокожая молодежь, для которой потребление наркотиков носит проблемный характер, оказывается под арестом чаще, чем все остальные лица, для которых потребление наркотиков также носит проблемный характер. Поэтому с неоднородностью приходится мириться. В случае достаточно большого числа обследуемых лиц и достаточно подробной информации о них для оценки их популяции ее можно стратифицировать. Можно

произвести оценку в разбивке по мужчинам и женщинам, возрастным группам и этнической принадлежности. Если будет обеспечена статистическая поддержка, то можно задействовать более сложные модели (см. раздел «Использование современных методов моделирования», ниже, посвященный развитию базовых методов).

**Репрезентативность — источники данных являются репрезентативными с точки зрения обследуемой популяции лиц, для которых потребление наркотиков носит проблемный характер**

Между исследованиями проблемного потребления наркотиков по методу двойного охвата и рядом других эпидемиологических исследований, например диабета, в ходе которых в качестве источников данных используются списки диагностированных лиц, существует определенное различие. Если допустить, что классификационные ошибки исключаются, то такое исследование будет направлено на то, чтобы определить число неучтенных диагнозов, относящихся к каждому из конкретных источников данных в рамках исследования. Поэтому это будет исследование «неустановленных» данных.

Можно ли с помощью источников данных, которые предполагается использовать, должным образом определить целевую популяцию лиц, для которых употребление наркотиков носит проблемный характер? Определение такой популяции, естественно, отличается от определения, например, лиц, больных диабетом. Во-первых, она определяется рядом клинических, социальных и уголовных аспектов, а это значит, что необходимо обеспечить как можно более широкий охват, в идеальном случае — сбор данных из источников системы здравоохранения и уголовного правосудия. Во-вторых, никакие объединенные списки, даже полные, в том, что касается их собственных целевых групп, не помогут определить в рамках исследуемой популяции общего числа лиц, для которых потребление наркотиков носит проблемный характер. Именно поэтому оценка распространенности является важной, интересной и более трудной, чем оценка многих других проблем в системе общественного здравоохранения.

**Отсутствие зависимости в исследованиях с использованием двух источников данных — оба источника данных независимы**

Провести исследование по двум выборкам легче, однако это допущение является одновременно и основным сдерживающим фактором. Как правило, не известно, зависят ли оба источника данных один от другого, а проверить это нельзя. Если мы все-таки

знаем направление и приблизительный размер смещения, то с помощью исследований по двум выборкам можно достаточно точно определить минимальные или максимальные размеры популяции. Это происходит в связи с тем, что «отрицательная зависимость» — если какое-либо лицо включено в источник данных А, то вероятность того, что оно будет также указано в источнике данных В, будет меньше — означает завышенную оценку фактического размера совокупности. Положительная зависимость — когда какое-либо лицо, включенное в источник данных А, весьма вероятно, также окажется включенным в источник данных В, — ведет к недооценке фактической распространенности.

Для исследований по трем выборкам такое допущение уже неактуально, однако в этом случае возрастает сложность подсчетов. Для анализа данных используются логарифмически-линейные модели, а это в большинстве случаев означает, что нам необходимы пакет статистических программ и, возможно, статистическая поддержка. Двойной охват на основе множественных источников данных предполагает, что взаимодействия между этими источниками не существует, т.е. если используются три источника данных,

то допускается, что трехсторонней зависимости между всеми источниками не имеется. Если, однако, подтвердится наличие связи между каждой парой источников, то такое допущение об отсутствии трехсторонней связи может оказаться недостоверным.

По мере роста числа используемых источников данных возрастает даже еще более быстрыми темпами число возможных моделей, подлежащих проверке. Так, по трем источникам данных возможна подгонка семи моделей, а по пяти — в принципе 31 модели. Есть одна модель, подогнать которую невозможно, — это так называемая модель «полного насыщения», когда все источники данных взаимодействуют между собой. Поэтому по трем источникам данных невозможно трехстороннее взаимодействие, а по четырем источникам — четырехстороннее взаимодействие.

Для оценки по методу двойного охвата с использованием множественных источников данных рекомендуется, если это практически возможно, получить, по крайней мере, один источник данных, предположительно не зависимый от всех остальных источников.

*Продолжение публикации  
читайте в следующем номере*