

## ГЛАВА 4

# РАЗРАБОТКА И КОМПЛЕКТАЦИЯ ВЫБОРКИ

*Настоящая глава, посвященная техническим вопросам<sup>1</sup>, предназначена главным образом для специалистов по построению выборок, а также координаторов обследования и других технических специалистов. Она поможет вам:*

- определить размер выборки;
- оценить возможность использования имеющейся выборки или принять решение в отношении соответствующей основы выборки для новой выборки;
- принять решение в отношении структуры новой выборки;
- ознакомиться с субнациональными показателями и оценками в сфере водоснабжения и санитарно-гигиенических условий;
- получить сведения об ошибках взвешивания, оценки и выборки;
- ознакомиться с методом составления выборки ВПР и скрытой стратификацией;
- получить сведения о структуре выборок в трех странах при проведении МИКС в 2000 году.

Проведение МИКСЗ в вашей стране будет осуществляться на выборочной основе, а не на основе сбора данных по всей целевой группе населения. В обследовании имеются различные целевые группы – домашние хозяйства, женщины в возрасте от 15 до 49 лет, а также дети в возрасте до пяти лет и дети других возрастных групп. Однако в качестве *респондентов* будут, как правило, выступать матери или воспитатели в каждом опрашиваемом домашнем хозяйстве<sup>2</sup>. При этом следует учитывать, что МИКСЗ является обследованием общенационального уровня и выборка будет создаваться на основе всех домашних хозяйств страны, а не только тех, где проживают малолетние дети или женщины детородного возраста.

---

<sup>1</sup> Пользователи Руководства по проведению предыдущего раунда кластерных обследований по многим показателям заметят, что настоящая глава несколько пересмотрена, с тем чтобы, в частности, отразить различные изменения, в том числе формулы для расчета размера выборки, больше внимания уделено разработке и обновлению основы и подсчету погрешности выборки, а также добавлены страновые примеры из раунда МИКС 2000 года.

<sup>2</sup> В *обследовании* по многим показателям домашнее хозяйство определяется как группа людей, которые совместно проживают и питаются. Каждый обладающий необходимыми сведениями взрослый (определяемый для целей МИКСЗ как лицо не моложе 15 лет) отвечает критериям, по которым он может быть основным респондентом по вопроснику домохозяйства. Вместе с тем во многих случаях в качестве респондента будет выступать мать или основной воспитатель ребенка, так как эти лица с наибольшей вероятностью будут находиться дома во время проведения опроса.

## ОПРЕДЕЛЕНИЕ РАЗМЕРА ВЫБОРКИ

Размер выборки является, возможно, наиболее важным параметром для построения выборки, поскольку от него больше, чем от какого-либо иного фактора, зависит точность, затраты и продолжительность обследования. Размер выборки должен определяться с учетом как имеющихся для проведения обследования средств, так и требований в отношении точности его результатов. Последний параметр в свою очередь должен рассматриваться в контексте необходимости подготовки как национальных, так и субнациональных оценок. Кроме того, общий размер выборки не может рассматриваться без учета количества районов выборки – первичных выборочных единиц (ПВЕ) – и размера непосредственных кластеров. Таким образом, несмотря на существование математических формул для расчета размеров выборки, при принятии окончательного решения необходимо учитывать все эти факторы.

### Справочная информация

Настоящая глава Руководства, несмотря на достаточно подробное изложение материала, не имеет целью превратить читателей в специалистов в области статистических выборок. Многие аспекты построения выборки, вероятно, потребуют помощи специалистов, либо государственного национального статистического управления, либо внешних организаций. К числу этих аспектов могут относиться расчет размеров выборки, разработка основы (основ), оценка различных вариантов построения выборки, использование схемы ВПР статистической выборки, определение весовых коэффициентов и подготовка оценок ошибок выборки. В любом случае настоятельно рекомендуется консультироваться с национальным статистическим управлением по вопросам построения выборки.

Два общих эмпирических правила определяют выбор количества ПВЕ и размеры кластеров: чем большее количество ПВЕ вы отберете, тем лучше по критерию географической репрезентативности или разброса, а также по критерию общей достоверности; чем меньше размер кластера выборки, тем более достоверными будут оценки.

#### ПРИМЕР:

При проведении общенационального обследования 600 ПВЕ с размером кластеров в 10 домашних хозяйств каждый дадут более достоверные результаты обследования, чем использование 400 ПВЕ с размерами кластеров в 15 домашних хозяйств каждый, хотя в обоих случаях будет охвачен тот же общий размер выборки – 6000 домашних хозяйств. Кроме того, размер кластера, равный 10, предпочтительнее, чем 15, поскольку меньшие размеры кластера выборки обеспечивают повышение достоверности результатов обследования. Таким образом, в итоге следует стремиться к большему числу ПВЕ при меньшем размере кластеров при прочих равных.

При том, что в целом, чем больше ПВЕ, тем лучше, количество ПВЕ в вашем обследовании будет зависеть в значительной степени от стоимостных параметров и от того, требуется ли проведение оценок на субнациональном уровне (анализ оценок на субнациональном уровне

см. ниже в настоящей главе). Важнейшим фактором являются транспортные расходы. Если расстояние между используемыми в выборке ПВЕ достаточно велико, и проводящие опрос команды переезжают с места на место (а не используются для проведения опроса лица, проживающие в каждой ПВЕ), то сокращение количества отобранных ПВЕ приведет к существенному сокращению общих затрат на проведение обследования. В отличие от этого, если требования обследования предполагают подготовку субнациональных оценок, то это будет серьезным аргументом в пользу отбора большего количества ПВЕ.

Еще одним параметром, который необходимо учитывать при определении размера выборки, является размер кластера, используемого в рамках проведения вашего обследования. Его влияние может быть оценено с использованием так называемого эффекта построения выборки, или ЭПВ. ЭПВ является показателем, отражающим соотношения выборочной дисперсии фактической стратифицированной выборки с использованием кластерного обследования (в данном случае МИКСЗ) и простой случайной выборки<sup>3</sup> при таком же общем размере выборки. Например, если подсчитанная величина ЭПВ индикаторного обследования равна 2,0, это показывает, что осуществленные по результатам этого обследования оценки имеют вдвое более высокий уровень выборочной дисперсии, чем простая случайная выборка такого же масштаба.

Некоторые конкретные примеры выбора количества ПВЕ и определения размера кластера приводятся в конце основного раздела по размерам выборки.

Объем затрат на составление простой случайной выборки не позволяет считать ее целесообразным вариантом для МИКСЗ и для обследований домашних хозяйств в целом, в связи с чем используется метод *кластерной* выборки. К числу факторов, которые оказывают влияние на эффект планирования выборки, относятся стратификация, размер кластеров и их однородность, то есть вероятность того, что два индивидуума (или домашних хозяйства) в кластере имеют одинаковые характеристики. Например, более высокая вероятность того, что двое детей, проживающих поблизости друг от друга, прошли одну и ту же вакцинацию, чем два ребенка, проживающих в случайно выбранных населенных пунктах популяции, является примером однородности кластера.

Стратификация обычно понижает уровень выборочной дисперсии, в то время как степень однородности и размер кластера повышают этот показатель. В связи с этим при построении выборки задача заключается в определении размеров кластера таким образом, чтобы сбалансировать однородность, для целей которой предпочтительнее меньший размер кластеров, и объем затрат, для минимизации которого предпочтительным обычно является более крупный размер кластеров.

Для расчета размера выборки при проведении обследования необходимо учитывать ЭПВ в рамках формулы расчета. Вместе с тем здесь существуют две проблемы. Во-первых, хотя величину ЭПВ легко рассчитать *после проведения обследования*, зачастую она неизвестна до его проведения, если только предыдущие обследования не проводились на базе тех же переменных.

---

<sup>3</sup> Вид вероятностной выборки, в которой  $n$  выборочных единиц отбираются с равной вероятностью из популяции в  $N$  единиц обычно без их замены и с использованием таблицы случайных чисел.

Во-вторых, величина ЭПВ различна по каждому из показателей и, практически, по каждой целевой группе, поскольку однородность кластера различна в зависимости от характеристики. Безусловно нерационально проводить обследование с использованием *различных* размеров выборки для каждой из характеристик на основании их переменных ЭПВ, даже если бы их значения были нам известны.

ЭПВ обычно неизвестны в отношении показателей до проведения обследования, однако предполагается, что они будут весьма незначительными в отношении многих показателей, то есть тех, которые основываются на *редких подклассах* (например, в отношении детей в возрасте от 12 до 23 месяцев)<sup>4</sup>. Если ранее проводилось обследование домашних хозяйств, в ходе которого собирались данные, аналогичные предполагаемым к получению в рамках МИКС, и при его проведении использовалась весьма похожая структура выборки, вы сможете использовать ЭПВ этого предыдущего обследования для оценки возможного влияния структуры на результаты МИКС3. Подсчет ЭПВ осуществляется в рамках немногих обследований домашних хозяйств, но однако программа обследований в области народонаселения и здравоохранения (ОНЗ) является надежным источником такого рода информации.

При расчете формулы и составлении таблицы в отношении размера выборки в нижеследующих подразделах мы исходили из предположения, что эффект планирования выборки равен 1,5 (что может быть несколько завышенным показателем и в связи с этим отражает консервативный подход). Останавливая свой выбор на ЭПВ с завышением погрешности, мы хотим добиться того, чтобы размер выборки был достаточно большим для определения величины всех основных показателей. Тем не менее эмпирическое правило при выборе размера кластера и, косвенным образом, количества кластеров заключается в том, что *размер кластера должен быть настолько мал, насколько можно эффективно обеспечить при проведении обследования на местах* с учетом таких факторов, как количество ПВЕ и затраты на проведение обследования на местах (которые были рассмотрены выше), а также необходимость обеспечения соответствующей нагрузки проводящих опрос сотрудников.

## РАСЧЕТ РАЗМЕРА ВЫБОРКИ

Для расчета размера выборки с использованием соответствующей математической формулы необходимо точно знать значения нескольких факторов, а в отношении значений других факторов необходимо сделать допущения или взять их из материалов предыдущих или аналогичных обследований. К числу этих факторов относятся нижеследующие:

- требуемая точность или *относительная* ошибка выборки;
- желательный уровень достоверности;
- оценочная (или известная) доля населения, относящаяся к выбранной целевой группе, в общей численности населения;

---

<sup>4</sup> Математическим выражением ЭПВ является функция произведения однородности кластера и его размера. Даже если размер кластера велик в отношении общего числа входящих в него домашних хозяйств, он будет незначительным в отношении конкретной целевой группы населения (дети в возрасте одного года), в результате чего ЭПВ будет, скорее всего, также мал.

- прогнозируемый или ожидаемый уровень полноты охвата, или уровень распространенности указанного показателя;
- ЭПВ выборки;
- средний размер домашнего хозяйства;
- корректировка для учета возможного уменьшения количества представленных в выборке домашних хозяйств в связи с неполучением от них ответов.

Расчет размера выборки осложняется тем обстоятельством, что некоторые из этих факторов различны у разных показателей. Мы уже упоминали, что значения ЭПВ могут быть различны. Даже допустимый предел погрешности вряд ли будет одинаковым для различных показателей (и на практике не может быть таковым). Это означает, что для различных показателей потребуются разные размеры выборки для обеспечения требуемой точности. Очевидно, что в рамках проведения обследования, мы должны остановиться на *одном* размере выборки.

Расчет размера выборки используется лишь для переменных, действующих в отношении индивидуумов, несмотря на то что он выражается в количестве домашних хозяйств, которые необходимо посетить для проведения опроса индивидуумов. Это связано с тем, что большинство важнейших показателей, используемых при проведении оценок в рамках МИКСЗ, относятся к физическим лицам. Переменные, связанные с домашними хозяйствами, не должны использоваться при подсчетах размера выборки, в связи с тем что для них необходимы другая формула, а также совсем другие значения эффекта планирования выборки (ЭПВ) – от 10 и выше.

Формула для расчета выглядит следующим образом:

$$n = \frac{[ 4 (r) (1-r) (f) (1,1) ]}{[ (0,12r)^2 (p) (n_h) ]},$$

где

- **n** – требуемый размер выборки, выраженный в числе домашних хозяйств, для получения КЛЮЧЕВОГО показателя (см. в следующем подразделе определение ключевого показателя);
- **4** – множитель, позволяющий обеспечить доверительный уровень, равный 95 процентам;
- **r** – прогнозируемый или ожидаемый уровень распространенности (уровень охвата) для оцениваемого показателя;
- **1,1** – множитель, необходимый для увеличения объема выборки на 10 процентов с учетом неполучения ответов;
- **f** – сокращенное обозначение ЭПВ;
- **0,12r** – допустимый предел погрешности при 95-процентном доверительном интервале, определяемом как 12 процентов от r (таким образом, 12 процентов представляют собой относительную ошибку выборки для показателя r);
- **p** – доля общей численности населения, на которой базируется показатель r;
- **n<sub>h</sub>** – средний размер домашнего хозяйства.

Если размер выборки для проведения обследования рассчитывается с использованием того или иного ключевого показателя, который основывается на наименьшей по численности целевой группе населения в отношении ее доли в общей численности населения, то в этом случае точность оценок большинства других основных показателей, полученных в результате обследования, будет более высокой.

Внимательные пользователи Руководства по проведению МИКС2 обратят внимание на то, что эта формула отличается тем, что показатель *относительной ошибки выборки* ( $0,12 r$ ) заменил показатель *пределов ошибок* (в предыдущем издании это был показатель  $\epsilon$  с величиной 0,03 или 0,05 в отношении, соответственно, показателей с низким и высоким уровнем охвата). В Руководстве по проведению МИКС2 достоверная оценка результатов проведения обследования определялась иначе, в зависимости от того, отражала ли она высокий или низкий уровень охвата. Для оценки величины показателей рекомендовалось, чтобы предел ошибки, или точность, были установлены на уровне 5 процентных пунктов в отношении уровней охвата (например, иммунизации), которые являются относительно высокими, то есть более 25 процентов, и в размере 3 процентных пунктов в отношении уровней охвата, которые являются более низкими, то есть 25 процентов или ниже. При том, что использование двух определенных таким образом пределов ошибки было обоснованно, пользователи, тем не менее, сталкивались с достаточно сложным выбором в отношении того, какой из них следует использовать в рамках своего обследования, особенно в тех случаях, когда рассчитанные размеры выборок разительно отличались друг от друга. Благодаря использованию относительной ошибки выборки<sup>5</sup> этой проблемы удастся полностью избежать, поскольку это сводит предел ошибки к величине сопоставимой точности, вне зависимости от того выбран ли в качестве ключевого показателя для определения размера выборки показатель высокого или низкого уровня охвата. Следует, тем не менее, отметить, что размер выборки больше в отношении показателей, характеризующихся низким уровнем охвата, в связи с чем важно тщательно продумать, какой показатель является *ключевым* для обследования (см. следующий подраздел).

### **ОПРЕДЕЛЕНИЕ И ВЫБОР КЛЮЧЕВОГО ПОКАЗАТЕЛЯ ПРИ РАСЧЕТЕ РАЗМЕРА ВЫБОРКИ**

Рекомендуемая стратегия при расчете размера выборки заключается в выборе одного из важных показателей, который обеспечит наибольший размер выборки. Для этого надо сначала выбрать целевую группу населения, которая представляет собой небольшую долю всего населения ( $r$  в приведенной выше формуле). Эта группа обычно представляет собой *целевую группу населения одного года рождения*<sup>6</sup>. В рамках МИКС3 этой группой является группа детей в возрасте от 12 до 23 месяцев, на которую во многих странах, проводящих МИКС3, приходится около 2,5 процента общей численности населения. *Мы рекомендуем использовать показатель, равный 2,5 процента, если вы не располагаете более точными оценками для своей страны.* Например, если ваш

---

<sup>5</sup> В отношении статистики относительная ошибка выборки известна как коэффициент вариации и определяется как частное от деления стандартной ошибки оценок по результатам обследования на величину самой оценки.

<sup>6</sup> При выборе групп, представляющих собой наименьшую долю от общей численности населения, настоятельно рекомендуется исключить из анализа группы детей в возрасте четырех месяцев, которые формируют основу для определения показателей грудного вскармливания, поскольку необходимые размеры выборки будут нереально большими.

показатель выше (3,5; 4 или 5 процентов), то размеры вашей выборки будут намного меньше, чем в таблице 4.3, в связи с чем чрезвычайно важно использовать наиболее точную имеющуюся оценку величины  $p$  для этой целевой группы населения. Во-вторых, в отношении этой же целевой группы населения должен быть выбран конкретный показатель. Мы назовем его *ключевым показателем* (однако лишь для целей расчета размера выборки).

**Таблица 4.1**  
**Коэффициенты охвата показателя, уровень распространенности или доля**

<p>Низкий уровень охвата является <i>нежелательным</i></p> <ul style="list-style-type: none"><li>• Использование усовершенствованных источников водоснабжения или санитарно-гигиенических устройств</li><li>• Посещаемость учебных заведений</li><li>• Дородовой уход и родоразрешение в стационарных условиях</li><li>• Доля грудного вскармливания</li><li>• Показатели охвата иммунизацией</li></ul> <p>Низкий уровень охвата является <i>желательным</i></p> <ul style="list-style-type: none"><li>• Уровень смертности</li><li>• Распространенность недостаточного веса, отставание в росте и истощение</li><li>• Детский труд</li></ul>
---

При принятии решения в отношении выбора ключевого показателя вам придется выбрать показатель, характеризующийся низким уровнем охвата. Вместе с тем некоторые из показателей, характеризующихся низким уровнем охвата, должны быть исключены из рассмотрения. Это можно понять при рассмотрении показателей, содержащихся в таблице 4.1, где даны примеры показателей, в отношении которых низкий уровень охвата является *нежелательным* и соответствующая цель состоит в *повышении* величины показателя (например, уровня охвата иммунизацией КДС). Вторая группа показателей в таблице 4.1 относится к противоположной модели – низкий уровень охвата является *желательным* и цель состоит в том, чтобы далее понизить его (примером является распространенность задержки роста). *Нет смысла основывать размер выборки на показателях, в отношении которых желателен низкий уровень охвата и охват уже является весьма низким; такого рода показатели должны быть исключены из рассмотрения при выборе ключевого показателя.*

В таблице 4.2 содержатся варианты выбора целевой группы и ключевого показателя в целях прямого расчета размера выборки или выбора размера статистической выборки в таблице 4.3. Следует обратить внимание на то, что в качестве возможного ключевого показателя не рекомендуется использовать коэффициент младенческой смертности или коэффициент материнской смертности<sup>7</sup>. Это связано с тем, что размеры статистических выборок, которые

<sup>7</sup> В отношении размера статистической выборки для коэффициента материнской смертности в руководстве ВОЗ и ЮНИСЕФ от 1997 года "The Sisterhood Method for Estimating Maternal Mortality" рекомендуется, если КМС составляет 300 (на 100 тыс. живорождений), определять его на основании выборки объемом приблизительно в 4 тыс. респондентов с пределом ошибки, равным приблизительно 60, с использованием косвенного метода "сестер".

необходимы для измерения этих показателей, являются слишком большими и исчисляются десятками тысяч респондентов, в связи с чем неразумно их рассматривать. Это не должно означать, что такие показатели не должны измеряться в рамках обследования, – это лишь предполагает, что размер выборки обследования не должен основываться на этих показателях. Результаты обследования в отношении этих показателей будут иметь больший размер ошибок выборки и, соответственно, более широкий доверительный интервал по сравнению с другими показателями.

**Таблица 4.2**

**Контрольный перечень в отношении выбора целевой группы и показателя**

Для выбора соответствующей целевой группы и показателя, который вам необходим для определения размера выборки, следует:

1. Выбрать две или три целевые группы населения, на которые приходится незначительная доля населения в целом. Обычно эти целевые группы не должны быть более узкими по сравнению с возрастной группой одного года рождения или более широкими по сравнению с возрастной группой с пятилетним диапазоном года рождения. В рамках МИКСЗ это, как правило, будут дети в возрасте от 12 до 23 месяцев или дети в возрасте до пяти лет, на которых во многих странах приходится, соответственно, от 2 до 4 процентов и от 10 до 20 процентов всего населения.
2. Проанализировать основные показатели на основании этих групп и отказаться от показателей, которые характеризуются очень низким (менее 5 процентов) или очень высоким (более 50 процентов) уровнем распространенности. Начать расчеты с наименьшей по численности группы – если показатели, базирующиеся на этой группе, характеризуются высоким уровнем охвата, провести расчеты для более широкой возрастной группы, для которой показатели могут иметь более низкий уровень охвата.
3. Выбирать, как правило, показатель, который характеризуется относительно низким уровнем охвата в диапазоне 15–20 процентов в отношении целевой группы, на которую приходится от 10 до 15 процентов населения. Для целевых групп, на которые приходится менее 5 процентов населения, выбрать показатель, характеризующийся несколько более высоким уровнем охвата – более 20, но менее 50 процентов.
4. Не выбирать из показателей, характеризующихся требуемым низким уровнем охвата, тот, который и так является допустимо низким.

При выборе вы также должны учитывать *относительную* важность различных показателей в своей стране. Например, вы не захотите использовать показатель, для оценки которого потребуется очень большой размер выборки, если этот показатель имеет относительно небольшое значение для вашей страны.

**ИСПОЛЬЗОВАНИЕ ТАБЛИЦЫ РАЗМЕРА ВЫБОРКИ**

В таблице 4.3 приведены величины размера выборок, уже рассчитанные с учетом требований МИКСЗ, а также некоторые допущения. Вы можете использовать содержащиеся в таблице величины для определения размера вашей выборки, если они соответствуют вашей ситуации;

в противном случае вы или ваш специалист по подготовке выборки можете самостоятельно рассчитать размер выборки, используя приводимую ниже формулу.

Если содержащиеся в таблице 4.3 параметры соответствуют ситуации в вашей стране, вы можете определить размер выборки без расчетов с использованием приводимой выше формулы. В таблице 4.3 доверительный интервал в отношении точности оценок заранее задан на уровне 95 процентов. В таблице приводятся различные величины среднего размера домашнего хозяйства и уровня охвата –  $r$ , соответственно, от 4,0 до 6,0 и от 0,25 до 0,40. Величина ЭПВ принимается равной 1,5, а уровень точности (предел погрешности) установлен в размере 12 процентов от  $r$ , то есть относительная погрешность выборки равна 12 процентам или  $r$ . В таблице отражена 10-процентная корректировка размера выборки в большую сторону с учетом возможности неполучения ответов от респондентов в ходе обследования.

*Необходимо отметить, что в таблице также делается допущение, согласно которому целевая группа населения для вашего ключевого показателя составляет 2,5 процента от всего населения. Если этот показатель имеет другое значение, вы не можете использовать таблицу для определения необходимого размера выборки. В целом, таблицу нельзя использовать, если любая из предполагаемых величин параметров в формуле не соответствует вашей ситуации; более подробно о том, что делать в этом случае, говорится ниже в настоящем разделе.*

**Таблица 4.3**

**Размер выборки (домашние хозяйства) для определения уровней охвата в отношении наименьшей по численности целевой группы населения (при относительной ошибке выборки в размере 12 процентов показателя уровня охвата при 95-процентном доверительном интервале)**

Средний размер домашнего хозяйства (число человек)	Уровни охвата ( $r$ )			
	$r = 0,25$	$R = 0,30$	$R = 0,35$	$r = 0,40$
4,0	13 750	10 694	8 512	6 875
4,5	12 222	9 506	7 566	6 111
5,0	11 000	8 556	6 810	5 500
5,5	10 000	7 778	<b>6 191</b>	5 000
6,0	9 167	7 130	5 675	4 583

Используйте эту таблицу в том случае, если:

- Целевая группа населения составляет 2,5 процента от численности населения в целом, обычно это дети в возрасте от 12 до 23 месяцев.
- Предполагается, что эффект планирования выборки (ЭПВ) составляет 1,5, а процент неполучения ответов от респондентов ожидается на уровне 10 процентов.
- Величина относительной ошибки выборки установлена в размере 12 процентов от предполагаемого уровня охвата,  $r$

Если все допущения в отношении величины параметров в формуле соответствуют условиям вашей страны, один из размеров статистической выборки, приведенных в таблице 4.3, должен быть применен к вашей ситуации. В некоторых случаях параметры могут быть применимы, однако выбранный вами показатель уровня охвата следует интерполировать.

Например, если уровень охвата в вашем случае составляет от 30 до 35 процентов, вы можете определить размер выборки, интерполируя числа между третьим и четвертым столбцами таблицы. Для иллюстрации: в нижней строке при уровне охвата в 32,5 процента ваш размер выборки будет занимать среднее положение между 7130 и 5675, то есть он составит около 6403 домашних хозяйств.

Поэтапная иллюстрация порядка использования таблицы 4.3 может быть представлена в следующем виде:

- Во-первых, удостоверьтесь, что все значения содержащихся в таблице 4.3 параметров актуальны для ваших условий.
- Затем выберите из таблицы 4.2 показатель, характеризующийся наименьшим уровнем охвата, исключив при этом те показатели, которые и так являются приемлемо малыми. Предположим, что это показатель уровня иммунизации от кори в размере 35 процентов.
- После этого найдите в таблице 4.3 величину среднего размера домашнего хозяйства, которая ближе всего к величине этого показателя в вашей стране (при условии что эта величина находится в указанных диапазонах). Предположим, что она равна 5,5 человек.
- Наконец, найдите в таблице 4.3 величину, которая соответствует величине среднего размера домашнего хозяйства, равной 5,5 человек, и уровню охвата в 35 процентов. Эта величина равна 6191.

Однако приведенные цифры должны использоваться не для точного, а только для приблизительного подсчета размера выборки; следует помнить, что при расчете размеров выборок использовалось несколько допущений. Целесообразно округлять размеры статистических выборок в большую или меньшую сторону в зависимости от объема бюджета. В данном примере вы можете принять решение о том, 6100 или 6200 домашних хозяйств следует использовать с учетом расходов на переезды между ПВЕ, а также размеров кластеров и нагрузки интервьюеров.

### **ИСПОЛЬЗОВАНИЕ ФОРМУЛЫ РАСЧЕТА РАЗМЕРА ВЫБОРКИ<sup>8</sup>**

Что происходит с расчетами размера выборки, если все допущения в отношении величины параметров являются действительными, за исключением того что доля детей в возрасте от 12 до 23 месяцев в общей численности населения в вашей стране не составляет 2,5 процента, а ближе к 2 процентам? В этом случае для получения размера выборки вы можете просто умножить все числа в таблице 4.3 на 2,5/2 или на 1,25. Это является, безусловно, важным, поскольку размеры становятся существенно больше, увеличиваясь на 25 процентов.

Однако возможны случаи, когда лучше отказаться от использования таблицы 4.3 и самостоятельно рассчитать размер выборки, используя приведенную ниже формулу. Формула должна использоваться в том случае, если любые из величин параметров в вашей стране отличаются от допущений, содержащихся в таблице 4.3. В таблице 4.4 перечислены условия, при наличии которых должна применяться эта формула.

---

<sup>8</sup> Шаблон Excel для расчета размера выборки можно найти по адресу: [www.childinfo.org](http://www.childinfo.org).

Мы уже рассмотрели ситуацию увеличения размера выборки при значении  $p$  меньше 0,025, и если вы не можете использовать приведенный выше пример, то для расчета размера выборки должна использоваться формула. Повторяем еще раз, она должна использоваться, если любое из значений других параметров соответствует величинам, содержащимся в таблице 4.4.

**Таблица 4.4**  
**Контрольный перечень для использования формулы расчета размера статистической выборки**

Формула для расчета размера вашей выборки выглядит следующим образом:

$$n = \frac{[ 4 (r) (1-r) (f) (1,1) ]}{[ (0,12r)^2 (p) (n_h) ]}$$

Используйте ее, если какие-либо (одно или более) из приведенных ниже условий соответствуют ситуации в вашей стране:

- доля детей в возрасте одного года в общей численности населения ( $p$ ) не составляет 0,025;
- средний размер домашнего хозяйства ( $n_h$ ) меньше 4,0 человек или больше 6,0 человек;
- уровень охвата вашего ключевого показателя ( $r$ ) меньше 25 процентов;
- эффект планирования выборки ( $f$ ) в отношении вашего ключевого показателя не составляет 1,5 в соответствии с принятыми оценками по результатам других обследований в вашей стране;
- вы предполагаете, что уровень отказа респондентов от ответов составит более или менее 10 процентов.

Не изменяйте уровень достоверности в формуле и сохраните его значение равным 4.

Применение формулы является достаточно простым, поскольку после внесения числовых значений параметров остается лишь совершить простые арифметические действия. Например, при  $r = 0,25$ ,  $f = 1,6$ , коэффициенте корректировки с учетом отказов респондентов от ответов = 1,05,  $p = 0,035$  и  $n_h = 6$ , мы имеем

$$n = \frac{[ 4 (0,25) (1-0,25) (1,6) (1,05) ]}{[ (0,12 * 0,25)^2 (0,035) (6) ]} = \frac{1,26}{0,000189} = 6667$$

*Во многих предыдущих обследованиях МИКС стандартный размер статистической выборки колебался в диапазоне 4000–8000 домашних хозяйств. Этот диапазон является для вас целью, к достижению которой стоит стремиться при осуществлении расчета размера выборки, учитывающего как требования в отношении достоверности, так и ограничения в области финансирования. Как мы уже отмечали, в рамках проведения МИКС3 будут получены оценочные величины многих показателей, каждая из которых будет характеризоваться собственной степенью точности. В связи с этим весьма полезным может оказаться проведение анализа соответствующих уровней достоверности – стандартных ошибок и доверительного интервала – в отношении ваших показателей при конкретном размере выборки.*

Таблица 4.5 содержит примеры таких уровней достоверности для выборки объемом в 6000 домашних хозяйств, которая может рассматриваться в качестве типичного размера выборки для подготовки относительно достоверных оценок в отношении большинства показателей, представляющих интерес в рамках проведения МИКС3.

Таблица 4.5

**Ожидаемые показатели достоверности (стандартная ошибка и доверительный интервал) для выборки объемом 6000 домашних хозяйств по различным демографическим сценариям**

Средний размер домашнего хозяйства	Величина показателя r	Размер подгруппы населения Р	Число отобранных индивидуумов в подгруппе населения	Число лиц, характеризующихся показателем	Стандартная ошибка	Доверительный интервал (95% уровень)	
						Нижний	Верхний
4 человека	0,10	,025	540	54	,016	,068	,132
		,05	1080	108	,011	,078	,122
		,125	2700	270	,007	,086	,114
		,20	4320	432	,006	,089	,111
	0,20	,025	540	108	,021	,158	,242
		,05	1080	216	,015	,170	,230
		,125	2700	540	,009	,181	,219
		,20	4320	864	,007	,185	,215
	0,30	,025	540	162	,024	,252	,348
		,05	1080	324	,017	,266	,334
		,125	2700	810	,011	,278	,322
		,20	4320	1296	,009	,283	,317
	0,50	,025	540	270	,026	,447	,553
		,05	1080	540	,019	,463	,537
		,125	2700	1350	,012	,476	,524
		,20	4320	2160	,009	,481	,519
5 человек	0,10	,025	675	68	,014	,072	,128
		,05	1350	135	,010	,080	,120
		,125	3375	338	,006	,087	,113
		,20	5400	540	,005	,090	,110
	0,20	,025	675	135	,019	,162	,238
		,05	1350	270	,013	,173	,227
		,125	3375	675	,008	,183	,217
		,20	5400	1080	,007	,187	,213
	0,30	,025	675	203	,022	,257	,343
		,05	1350	405	,015	,269	,331
		,125	3375	1013	,010	,281	,319
		,20	5400	1620	,008	,285	,315
	0,50	,025	675	338	,024	,453	,547
		,05	1350	675	,017	,467	,533
		,125	3375	1688	,011	,479	,521
		,20	5400	2700	,008	,483	,517
6 человек	0,10	,025	810	81	,013	,074	,126
		,05	1620	162	,009	,082	,118
		,125	4050	405	,006	,088	,112
		,20	6480	648	,005	,091	,109
	0,20	,025	810	162	,017	,166	,234
		,05	1620	324	,012	,176	,224
		,125	4050	810	,008	,185	,215
		,20	6480	1296	,006	,188	,212
	0,30	,025	810	243	,020	,261	,339
		,05	1620	486	,014	,272	,328
		,125	4050	1215	,009	,282	,318
		,20	6480	1944	,007	,286	,314
	0,50	,025	810	405	,022	,457	,543
		,05	1620	810	,015	,470	,530
		,125	4050	2025	,010	,481	,519
		,20	6480	3240	,008	,485	,515

В таблице 4.5 в столбце 4 приведены данные об ожидаемом числе лиц, которые будут опрошены в рамках выборки, равной 6000 домашних хозяйств, при допущении количества случаев неполучения ответов на вопросы в размере 10 процентов. Например, в стране, где средний размер домашнего хозяйства составляет 4 человека, число выбранных лиц в подгруппе населения, составляющей 2,5 процента от численности всего населения (например, дети в возрасте от 12 до 23 месяцев), будет равно приблизительно 540, а не 600, с учетом случаев неполучения ответов на вопросы в рамках проведения обследования. Для этой группы лиц столбец 5 демонстрирует ожидаемое число лиц в выборке, которые будут характеризоваться следующими показателями:  $g = 54$ , если  $g$  равно 10 процентам; 108, если  $g$  равно 20 процентам; 162 для  $g$ , равного 30 процентам; и 270 – для 50 процентов.

Следует отметить, что ожидаемая стандартная ошибка существенно колеблется в зависимости от размера отобранной подгруппы населения и величины показателя. Важным показателем достоверности при оценке ваших результатов является доверительный интервал, отражаемый в последнем столбце таблицы 4.5. Доверительный интервал, или ДИ, показывает диапазон, в котором можно ожидать колебания значений вашей оценки относительного истинного значения для всего населения с учетом величины стандартной ошибки. Он подсчитывается путем сложения и вычитания два раза стандартной ошибки (в отношении 95-процентного доверительного интервала) из оцениваемой величины показателя. В нижней строке таблицы 4.5 приводится доверительный интервал  $[0,485-0,515]$  для величины показателя, оцениваемой в 0,50. Это означает, что если вы оцениваете уровень охвата показателя в 50 процентов, вы можете быть уверены на 95 процентов, что реальная величина показателя в отношении всего населения составляет от 48,5 до 51,5 процента.

### ПРИНЯТИЕ РЕШЕНИЯ О КОЛИЧЕСТВЕ ПВЕ И РАЗМЕРАХ КЛАСТЕРОВ – ПРИМЕРЫ

В начале раздела, посвященного размеру выборки, мы рассмотрели вопрос о том, насколько количество ПВЕ и размеры кластеров также оказывают влияние на размер выборки, обратив при этом внимание на то, что достоверность выборки повышается при увеличении количества ПВЕ и уменьшении размеров кластера. Мы завершаем настоящий раздел тремя примерами, в которых используются различные сценарии для демонстрации взаимозависимости размера выборки, количества ПВЕ и размера кластера.

#### ПРИМЕР 1

Целевая группа:	дети в возрасте от 12 до 23 месяцев
Доля в общей численности населения:	2,6 процента
Ключевой показатель:	уровень охвата иммунизацией против КДС
Распространенность (охват):	40 процентов
ЭПВ:	информация отсутствует
Средний размер домашнего хозяйства:	6 человек

В рамках этого сценария можно использовать таблицу 4.3, поскольку цифровые значения уровня охвата в отношении ключевого показателя и размера домашних хозяйств представлены в таблице. Целевая группа населения, на которую приходится 2,6 процента общей численности населения, также очень близка по своей численности к цифре в 3 процента, на которой основывается таблица 4.3. При отсутствии информации об эффекте планирования предполагается, что его величина составляет 1,5, а коэффициент

корректировки с учетом возможности неполучения ответов респондентов устанавливается в размере 1,1, что соответствует ожидаемому уровню отказа респондентов от ответов в размере 10 процентов. Размер выборки при среднем размере домашнего хозяйства в 6 человек и при уровне охвата, равном 40 процентам, составляет, таким образом, 4583 домашних хозяйства.

Предположим, что ваша страна имеет относительно большую площадь и, кроме того, делится на множество провинций, например 15. С учетом этого вы и ваш персонал, занимающийся вопросами выборки, приняли решение о том, что в целях обеспечения необходимого географического охвата и достаточной репрезентативности в каждой провинции вам необходимо не менее 300 ПВЕ. Кроме того, вы считаете, что бюджет обследования достаточен для такого количества ПВЕ. В этом случае размер кластера должен быть рассчитан путем деления 4583 на 300, что составит приблизительно 15–16 домашних хозяйств.

Вместо того чтобы стремиться к охвату 300 ПВЕ, вы с персоналом, занимающимся проведением обследования и составлением выборки, можете принять альтернативное решение о том, что вам нужны кластеры определенного размера, например 10, в целях соблюдения таких оперативных требований, как распределение нагрузки среди интервьюеров. В этом случае вы разделите 4583 на 10, что даст вам количество ПВЕ, равное приблизительно 458. После этого вы должны провести анализ этого количества по критерию затрат и по другим факторам и либо принять его, либо скорректировать размер кластеров. Вы можете решить, что с учетом транспортных расходов максимальное количество ПВЕ, которые вы можете обследовать, равно 425, и в этом случае вы должны скорректировать размер кластера до 11 (то есть  $4583/425$ ).

#### **ПРИМЕР 2**

Целевая группа:	дети в возрасте от 12 до 23 месяцев
Доля в общей численности населения:	2,5 процента
Ключевой показатель:	уровень охвата иммунизацией от полиомиелита
Распространенность (охват):	26 процентов
ЭПВ:	информация отсутствует
Средний размер домашнего хозяйства:	6 человек

В рамках этого сценария вы также можете использовать таблицу 4.3, потому что, за исключением уровня охвата в отношении ключевого показателя, все параметры таблицы актуальны, при условии что мы опять делаем допущение, что эффект планирования равен 1,5, а коэффициент корректировки на случай неполучения ответов на вопросы составляет 1,1. В отношении охвата,  $\gamma$ , мы можем использовать столбец по 25 процентам, поскольку оцениваемая величина в 26 процентов является весьма близкой. Приводимый в таблице размер выборки для среднего размера домашнего хозяйства в 6,0 человек составляет 9167 домашних хозяйств.

Предположим, что с учетом стоимостных показателей и уровня производственной нагрузки на местах, проводящая обследование команда принимает решение о том, что ей необходимо, по возможности, размер кластеров, равный 30 домашним хозяйствам. Исходя из этого, путем деления 9167 на 30 мы получаем 306 ПВЕ, и вы можете принять решение о том, что это является приемлемым количеством для проведения обследования на местах. Если, с другой стороны, вы принимаете решение о том, что для обеспечения географического охвата, а также для использования большего количества ПВЕ в целях подготовки субнациональных оценок для пяти регионов вам необходимы 400 кластеров, вы должны разделить 9167 на 400, что даст размер вашего кластера, равный приблизительно 23. Помните о том, что чем меньше размер кластера, тем более достоверными будут оценки величины показателя (для всех показателей, а не только для ключевого показателя).

В связи с этим вы можете принять решение об использовании схемы выборки с 400 ПВЕ и средним размером кластера, равным 23 домашним хозяйствам, помня при этом, что в связи с необходимыми транспортными расходами он будет предполагать больший объем затрат по сравнению с планом, содержащим 306 ПВЕ.

### ПРИМЕР 3

Целевая группа:	дети в возрасте от 0 до 11 месяцев
Доля в общей численности населения:	3,5 процента
Ключевой показатель:	дети с полноценным питанием
Распространенность (охват):	24 процента
ЭПВ	1,4 (на основании данных предыдущего обследования)
Средний размер домашнего хозяйства:	4 человека
Ожидаемый уровень неполучения ответов:	10 процентов

В рамках этого сценария вы должны сами подсчитать размер выборки с использованием формулы, предложенной в настоящем разделе, поскольку некоторые из параметров отличаются от тех, которые использованы в таблице 4.3 или в отношении которых были сделаны определенные допущения. В их число входят значения  $p$ ,  $f$  и коэффициента корректировки в отношении неполучения ответов, последний из которых основывается на ожидаемом уровне неполучения ответов в размере 5 процентов в отличие от 10 процентов, полученных в результате проведения аналогичного обследования в вашей стране. Использование формулы дает результат, соответствующий 10 303 домашним хозяйствам.

Предположим, что проводящий обследование персонал пришел к выводу, что с учетом финансовых соображений в рамках обследования может быть опрошено не более 300 ПВЕ. В этом случае вы принимаете 300 в качестве фиксированной величины и рассчитаете размер кластера путем деления 10 303 на 300, что будет иметь результатом 34 домашних хозяйства в качестве размера кластера. В этом случае вам будет нужно оценить, позволит ли такой размер кластера осуществить достаточно достоверные оценки других показателей, помимо ключевого<sup>9</sup>. Если мы предполагаем, что максимальный размер кластера не должен превышать 30 домашних хозяйств, то количество ПВЕ, необходимых для 10 303 домашних хозяйств, будет равно 343. Таким образом, придется сделать выбор о том, следует ли смириться с более низким уровнем достоверности плана выборки с 300 ПВЕ или с более высокими затратами схемы выборки с 343 ПВЕ.

## ПРИНЯТИЕ РЕШЕНИЯ ОТНОСИТЕЛЬНО ИСПОЛЬЗУЕМОЙ ВЫБОРКИ

После того как вы приняли решение о размере выборки и в предварительном порядке определили количество ПВЕ, следующая задача заключается в принятии решения о том, какую выборку использовать при проведении обследования. Построение, отбор и проведение корректной вероятностной выборки с начала до конца требуют больших затрат времени и финансовых ресурсов (проведение вероятностных выборок рассматривается в следующем разделе). Для целей МИКС3 необходимо провести оценки показателей в относительно короткий срок, и у вас может не оказаться достаточно времени для построения новой выборки для проведения обследования.

<sup>9</sup> Хотя в этом примере величина эффекта планирования является весьма незначительной для основного показателя и в связи с этим достоверность оценки, как предполагается, будет соответствовать установленным требованиям точности, другие показатели, которые имеют гораздо более высокую межкластерную корреляцию по сравнению с показателем в отношении детей младше одного года, будут, как ожидается, характеризоваться намного более высоким уровнем ошибки выборки при размере кластера больше 30 по сравнению с размером кластера в 20 или 25.

Поэтому необходимо предпринять два основных шага для определения того, какую выборку следует использовать при проведении обследования:

**Этап 1:** Определить, можно ли использовать какую-либо существующую выборку.

**Этап 2:** В случае отсутствия подходящей выборки необходимо разработать выборку специально для МИКСЗ.

В настоящем разделе мы рассматриваем первый этап. При наличии подходящей для МИКСЗ выборки вы можете не анализировать возможные варианты построения выборок, предлагаемые для этапа 2, которые рассматриваются в следующем разделе. Тем не менее следующий раздел полезно изучить, чтобы удостовериться, что существующая выборка, которую вы планируете использовать, является корректной вероятностной выборкой с относительно современной основой выборки.

### **ИСПОЛЬЗОВАНИЕ СУЩЕСТВУЮЩЕЙ ВЫБОРКИ – ВАРИАНТ 1**

К счастью, большинство стран располагают хорошо проработанными программами проведения обследований силами национальных статистических служб или министерств здравоохранения. В связи с этим в вашей стране может оказаться возможным использовать уже существующую выборку, построенную для других целей. *Это является рекомендуемым вариантом для проведения вашего обследования, при условии что существующая выборка является достоверной вероятностной выборкой и имеется возможность ее использования.* Необходимо провести оценку существующей выборки, для того чтобы удостовериться, насколько полно она отвечает требованиям проведения вероятностной выборки (как уже отмечалось, этот вопрос рассматривается в следующем разделе Руководства).

Существуют различные способы использования имеющейся выборки, например:

- добавить модули вопросников МИКСЗ к вопросам, которые планируется использовать в рамках проведения другого обследования;
- использовать выборку, или подмножество, какого-либо обследования;
- использовать описи домохозяйств в районах проведения (или кластерах) какого-либо другого обследования;
- использовать районы проведения или кластеры предыдущего обследования с обновленной описью домохозяйств.

У каждого из этих вариантов есть свои преимущества и недостатки. Важнейшее значение имеет также и временной фактор. Например, первый вариант нельзя будет использовать в том случае, если не планируется проведения какого-либо другого исследования в период времени, предусмотренный для проведения МИКСЗ. Это вариант – добавление модулей вопросников к вопросам другого обследования, иногда называемый "комбинированным вариантом", поскольку данные двух обследований собираются одновременно, имеет очевидные преимущества, поскольку выборка уже подготовлена, что сокращает затраты на подготовку выборки в рамках проведения МИКСЗ. Существенным недостатком при этом может быть проблема нагрузки на респондентов, поскольку вопросники МИКСЗ являются достаточно

объемными, а основное обследование может использовать собственный объемный вопросник. Эти проблемы должны быть тщательно проанализированы, и их следует обсудить с организаторами базового обследования и с его руководством.

Второй вариант, то есть использование выборки предшествующего обследования, также имеет своим преимуществом наличие построенной выборки, что опять-таки сокращает расходы на подготовку выборки. Если размеры выборки для предшествующего обследования были слишком большими, то специалист в области построения выборок легко сможет создать подвыборку первоначальной выборки и привести ее размер в соответствие с требованиями МИКСЗ. При слишком маленьком размере выборки его, напротив, будет сложнее увеличить. Существуют также ограничения в части повторного посещения тех домашних хозяйств, которые опрашивались в рамках предыдущего обследования, вновь в связи с возможным возникновением проблем, связанных с нагрузкой на респондентов/или выработкой у них навыка ответов на вопросы. Наконец, для того чтобы имело смысл использовать этот вариант, предыдущее обследование должно было быть проведено достаточно недавно.

Третий вариант, то есть использование описей домохозяйств в районах проведения предыдущего обследования в качестве основы для подготовки выборки для целей МИКСЗ, имеет двойное преимущество, заключающееся в том, что 1) уже проведено предварительное обследование первичных единиц, и 2) описи домохозяйств уже подготовлены. Таким образом, опять-таки уже осуществлена значительная часть действий и произведены соответствующие расходы по подготовке выборки. Преимущество заключается в том, что для целей МИКСЗ будут отобраны домашние хозяйства, не опрашивавшиеся в рамках предыдущего обследования, что позволит избежать проблем нагрузки на респондентов и их усталости, или выработки у них навыков ответов на вопросы. Недостаток заключается в том, что описи домохозяйств устареют, если предыдущее обследование проводилось год или два назад, в каком-либо случае этот вариант является неприемлемым. В том случае, если описи домохозяйств устарели, можно рассмотреть четвертый из предложенных выше вариантов. Этот вариант предполагает составление новых описей домохозяйств в районах выборки до начала составления выборки. Хотя недостатком при этом является необходимость проведения мероприятий и производства соответствующих затрат, связанных с подготовкой новых описей домохозяйств, преимущество заключается в том, что первичные единицы уже отобраны и план выборки в целом уже подготовлен, что снимает необходимость его разработки.

Каждый из этих пунктов должен быть тщательно проанализирован, и следует сделать вывод о целесообразности осуществления необходимых изменений до принятия решения об использовании какой-либо существующей выборки.

Существующая выборка, которая идеально подходит, является выборка для обследования в области народонаселения и здравоохранения (ОНЗ)<sup>10</sup>. Многие страны недавно проводили такого

---

<sup>10</sup> Проблемы построения статистической выборки рассмотрены в: Demographic and Health Surveys: Sampling Manual, Basic Documentation - 8, Macro International Inc., Calverton, Maryland 1987.

рода обследования, а другие государства планируют провести их в ближайшие месяцы<sup>11</sup>. Задачи измерений, проводимых в рамках ОНЗ, весьма похожи на задачи МИКСЗ. В связи с этим структура выборки, используемая в рамках ОНЗ, вероятно, вполне соответствует вашим целям.

**Таблица 4.6**  
**Вариант 1 – существующая выборка**

**Преимущества**

- Экономия времени и средств
- Можно рассчитывать на то, что она построена должным образом с использованием методов вероятностной выборки
- Можно легко осуществить корректировку с учетом требований МИКСЗ

**Недостатки**

- Требуется обновление в том случае, если выборка составлялась достаточно давно
- Респонденты могут быть перегружены
- Вопросник, используемый для получения данных по необходимому показателю, может быть слишком объемным в случае использования "комбинированного метода"
- Корректировка с учетом требований МИКСЗ может оказаться достаточно сложной

При каких обстоятельствах следует использовать выборку ОНЗ? Вы должны оценить ее доступность, а также ее актуальность и пригодность с учетом ваших потребностей. *Для проведения обследования МИКСЗ на местах может использоваться недавняя, но подготовленная до 2003 года выборка для ОНЗ, или предстоящее в ближайшее время ОНЗ, в рамках которого вопросник МИКСЗ может использоваться в качестве дополнения.* ОНЗ, несомненно, будет построено как вероятностная выборка. Таким образом, вам необходимо только провести оценку того, а) является ли размер выборки достаточным для МИКС, и б) находятся ли количество ПВЕ и размер кластеров в диапазонах значений, рассматриваемых в настоящем Руководстве. Наконец, это потребует достижения договоренностей и сотрудничества с организаторами ОНЗ или учреждением-исполнителем в вашей стране с учетом упомянутых выше трудностей, связанных с излишней нагрузкой на респондентов.

Еще одним обследованием, которое проводили многие страны и выборка которого может быть вами использована, является обследование трудовых ресурсов. При том, что цели проведения обследований трудовых ресурсов существенным образом отличаются от задач МИКСЗ, обследования трудовых ресурсов весьма часто имеют структуру, весьма схожую со структурой обследований МИКС в отношении стратификации, размера выборки и других критериев проведения выборки.

<sup>11</sup> Тем не менее следует отметить, что не рекомендуется проводить МИКСЗ в том случае, если ОНЗ проводилось после 2003 года или планируется к проведению в 2005 году или в начале 2006 года.

## ПОДГОТОВКА ОСНОВЫ ВЫБОРКИ ДЛЯ НОВОЙ ВЫБОРКИ

В том случае когда нельзя использовать существующую выборку, необходимо использовать и/или подготовить основу выборки домашних хозяйств, на базе которой можно построить новую выборку для МИКСЗ. Основа должна строиться на началах вероятностной выборки.

### НАДЛЕЖАЩАЯ СХЕМА ВЕРОЯТНОСТНОЙ ВЫБОРКИ И ОСНОВА ВЫБОРКИ

В отношении получения достоверных результатов, в максимально возможной степени свободных от предвзятости, схема соответствующей *вероятностной* выборки для проведения обследования настолько же важна, как и разработка различных модулей вопросников. Существуют несколько способов планирования вероятностной выборки, и каждая страна наверняка характеризуется собственной ситуацией, условиями и потребностями в данных, определяющими *конкретную* схему выборки, которую эта страна принимает. Однако существуют общие особенности, которые должны учитываться всеми странами в целях соблюдения *требований* научно обоснованной вероятностной выборки:

- использование общепринятых методов построения вероятностной выборки на каждом этапе создания выборки;
- подбор выборки, являющейся репрезентативной в национальном масштабе;
- обеспечение того, чтобы проведение опроса на местах соответствовало схеме выборки;
- обеспечение того, чтобы размер выборки являлся достаточным для выполнения требований в отношении достоверности.

В дополнение к этим четырем *требованиям* существуют и другие правила построения схемы выборки, которым вам настоятельно рекомендуется следовать, несмотря на то что каждое из них может быть определенным образом изменено в зависимости от ситуации в вашей стране и соответствующих потребностей. В их число входит:

- использование простых, а не сложных процедур построения выборки;
- использование последней переписи населения в качестве *основы выборки*;
- по возможности *самовзвешенная* выборка.

Использование научно обоснованных методов построения вероятностной выборки для обследований практикуется в большинстве стран мира на протяжении десятилетий. Если выборка из всего представляющего интерес населения не сделана точно, с использованием широко известных методов построения вероятностной выборки, оценки обследования окажутся *смещенными*. При этом масштаб этих смещений будет неизвестным. Чрезвычайно важно обеспечить, чтобы в методологии построения выборки использовались методы вероятностного отбора на каждом этапе отбора.

---

**Во избежание смещения выборки для отбора респондентов необходимо использовать *вероятностное* построение выборки. Смещение выборки определяется не размером выборки, а используемыми *методами* отбора. Увеличение размера выборки не устранил смещения выборки, если методы отбора неверны.**

---

Построение вероятностной выборки является средством обеспечения того, чтобы все представители целевой группы населения<sup>12</sup> имели известную степень вероятности быть включенными в выборку. Кроме того, эта вероятность *не должна равняться нулю* и должна *поддаваться подсчету*. Явным признаком отсутствия вероятностного характера выборки является ситуация, при которой специалист в области статистики не может рассчитать вероятность отбора в рамках используемой схемы выборки.

Примерами методов построения выборки, которые не основываются на методах вероятности, являются *опирающаяся на оценочные суждения* выборка, *направленная* выборка и выборка *на основе квот*. Метод случайного блуждания при отборе детей является методом выборки на основе квот. Важно помнить, что вы *не должны* использовать эти методы при проведении МИКСЗ.

Наилучший способ избежания смещения при построении выборки заключается в строгом следовании методам вероятностного построения выборки. Существуют другие смещения, не связанные с методами построения выборки, такие как отказ от ответа, ошибочный ответ, ошибки интервьюеров, однако эти ошибки в различных масштабах будут присутствовать вне зависимости от используемых методов построения выборки. Необходимо принять соответствующие меры и для предотвращения такого рода *не связанных с построением выборки смещений*, включая такие меры, как предварительные испытания, тщательная подготовка интервьюеров, а также контроль качества работы на местах.

---

**При вероятностных выборках каждый представитель целевой группы населения имеет шанс быть отобранным, вероятность отбора не является нулевой и может быть подсчитана математическим способом, а вероятностные методы используются на каждом этапе осуществления отбора.**

---

Второе требование в отношении схемы выборки МИКСЗ заключается в том, чтобы выборка имела *общенациональный* характер применительно к масштабу и охвату. Это необходимо, чтобы оценки показателя отражали ситуацию в стране в целом. Для обеспечения общенационального охвата важно в максимально возможной степени включить в выборку трудно подсчитываемые группы населения. Такими группами населения могут быть кочевники, бездомные или бродяги, а также обитатели лагерей беженцев, казарм и населенных пунктов в изолированных районах, до которых сложно добраться по тем или иным причинам. Вполне вероятно, что состояние здоровья лиц, в частности детей, проживающих в этих условиях, отличается от показателей здоровья детей, проживающих в более стабильных или традиционных условиях, и исключение этих категорий населения приведет к смещению величин показателей.

Один из важнейших способов обеспечения действительно общенационального масштаба выборки и его соответствия методам вероятностного построения выборки заключается в том, что используемая основа выборки *охватывает* все население страны. Основа выборки подробнее рассматривается ниже.

---

<sup>12</sup> При проведении МИКСЗ существуют различные целевые группы населения в зависимости от соответствующего показателя. Примерами являются дети в возрасте от 0 до 11 месяцев, в возрасте от 12 до 23 месяцев, дети в возрасте до пяти лет, дети в возрасте до пяти лет, страдающие диареей, женщины в возрасте от 15 до 49 лет и население в целом.

Для обеспечения эффективности вероятностного построения выборки необходимо обеспечить, чтобы реализация на местах плана составления выборки, включая процедуры опроса, соответствовала бы этой схеме. Отмечены многочисленные случаи, когда небрежная работа на местах лишала смысла прекрасную во всем остальном схему выборки. Руководители на местах должны обеспечить строгое соблюдение процедур составления выборки.

Важнейшим условием обеспечения достоверности вероятностного составления выборки является точное изложение требований к точности в целях расчета размера выборки. Этот вопрос рассматривался в разделе выше, посвященном определению размера выборки. Мы высказали рекомендации о том, чтобы уровень точности ключевого показателя был установлен на 12 процентах относительной ошибки выборки при 95-процентном доверительном интервале, и это те критерии, на которых базируется формула для подсчета размера выборки, приведенная в упомянутом выше разделе. Если, например, ваш ключевой показатель характеризуется уровнем охвата или распространенности, равным 20 процентам, то в этом случае относительная ошибка, равная 12 процентам, даст предел ошибок в размере 2,4 процентных пункта, а доверительный интервал оценки, полученной в результате вашего обследования, будет равен  $|17,6-22,4|$ .

Ваша выборка должна иметь возможно более простую схему. Хорошо известно, что чем более сложным является план выборки, тем больше вероятность того, что его реализация будет неправильной. Это может создать особенно много проблем в случае проведения обследования на местах при необходимости осуществления сложных процедур построения выборки. Кроме того, задача по своевременному получению результатов обследования может оказаться невыполненной.

План выборки называется *самовзвешенным* в том случае, когда каждый член целевой группы населения отбирается с одинаковой *общей* вероятностью. Общая вероятность является произведением вероятностей на каждом этапе отбора. Самовзвешенная выборка является желательной, поскольку могут быть подготовлены различные оценки, например оценки процентного распределения на основании данных выборки без их взвешивания или расширения. Учитывая стремление к простоте при построении выборки, предпочтительнее располагать самовзвешенной схемой выборки, а не более сложной не самовзвешенной схемой. Тем не менее *требование в отношении самовзвешенности не должно является строгим*, поскольку взвешивание результатов выборки для подготовки оценок может быть легко осуществлено при помощи современных компьютеров. Кроме того, существуют ситуации, когда схема выборки не может быть самовзвешенной.

**ПРИМЕР:**

Предположим, что в вашей стране вам будет необходимо получить отдельные значения того или иного показателя для городских и сельских районов, и, предположим далее, что вы хотите, чтоб эти оценки были *одинаково* достоверными. Это определит необходимость отбора выборки одинакового размера в городских и сельских районах. За исключением случая равной численности городского и сельского населения, частота выборки для каждой группы населения будет различной. В связи с этим общенациональная выборка потребует взвешивания для получения достоверных результатов и, таким образом, выборка для проведения обследования не будет самовзвешенной.

## ПЕРЕПИСЬ НАСЕЛЕНИЯ КАК ОСНОВА ВЫБОРКИ И СЛУЧАИ, КОГДА НЕОБХОДИМО ОБНОВЛЕНИЕ ДАННЫХ

Настоятельно рекомендуется использовать в качестве *основы выборки* материалы последней по времени переписи населения, которые в случае необходимости следует обновить. Почти во всех странах мира за последние десять лет прошли переписи населения. Основа выборки представляет собой набор материалов, из которых осуществляется отбор выборки для проведения обследований. Образцовая основа выборки должна быть полной, точной и актуальной, хотя ни одна из основ выборок не является идеальной на 100 процентов, ближе всего этим критериям в большинстве стран соответствуют переписи населения. *Основная цель использования переписи населения для проведения нашего обследования заключается в получении полного списка счетных участков (СУ) с количественными показателями, такими как численность населения или количество домашних хозяйств, в целях осуществления отбора выборочных единиц первого этапа.* В большинстве стран составными элементами переписей населения обычно являются карты, и в их число могут входить схемы счетных участков. Эти карты являются полезным инструментом, поскольку данные по отобранным счетным участкам, скорее всего, понадобится уточнить в отношении находящихся в них в *настоящее время* домашних хозяйств, особенно в том случае, если перепись населения проводилась год или два назад.

Некоторые страны провели свои переписи населения серии 2000 еще в 1999 году, а многие другие страны провели эти переписи в период 2000–2002 годов. Это поднимает весьма важный вопрос, упоминавшийся в предыдущем абзаце, нужно ли для МИКС обновлять данные по основе выборки. *В целом не рекомендуется проводить обновление, если основа выборки была создана в 2003 году или позднее, но с одним исключением.* В странах, где имело место *значительное* перемещение населения после 2003 года, особенно в высокоурбанизированных районах, которые расширились в конкретных зонах в связи с массовым новым жилищным строительством, необходимо провести уточнение данных в отношении этих зон. Однако вы можете принять решение, что это не нужно делать в том случае, если перепись населения в вашей стране проводилась совсем недавно – за 12 или менее месяцев до вашего обследования.

---

**Если основа выборки в вашей стране была подготовлена до 2003 года, рекомендуется провести ее обновление.**

---

Причина, по которой необходимо провести обновление, по-видимому, очевидна. Следует обеспечить, чтобы охват всего населения был в максимально возможной степени точным и полным. Рекомендуемые шаги по обновлению основы выборки одинаковы для обоих сценариев, то есть для случаев масштабного развития городских районов после 2003 года и для общего обновления старой основы выборки, подготовленной до 2003 года. Различие заключается в объеме и масштабе работы по обновлению данных. Обновление старой основы выборки, разработанной до 2003 года, требует больших затрат сил и средств, чем обновление основ выборки, разработанных позднее. Однако в любом случае работа должна быть проведена в отношении всей основы выборки, а не только в отношении тех счетных участков – ПВЕ, – которые оказались отобранными в выборку; собранная в процессе обновления информация используется для отбора выборки.

Важно помнить, что обновление основы выборки является важнейшей статистической операцией, и если в нем возникает необходимость, то проведение такого рода обновления нельзя игнорировать при расчете затрат в рамках подготовки бюджета. Наряду с этим вам настоятельно рекомендуется воспользоваться услугами вашей национальной статистической организации, если обновление сочтено необходимым. Конкретные этапы заключаются в следующем:

1. Выявить зоны, особенно в крупных городах, где после проведения переписи населения осуществлялось масштабное жилищное строительство, вне зависимости от того проводилась ли перепись населения в вашей стране до или после 2003 года.
2. Выявить новые районы, такие как общины "скваттеров", население которых значительно увеличилось после проведения переписи; в их число могут входить районы, которые были "пустыми" или весьма слабо заселенными во время проведения переписи населения.
3. Игнорировать старые, стабильные жилые районы, где со временем произошли незначительные изменения.
4. Сопоставить зоны, выявленные на этапах 1 и 2, с их счетными участками, не забывая учитывать перекрывающиеся друг друга границы.
5. В требующих обновления данных счетных участках собрать сведения по каждому из них и произвести быстрый подсчет единиц жилья. Учтите, что быстрый подсчет подразумевает лишь приблизительный подсчет единиц жилья без фактического перечисления проживающих. Быстрый подсчет отнюдь не предполагает посещение всех домашних хозяйств, за исключением случаев многоквартирных домов, когда при взгляде с улицы не ясно, сколько в них имеется квартир.

Используйте данные быстрого подсчета единиц жилья<sup>13</sup> вместо первоначального подсчета домашних хозяйств в основе переписи населения. Это будет являться новым "измерением размера", то есть подсчетом, необходимым для определения вероятности отбора счетных участков при подготовке выборки.

Очевидно, что обновление основы выборки до составления выборки является достаточно сложной задачей, выполнение которой требует значительных затрат времени и средств. Это является одной из причин, по которой рекомендуется при возможности использовать имеющуюся выборку.

---

<sup>13</sup> Общеизвестно, что количество единиц жилья может не совпадать с количеством домашних хозяйств, однако важно получить их приблизительную оценку в целях определения масштабов размера. Например, если 120 единиц жилья были "быстро подсчитаны" в счетном участке, который был выбран для составления выборки, а впоследствии обнаружилось, что в этих единицах жилья находятся 132 домашних хозяйства, то это не окажет серьезного воздействия на обоснованность и достоверность результатов выборки.

### **ИСПОЛЬЗОВАНИЕ НОВОЙ ВЫБОРКИ ДЛЯ МИКС И ПРИНЯТИЕ РЕШЕНИЯ В ОТНОШЕНИЕ ЕЕ СХЕМЫ**

Как уже отмечалось в настоящей главе, в случае отсутствия соответствующей выборки для ее использования в рамках МИКС3, либо в рамках проведения самостоятельного обследования, либо в рамках использования в качестве дополнения к другому обследованию, необходимо спланировать и создать новую выборку, начиная с подготовки основы выборки (рассматривалось выше).

В настоящем разделе Руководства мы даем рекомендации по основным характеристикам схемы выборки МИКС3. Размер выборки, безусловно, является важнейшим показателем, и он подробно рассматривался в начале главы. Ниже предлагаются два варианта, но сначала кратко описываются общие характеристики.

В наиболее общих чертах используемая вами при проведении обследования выборка должна характеризоваться параметрами, рассмотренными выше в настоящей главе. Она должна быть вероятностной выборкой на всех этапах отбора и общенациональной по охвату и должна иметь возможно более простую структуру, для того чтобы ее реализация на местах могла быть легко и добросовестно осуществлена с минимальными возможностями отступления от схемы. Для обеспечения простоты как стратификация, так и количество этапов отбора должны быть минимальными. Основная цель стратификации заключается в повышении точности результатов обследования, а также в обеспечении выборки с запасом в отношении субнациональных районов, когда эти районы представляют особый интерес. Стратификацией, которую легко осуществить и которая является чрезвычайно эффективной в том случае, когда основной интерес представляют оценки на общенациональном уровне, является неявная стратификация. Это разновидность географической стратификации, которая при ее использовании вместе с систематическим отбором по принципу ВПР<sup>14</sup> (см. примеры в конце настоящей главы) автоматически пропорционально распределяет выборку между национальными административными образованиями, а также между городскими и сельскими районами. Неявная стратификация проводится путем географического упорядочивания основы выборки в серпантинной форме с разделением городских и сельских районов до использования систематической ВПР.

Далее, построение выборки должно осуществляться в три этапа. На первом этапе должны быть определены первичные выборочные единицы (ПВЕ) по возможности в качестве счетных участков, и они должны быть отобраны с использованием ВПР. Рекомендуется использовать счетные участки, поскольку ПВЕ должна быть районом, вокруг которого можно эффективно организовать работу на местах; она должна быть достаточно маленькой для составления карт, сегментации или составления списков проживающих в домашних хозяйствах, но достаточно большой для ее быстрой идентификации при проведении работы на местах.

---

<sup>14</sup> Это означает вероятность, пропорциональную размеру (ВПР), и относится к методу отбора районов выборки пропорционально численности населения; в результате чего район с численностью населения в 600 человек имеет в два раза более высокую вероятность отбора, чем район с населением в 300 человек.

Второй этап заключается в выборе сегментов (кластеров), а на третьем этапе в рамках каждого сегмента осуществляется отбор отдельных домашних хозяйств, которые должны быть опрошены в ходе обследования. Эти домашние хозяйства могут быть отобраны различными способами, например путем составления подвыборки из имеющегося списка домашних хозяйств в каждом сегменте или из нового списка.

Эта схема выборки, несомненно, предполагает определенную гибкость, что позволяет учитывать конкретные условия страны и ее потребности. Схема, скорее всего, будет отличаться в разных странах в отношении количества отобранных ПВЕ, количества сегментов или кластеров из расчета на одну ПВЕ и количества домашних хозяйств в каждом сегменте и, соответственно, общего размера выборки.

В соответствии с весьма общим эмпирическим правилом:

- количество ПВЕ должно составлять от 250 до 350;
- размеры кластеров (то есть количество домашних хозяйств, которые должны быть опрошены в каждом сегменте) должны составлять от 10 до 30, в зависимости от того какой из описанных ниже вариантов используется;
- общий размер выборки должен составлять от 2500 до 14 000 домашних хозяйств.

Та или иная страна может в собственных целях принять решение о том, что ей необходимо получить оценки показателя для нескольких субрегионов в дополнение к общестрановым оценкам. В этом случае ее структура выборки должна, безусловно, предполагать иную схему стратификации и большее количество ПВЕ в целях обеспечения соответствующей географической репрезентативности районов выборки в каждом субрегионе. В дополнение к этому размер выборки обследования должен быть существенно увеличен для обеспечения достоверных оценок для каждого субрегиона и других субнациональных *территорий* (рассматривается более подробно ниже в настоящей главе).

## СТАНДАРТНАЯ СХЕМА СЕГМЕНТА – ВАРИАНТ 2

Выше упоминалось, что программа ОНЗ может предоставить существующую выборку, пригодную для использования в рамках МИКСЗ (мы упоминали об использовании существующей выборки в качестве варианта 1). Стандартная схема выборки ОНЗ является хорошей моделью для МИКСЗ в том случае, если вы принимаете решение о необходимости построения новой выборки. Модель выборки ОНЗ также использовалась в других программах обследований, имеющих отношение к сфере здравоохранения, таких как обследования ПАПЧАЙЛД в арабских странах<sup>15</sup>.

Модели выборок ОНЗ и ПАПЧАЙЛД основываются на так называемой *стандартной схеме сегмента*, которая характеризуется преимуществами методики вероятностной выборки, а также простотой и соответствием целям МИКСЗ как в отношении существа этих целей, так и в отношении статистики. В руководствах по построению выборок для ОНЗ и ПАПЧАЙЛД

---

<sup>15</sup> См. The Arab Maternal and Child Health Survey, Basic Documentation 5: Sampling Manual, League of Arab States, Cairo, 1990.

отмечается, что большинство стран располагают репрезентативными территориальными основами выборки в виде счетных участков последней переписи населения. По счетным участкам обычно существуют схемы, а также информация о численности населения и/или количестве домашних хозяйств. Счетные участки переписи обычно имеют приблизительно одинаковые размеры. Во многих странах не существует списков жилых кварталов или домашних хозяйств должного качества, в них также отсутствует адекватная система адресов, особенно в сельских районах. Соответственно, для обновления основы выборки необходимо подготовить новые списки проживающих в домашних хозяйствах.

Для применения стандартной схемы сегментов к МИКСЗ в первую очередь выстройте основу выборки счетных участков в географической последовательности для неявной стратификации. Некоторые счетные участки настолько велики, что экономически нецелесообразно составлять новые списки всех домашних хозяйств, в том случае если они уже отобраны. Вместо этого более эффективным представляется использование сегментов. Это осуществляется путем присвоения каждому счетному участку измерения его размера, равного желательному количеству содержащихся в нем "стандартных сегментов". В руководствах по составлению выборок ОНЗ и ПАПЧАЙЛД рекомендуется, чтобы количество стандартных сегментов было определено (и подсчитано) путем деления полученной в результате переписи численности населения в счетном участке на 500 и последующего округления результата до ближайшего целого числа. Следует иметь в виду, что в случаях обновления основы выборки переписи населения следует использовать не данные о численности населения, полученные в результате переписи, а *данные о количестве единиц жилья (умноженном на 5)*, которые вы получили на последнем этапе операции по обновлению данных, описанной в предшествующем разделе, посвященном основам выборок. Множитель, равный 5, нужен для приблизительного подсчета численности населения в обновленных счетных районах на текущий момент, с тем чтобы их измерения размера определялись так же, как и в счетных участках, данные по которым не обновлялись. *Этот размер стандартного сегмента рекомендуется в отношении МИКСЗ в том случае, если вы принимаете решение использовать вариант 2.*

Следующий этап заключается в отборе счетных участков с использованием вероятности, пропорциональной этому измерению размера. Следует иметь в виду, что измерение размера – это также количество сегментов. Во многих случаях вы обнаружите, что средний размер счетного участка составляет около 500 человек (или, соответственно, 100 домашних хозяйств при среднем размере домашнего хозяйства, равном 5 человек); соответственно типичным измерением размера является единица.

Следующим этапом является сегментация с использованием имеющихся карт. *Если количество сегментов в счетном участке выборки равно одному, сегментация не требуется*, поскольку сегмент и счетный участок совпадают. Если же количество сегментов превышает 1, необходима сегментация. Это предполагает деление отобранных счетных участков на части (по количеству сегментов), каждая из которых содержит приблизительно одинаковое количество домашних хозяйств. Сегментация может быть как камеральная работа, при условии что карты являются достаточно точными; в противном случае потребуется выезд на место, особенно в тех случаях, когда различимые внутренние границы в счетном участке недостаточно четки (см. в главе 6 более подробное изложение проблем составления карт и сегментации).

**Таблица 4.8**  
**Вариант 2 – Краткая характеристика стандартной схемы сегмента**

<p><b>Отличительные черты</b></p> <ul style="list-style-type: none"><li>• Трехэтапное построение выборки при неявной стратификации</li><li>• Отбор счетных участков методом ВПР</li><li>• Составление карт и сегментация в счетных участках, состоящих из нескольких стандартных сегментов</li><li>• Произвольная выборка одного сегмента в каждом счетном участке</li><li>• Составление списка домашних хозяйств в сегментах выборки</li><li>• Планомерный отбор выборочных домашних хозяйств в сегментах</li></ul> <p><b>Параметры</b></p> <ul style="list-style-type: none"><li>• Обычно от 250 до 400 счетных участков выборки (ПВЕ)</li><li>• Стандартные сегменты размером в 500 человек (около 100 домашних хозяйств)</li><li>• Некомпактный кластер размером от 10 до 35 домашних хозяйств (отличающийся от приведенного ниже варианта 3)</li><li>• Обычный объем выборки от 4000 до 14 000 домашних хозяйств*</li></ul> <p><small>* Следует отметить, что в целом мы не рекомендуем умножать минимальное количество ПВЕ на минимальный размер кластера (250 умножить на 10), поскольку объем выборки, равный 2500, в большинстве стран является, вероятно, слишком маленьким для достоверного определения величины важнейших показателей.</small></p>
--

После проведения сегментации произвольно выбирается один сегмент в каждом районе счетных участков выборки. Во всех отобранных сегментах *вновь составляется список домашних хозяйств*, в который обычно опять-таки входят около 100 домашних хозяйств. После этого на основании списков с использованием фиксированной доли производится систематический отбор домашних хозяйств в каждом отобранном сегменте для проведения опроса.

**Таблица 4.9**  
**Вариант 2 – Стандартная схема сегмента**

<p><b>Преимущества</b></p> <ul style="list-style-type: none"><li>• Вероятностная выборка</li><li>• Минимальный объем работы по составлению карт и сегментации</li><li>• Минимальный объем работы по составлению списков</li><li>• Несколько более достоверный, чем вариант 3 (ниже)</li><li>• Частично корректирует старую основу выборки</li><li>• Самовзвешенная схема</li></ul> <p><b>Недостатки</b></p> <ul style="list-style-type: none"><li>• В каждом сегменте выборки необходимо составление списков, хотя и в минимальном объеме</li><li>• Может привести к сильно различающимся по размерам сегментам, особенно если основа выборки составлена давно и не обновлялась</li></ul>
---

**ПРИМЕР:**

Может быть принято решение об отборе каждого пятого домашнего хозяйства, включенного в новый список, в каждом сегменте выборки. Таким образом, при наличии, например, 300 сегментов количество выбранных домашних хозяйств в каждом сегменте будет приблизительно равно 20 (при том, что это количество будет различным в разных ПВЕ), а общий объем выборки составит приблизительно 6000 домашних хозяйств.

Стандартная схема сегмента удобна и практична. В типичной стране, где размер учетного участка в среднем равен приблизительно 100 домашним хозяйствам, фактически потребуются проведение лишь незначительной сегментации. Кроме того, и объем работы по составлению списка домашних хозяйств будет ограничен.

Домашние хозяйства выборки в рамках варианта 2 находятся внутри некомпактных кластеров<sup>16</sup>, а выборка является самовзвешенной. Количество домашних хозяйств, отобранных в каждой выборочной ПВЕ, будет несколько различным, поскольку ПВЕ отбираются на основании данных об их размерах по результатам переписи населения (за исключением тех, данные по которым были уточнены), которые, вероятно, окажутся отличными от фактических размеров, когда будет подготовлен новый список домашних хозяйств.

**ПРИМЕР:**

Предположим, что квота отбора внутри сегмента рассчитана как 1 из 5 включенных в список домашних хозяйств. Если сегмент отобран, исходя из предположения, что в нем по данным переписи находится 98 домашних хозяйств, а в результате составления нового списка оказывается, что в нем в настоящее время имеется 112 домашних хозяйств, то в этом случае квота отбора, равная 1/5 домашних хозяйств, даст результат в размере 22 или 23 домашних хозяйств (истинное число) вместо ожидавшихся 19 или 20. Эта процедура не только правильным образом отражает изменение состава населения, но и сохраняет самовзвешенность выборки. Расхождения в среднем размере сегмента не должны быть большими, за исключением случаев использования старой основы переписи населения, которая не была обновлена<sup>17</sup>.

---

<sup>16</sup> Некомпактным кластером является кластер, в котором домашние хозяйства, отобранные для выборки, расположены систематически во всем районе выборки. Компактным кластером является кластер, в котором каждое отобранное домашнее хозяйство в конкретном сегменте граничит с соседним домашним хозяйством. Некомпактные кластеры дают более достоверные результаты, чем компактные кластеры, в связи с меньшим значением эффекта планирования.

<sup>17</sup> Существует альтернативная процедура, используемая в тех случаях, когда есть основания полагать, что структура населения существенным образом изменилась, в результате чего средний размер сегмента может быть слишком непостоянным для эффективной работы на местах. В этом случае фиксированным может быть размер сегмента, а не подлежащая отбору доля домашних хозяйств, что потребует расчета нового интервала выборки и его применения в каждом сегменте выборки. В результате каждый сегмент будет иметь различный вес, что должно быть учтено при подготовке оценок показателя.

### ИЗМЕНЕННАЯ СХЕМА СЕГМЕНТА – ВАРИАНТ 3

Мы рассмотрели использование существующей выборки в качестве предпочтительного варианта при проведении МИКС3 при условии наличия и достоверности имеющейся и должным образом построенной выборки. Мы также рассмотрели вопрос об использовании типового плана выборки ОНЗ и ПАПЧАЙЛД, а также стандартной схемы выборки в качестве следующего по качеству варианта в том случае, когда вашей стране необходимо с нуля построить выборку для проведения обследования.

Вариант 3 использует модификацию стандартной схемы сегмента. *Измененная схема сегмента* аналогична стандартной схеме сегмента, однако между ними существуют важные различия<sup>18</sup>. Вместо создания стандартных сегментов в размере 500 человек в каждом счетном участке выборки последний подразделяется на заранее определенное количество сегментов. Это заранее определенное количество сегментов равно количеству домашних хозяйств по данным переписи населения (или уточненному количеству единиц жилья) в счетном участке, разделенному на желаемый размер кластера и округленному в сторону ближайшего целого числа. Следует отметить, что в этом варианте мы используем домашние хозяйства (или единицы жилья в отношении районов с обновленными данными), а не население, которое использовалось в варианте 2. В связи с этим нет необходимости умножать на пять количество единиц жилья в районах, по которым данные обновлялись.

**ПРИМЕР:**

Если желаемый размер кластера равен 20 домашним хозяйствам, и при этом в счетном участке имеется 155 домашних хозяйств, в этом случае будет сформировано 8 сегментов.

Так же, как и в варианте 2, отбор счетных участков производится с вероятностью, пропорциональной количеству содержащихся в них сегментов. После этого каждый отобранный счетный участок делится на заранее определенное количество сегментов с использованием карт-схем и проведением быстрого подсчета имеющихся в настоящее время единиц жилья. В рамках проведения сегментации следует четко определить границы, а количество единиц жилья в каждом сегменте должно быть приблизительно одинаковым, хотя оно и не должно обязательно быть равным. Следует отметить, что быстрый подсчет может опять-таки основываться на количестве единиц жилья, а не домашних хозяйств, как и при обновлении основы выборки (более подробно см. соответствующий подраздел).

После проведения сегментации в рамках каждого счетного участка выборки произвольно отбирается один (и только один) сегмент. *Все домашние хозяйства, находящиеся в границах выбранного сегмента, после этого опрашиваются в рамках проведения обследования, в результате чего сегмент образует компактный кластер домашних хозяйств.*

---

<sup>18</sup> См. подробное описание измененной схемы сегмента (или кластера) в: Turner, A., R. Magnani, and M. Shuaib, "A Not Quite as Quick but Much Cleaner Alternative to the Expanded Programme on Immunization (EPI) Cluster Survey Design," *International Journal of Epidemiology*, 1996, Vol. 25, No.1.

Другие особенности измененной схемы сегмента являются практически такими же, как и при стандартной схеме сегмента – трехэтапное построение выборки, неявная стратификация, отбор счетных участков по методу ВПР.

Измененная методика построения сегмента имеет преимущество по сравнению со стандартной схемой сегмента в том, что в случае ее использования отсутствует необходимость в составлении списка домашних хозяйств, что позволяет избежать существенной части затрат на проведение обследования. Быстрый подсчет и составление карт-схем, тем не менее, предполагают дополнительные затраты, однако расходы на проведение быстрого подсчета являются минимальными, поскольку его можно осуществить визуально, а не посещая жилища и беседуя с респондентами. Кроме того, процедура в некоторой степени компенсирует использование, возможно, устаревшей основы выборки благодаря проведению опроса всех имеющихся в настоящее время домашних хозяйств в выбранном сегменте вне зависимости от того, сколько их было во время проведения переписи населения.

**Таблица 4.10**

**Вариант 3 – Краткое описание измененной схемы сегмента**

**Отличительные особенности**

- Трехэтапное построение выборки при неявной стратификации
- Предварительное определение количества сегментов с использованием ПВЕ
- Отбор счетных участков переписи методом ВПР
- Составление карт и сегментация во всех выбранных счетных участках выборки
- Произвольный выбор одного сегмента в каждом из счетных участков
- Опрос всех выбранных домашних хозяйств в отобранном сегменте

**Параметры**

- Обычно от 250 до 400 счетных участков выборки (ПВЕ)
- Компактный кластер в размере от 20 до 30 домашних хозяйств (минимальный размер составляет 20)
- Обычный размер выборки от 5 тыс. до 12 тыс. домашних хозяйств\*
- *Размеры сегмента и кластера совпадают (в отличие от варианта 2)*

*\* Следует обратить внимание на то, что диапазон размеров выборки отличается от аналогичных показателей в варианте 2, приведенных в таблице 4.8, в связи с иными рекомендуемыми размерами компактного кластера.*

Недостатком измененной схемы сегмента является то, что сегменты (кластеры) компактны. В связи с этим при том же размере выборки достоверность выборки при этой схеме будет несколько ниже, чем при стандартной схеме сегмента, когда кластеры не компактны. Вместе с тем это можно компенсировать путем отбора большего количества счетных участков с меньшим по размеру охватом выборки в рамках счетных участков. Еще один недостаток заключается в том, что для проведения самой сегментации требуется определение относительно небольших сегментов, что может оказаться нереалистичным в некоторых странах. Точную или даже адекватную сегментацию может быть чрезвычайно сложно осуществить в небольших районах, где

отсутствует достаточное количество естественных границ, таких как дороги, аллеи, реки и т. д. *В связи с этим мы рекомендуем, чтобы размер сегмента при данном варианте был равен не менее чем 20 домашним хозяйствам; а для компенсации снижения достоверности в рамках компактного сегмента он не должен превышать 30 домашних хозяйств.* В отношении контролирования смещения выборки чрезвычайно важно обозначить границы при образовании сегментов.

### **УПРОЩЕННЫЕ СХЕМЫ – НЕ РЕКОМЕНДУЮТСЯ**

В ходе первого раунда МИКС, который был проведен в 1995 году, большое внимание было уделено методу случайного блуждания, который используется в рамках Расширенной программы иммунизации (РПИ). Основной довод против использования метода случайного блуждания в контексте МИКС3 заключается в том, что выбор домашних хозяйств основан не на методах вероятностного определения выборки, но скорее на процедуре, результатом которой фактически является выборка на основе квот.

Поскольку для МИКС3 характерны крупные размеры выборки, процедура случайного блуждания является неприемлемой. Иногда утверждается, что в маломасштабных обследованиях РПИ, при характерных для них небольших размерах выборки, чаще присутствует дисперсия выборки, чем смещение, что до некоторой степени оправдывает использование метода случайного блуждания.

Однако в случае МИКС3 те же доводы приводят к противоположному выводу, заключающемуся в том, что смещение представляет собой более серьезную проблему, чем дисперсия выборки из-за гораздо более значительного размера выборки, и поэтому на каждой стадии отбора следует использовать более жесткие вероятностные методологии.

---

**Упрощенные процедуры, такие как случайное блуждание, которые являются отклонением от вероятностных схем, не рекомендуются для МИКС3, и их не следует использовать.**

---

**Таблица 4.11**

### **Краткий контрольный перечень для определения размера и схемы выборки**

- Определите целевую группу, составляющую небольшую долю общей численности населения
- Определите степень охвата для этой целевой группы
- Выберите размер выборки из таблицы 4.3, если ситуация в вашей стране соответствует изложенным в этой таблице допущениям и условиям; В ПРОТИВНОМ СЛУЧАЕ рассчитайте размер выборки по формуле, предложенной в настоящей главе
- Определите размер кластера, обычно от 10 до 35 домашних хозяйств
- Разделите размер выборки на размер кластера, для того чтобы выяснить число ПВЕ (районов выборки)
- Проанализируйте свой выбор числа  $n$ , размера кластера и числа ПВЕ, ДЛЯ ТОГО ЧТОБЫ выбрать для схемы выборки вариант 1, 2 или 3

## СПЕЦИАЛЬНЫЕ ТЕМЫ, КАСАЮЩИЕСЯ ВЫБОРКИ МИКСЗ

В настоящем разделе мы рассмотрим несколько других важных тем, которые следует учитывать в процессе планирования связанных с определением выборки аспектов МИКСЗ в вашей стране. Среди этих тем – субнациональные оценки, оценка динамики и аналитические подгруппы, а также оценки в области водоснабжения/санитарно-гигиенических условий.

### СУБНАЦИОНАЛЬНЫЕ ОЦЕНКИ

До сих пор нас интересовали размеры выборки, необходимые для выработки национальных оценок показателей. Вместе с тем многие страны также захотят использовать МИКСЗ для получения субнациональных данных, например на уровне городских/сельских районов, регионов, штатов или провинций или, возможно, округов. Такие данные можно использовать для определения тех областей, где необходимо приложить дополнительные усилия, а также для составления программ и оценки.

Одним из важнейших ограничивающих факторов в процессе получения достоверных субнациональных оценок является размер выборки. Для каждой *территориальной единицы*, по которой составляется отчет (т. е. субнациональной единицы, такой как регион или городской/сельский район), общий размер выборки необходимо существенно увеличить, для того чтобы получить достаточно достоверные результаты. Если по каждой территориальной единице желательно получить *одинаково* достоверные результаты, обычно увеличивают размер *национальной* выборки (число  $n$ ) на множитель, близкий к числу таких единиц, тем самым выбирая  $n$  случаев *по каждой территориальной единице*. Другими словами, если мы хотим получить одинаково достоверные данные по 5 регионам страны, мы должны *умножить* размер выборки, рассчитанный для национальных оценок на основе таблицы 4.3 или непосредственно с помощью формулы определения размера выборки, *на множитель, равный примерно 5*, для того чтобы получить региональные оценки. Это, естественно, приводит к существенному увеличению общего размера выборки (и расходов) и для многих стран может быть неосуществимо<sup>19</sup>.

При этом необходимо прибегать к компромиссным решениям, особенно если мы имеем дело с большим числом территориальных единиц. Можно рассмотреть различные альтернативные варианты. Одна из возможных альтернатив заключается в том, чтобы ограничить отдельные единицы, по которым составляется отчет, например такие, как провинции, только теми, численность населения которых превышает определенный порог. Оставшиеся субрегионы можно свести в региональные группы. Еще одна альтернатива состоит в том, чтобы допустить для оценок по таким территориальным единицам менее жесткую степень точности по сравнению с той, которая устанавливается для национальных оценок. Например, предел ошибки по ключевому показателю для национальной оценки установлен на уровне 12 процентов  $r$  (см. подраздел по размерам выборки), однако по отдельным территориальным единицам предел ошибки может быть

---

<sup>19</sup> Вместо этого Виджей Верма предлагает увеличивать размер выборки на множитель  $D^{65}$ , где  $D$  – число территориальных единиц. При таком подходе достоверность оценки по каждой территориальной единице является несколько меньшей, чем общенациональная оценка. См. "A Critical Review of MICS Sampling Methodology," Report by Verma to UNICEF, April 1995.

гораздо более высоким и доходить даже до величины от 25 до 30 процентов г. Кроме того, эти два альтернативных решения можно использовать в сочетании друг с другом.

### ОЦЕНКА ДИНАМИКИ И АНАЛИЗ ПОДГРУПП

Некоторые международные цели имеют вид ожидаемых сокращений, например сокращение недоедания в мире на 20 процентов за пятилетний период. В вашем распоряжении могут также быть результаты обследования МИКС или обследования ОНЗ, проведенного несколько лет назад, где содержатся оценки показателей, включенных в МИКСЗ, и вы можете захотеть провести оценку изменений, происшедших с момента проведения такого обследования. Для такого вида оценки необходимо иметь результаты двух обследований – на начало и на конец периода. Размер выборки, необходимый для оценки изменений, происшедших в интервале между этими двумя моментами, в значительной степени зависит от масштабов изменений, а также масштабов двух оценок, сделанных в эти моменты. Этот процесс является довольно сложным, и было бы нецелесообразно давать краткие и общие руководящие принципы оценки происшедших изменений. Рекомендуем вам обратиться в национальное статистическое управление или специализированный орган, занимавшийся определением выборки, если вы планируете провести оценку происшедших изменений.

Что касается анализа подгрупп, таких как показатели в разбивке по признаку пола или социально-экономическим группам, оценки показателей будут менее точными, чем оценки по выборке в целом.

На представленном ниже примере показано, как пределы ошибки увеличиваются по мере уменьшения численности подгрупп.

**ПРИМЕР:**

Основываясь на полной (общенациональной) выборке, если рассчитанная степень точности составляет, к примеру,  $\pm 5$  процентных пунктов для 50-процентного охвата, предел ошибки будет составлять примерно  $\pm$

- 6,3 процентных пункта для показателей в разбивке по признаку пола, если допустить, что 50 процентов выборки составляют мальчики и 50 процентов – девочки;
- 8,6 процентных пункта для подгруппы, составляющей 20 процентов от размера всей выборки<sup>20</sup>.

Таким образом, достаточно точные результаты можно получить для показателей в разбивке по признаку пола, а также по другим подгруппам, составляющим пятую или бóльшую часть от размера всей выборки.

---

<sup>20</sup> См. неопубликованную записку ЮНИСЕФ "Some Proposed Modifications for the WHO Simplified Cluster Sampling Method for Estimating Immunization Coverage," by Graham Kalton, September 1988, page 10.

## ПОКАЗАТЕЛИ В ОБЛАСТИ ВОДОСНАБЖЕНИЯ И САНИТАРНО-ГИГИЕНИЧЕСКИХ УСЛОВИЙ

Сбор данных о водоснабжении и санитарно-гигиенических условиях является важным компонентом обследования МИКСЗ. Вместе с тем при использовании и оценке данных возникает ряд статистических вопросов, которые необходимо учитывать.

Как уже говорилось выше, схема выборки для МИКС основана на сборе индивидуальных переменных, а не переменных о домашних хозяйствах. Доступ к водоснабжению и санитарно-гигиеническим устройствам является одной из характеристик домашних хозяйств в том смысле, что все домашнее хозяйство обладает одинаковым доступом к этим услугам. Кроме того, в случае территориального кластера чаще всего все домашние хозяйства, относящиеся к этому кластеру, имеют одинаковый доступ к таким услугам. Тем самым *эффект* построения выборки (ЭПВ), о котором говорилось ранее, существенно больше для показателей, касающихся доступа к водоснабжению и санитарно-гигиеническим устройствам, чем значение 1,5, которое мы приняли при расчете размеров выборки. В результате этого ошибки выборки по показателям доступа к водоснабжению и санитарно-гигиеническим устройствам будут намного выше, чем по индивидуальным показателям.

Следует отметить, что, если бы главная цель МИКСЗ заключалась в том, чтобы собрать данные о типах доступа к водоснабжению и санитарно-гигиеническим устройствам, схема выборки была бы существенно иной. План выборки, скорее всего, будет предусматривать проведение упрощенного общинного обследования, в котором одному респонденту в каждой общине выборки (зональном кластере) будут задавать вопросы о водоснабжении и санитарно-гигиенических условиях, в отличие от метода, используемого в обследовании МИКСЗ, при котором эти вопросы задаются каждому домашнему хозяйству, включенному в выборку.

Несмотря на ожидаемую высокую степень погрешности при определении выборки для оценки показателей, касающихся водоснабжения и санитарно-гигиенических условий, результаты, тем не менее, будут весьма полезными, особенно для выявления *тенденций*. Это связано с тем, что данные по этим показателям собирались и в ходе других обследований домашних хозяйств, таких как ОНЗ, структура которых аналогична МИКСЗ. Сравнение результатов МИКСЗ с результатами других обследований будет весьма полезным для целей выявления тенденций, поскольку эффекты построения выборки, хотя и значительные в обоих обследованиях, несколько сглаживаются при оценке динамики.

## ПОДГОТОВКА ОЦЕНОК И РАСЧЕТ ОШИБОК ВЫБОРКИ

В этом разделе мы обсудим различные варианты взвешивания для подготовки оценок, а также необходимость расчета ошибок выборки.

По мере необходимости при расчете оценок показателей можно последовательно применять два типа взвешивания. За исключением тех случаев, когда домашние хозяйства были отобраны с едиными общими вероятностными характеристиками (т. е. самовзвешенная структура), все данные по выборке должны взвешиваться с использованием обратных величин общих вероятностей выборки – так называемых *расчетных весов*. Вместе с тем расчетные веса

следует скорректировать для учета отказов от ответа, даже если выборка обладает самовзвешенной структурой. Это можно сделать разнообразными способами, включая взвешивание респондентов в каждой ПВЕ (или кластере), для того чтобы отразить число не ставших респондентами в данной ПВЕ. Главное преимущество этого подхода состоит в том, что для него не требуются данные из внешних источников. Эти два этапа – применение расчетных весов и корректировка для учета отказов от ответа – могут составить все взвешивание, необходимое для вашего обследования.

Дальнейшее взвешивание может производиться путем корректировки расчетных весов для обеспечения того, чтобы взвешенное распределение выборки по некоторым ключевым переменным, например по городским/сельским районам или регионам, соответствовало внешнему распределению населения, например последней по времени переписи населения. Возможность использования этого вида взвешивания после стратификации следует рассмотреть в случае существенных изменений схемы на этапе осуществления, когда из-за недостатков основы выборки приходится использовать процедуры приближения или когда в выборке не соблюдаются жесткие вероятностные критерии.

Фактические формулы и вычисления расчетных весов, множителей, корректировки для учета отказов от ответа и, если это необходимо, корректировки после стратификации в значительной степени зависят от используемой схемы выборки, т. е. количества этапов определения выборки, выбора между самовзвешивающимися и несамовзвешивающимися планами, определения того, являются ли отказы от ответов равномерно распространенными в различных подгруппах населения или регионах, или же их доля варьируется в значительных пределах наличия внешних данных корректировки после стратификации и т. д. Таким образом, нецелесообразно подробно разрабатывать этапы расчета весов для вашего МИКСЗ. Тем не менее см. пример расчета весов для отдельного МИКС на конкретном примере Ливана в последнем разделе настоящей главы. Статистик, занимающийся составлением структуры выборки, должен уметь составлять схему весов и готовить всю необходимую для этого документацию и нести за это ответственность.

## **РАСЧЕТ ОШИБОК ВЫБОРКИ**

Как уже неоднократно подчеркивалось в настоящей главе, размер выборки в рамках вашего обследования определяется ключевым показателем, для которого мы заранее определили ожидаемую степень точности. Однако в процессе проведения обследования будут получены сотни оценок, каждая из которых будет иметь свою собственную степень точности (ошибка выборки), и эта степень будет зависеть от национального или субнационального уровня оценки, а также от величины  $p$ ,  $g$  и ЭПВ, связанных с каждой из них (см. формулу определения размера выборки). Как уже отмечалось выше, к примеру, оценкам в области водоснабжения и санитарно-гигиенических условий, как ожидается, присуща гораздо более высокая степень ошибки выборки, чем другим показателям. Вот почему важно рассчитать ошибки выборки по некоторым (не всем) показателям.

Таким образом, в ходе обычной подготовки результатов обследования рекомендуется проводить оценку ошибок выборки и связанных с нею переменных, таких как ЭПВ по основным

показателям, – возможно, от 30 до 50 различных оценок. Это необходимо, для того чтобы определить достоверность оценок показателей. Следует помнить, что *доверительные интервалы, включающие предел ошибки оценок показателей обследования, невозможно определить, если не провести оценки ошибок выборки. Тем самым можно нанести серьезный ущерб толкованию оценок.*

Расчет ошибок выборки или *среднеквадратическая ошибка* может представлять собой достаточно сложную часть работы по проведению обследования. Среднеквадратические ошибки следует рассчитывать с учетом сложной схемы выборки (количества кластеров, стратификации и взвешивания). Неправильное использование простых формул случайного определения выборки будет, как правило, приводить к серьезному занижению среднеквадратических ошибок.

Поскольку в различных странах, которые примут участие в обследовании, мы, несомненно, будем иметь дело с самыми различными схемами выборки, включая те, которые основаны на уже существующих выборках, определить единую схему для оценки ошибок выборки в МИКСЗ не представляется возможным. Вместе с тем существует один способ, известный как *метод конечных кластеров*, который может оказаться пригодным для большинства схем выборки, используемых в рамках МИКСЗ. Для этой цели подготовлены простая электронная таблица Excel и синтаксический файл SPSS, а также инструкции, касающиеся применения данного метода, которые можно найти на веб-сайте МИКСЗ по адресу [www.childinfo.org](http://www.childinfo.org). В целом метод конечных кластеров можно использовать в том случае, если ваша выборка является самовзвешивающейся или приблизительно самовзвешивающейся.

В случае невозможности использовать электронную таблицу существует ряд разработанных программных пакетов, которые можно адаптировать для оценки переменных. К их числу относятся программа CLUSTERS, которая была изначально подготовлена для Всемирного обследования фертильности и которую можно получить в Университете Эссекса; программа CENVAR, которую можно бесплатно получить в Бюро переписи США; и программа WesVar, разработанная организацией WESTAT для использования вместе с программой SPSS. При сложных схемах можно также воспользоваться такими пакетами, как SAS, SUDAAN и Epi-Info. Последние версии программы SPSS также способны рассчитывать ошибки выборки при сложных схемах. Одни пакеты предоставляются бесплатно, и их даже можно загрузить из Сети, а другие можно приобрести на коммерческой основе<sup>21</sup>.

В любом случае независимо от того, используете ли вы электронную таблицу или пакет компьютерных программ, процесс можно упростить, если вам удастся обеспечить, чтобы отчеты по данным содержали идентификаторы ПВЕ. В случае использования типических групп эти группы следует определить для всех ПВЕ.

---

<sup>21</sup> См. комплексный анализ этих программ в “Sampling Error Software for Personal Computers,” by Jim Lepkowski and Judy Bowles of the University of Michigan. Эта статья опубликована в издании The Survey Statistician, No. 35, December 1996, 10-17 (см. [www.fas.harvard.edu/~stats/survey-soft/iass/html](http://www.fas.harvard.edu/~stats/survey-soft/iass/html)). Самую свежую информацию можно получить в издании Статистического отдела Организации Объединенных Наций под названием “Household Sample Surveys in Developing and Transition Countries”, март 2005 года, в особенности в главе 21 “Sampling Error Estimation for Survey Data”, by Donna Brogan ([http://unstats.un.org/unsd/HHSurveys/pdf/Household\\_surveys.pdf](http://unstats.un.org/unsd/HHSurveys/pdf/Household_surveys.pdf)).

## ПОДРОБНОЕ ОПИСАНИЕ МЕТОДОВ ОПРЕДЕЛЕНИЯ ВЫБОРКИ

В настоящем разделе мы представим подробную информацию и примеры того, как следует применять методы определения выборки, которые обсуждаются в данной главе.

### ПРОЦЕДУРЫ ОПРЕДЕЛЕНИЯ ВЫБОРКИ С ПОМОЩЬЮ ВПР – ВАРИАНТ 2

Первый метод, который мы хотим проиллюстрировать, касается порядка выбора единиц первого этапа с использованием ВПР. Эта иллюстрация также покажет вам, каким образом следует сочетать систематизированное определение выборки с использованием ВПР с географическим расположением основы выборки для обеспечения *неявной* стратификации.

Для иллюстрации воспользуемся вариантом 2, предложенным в настоящей главе, а именно схемой стандартного сегмента, и на его основе составим национальную выборку. Предположим, что 1) размер стандартного сегмента по варианту 2 составляет 500 человек или около 100 домашних хозяйств; 2) счетные участки переписи населения (СУ) принимаются за основу выборки; и 3) количество ПВЕ, которые следует отобрать, равняется 300. Шаги, предпринимаемые на первом этапе отбора, о которых говорится ниже, *следует выполнять в виде компьютерной операции*, хотя их можно выполнить и вручную.

- Шаг 1: разбейте файл СУ на городские и сельские районы;
- Шаг 2: в категории городских районов далее разбейте файл в географической серпантинной форме согласно административным единицам вашей страны (например, провинция или штат, округ, община и т. п.);
- Шаг 3: повторите шаг 2 в отношении категории сельских районов;
- Шаг 4: в первом столбце отразите численность населения СУ;
- Шаг 5: в следующем столбце выведите число стандартных сегментов, которое равно численности населения, поделенной на 500 и округленной до ближайшего целого числа. Это измерение размера СУ;
- Шаг 6: в следующем столбце сведите воедино измерения размера;
- Шаг 7: рассчитайте интервал выборки  $I$ , разделив общую сумму на 300, с точностью до одной десятой. В данном примере допустим, что общая сумма равна 5281. В этом случае интервал выборки  $I$  будет равен  $5281/300$  или 17,6;
- Шаг 8: выберите случайную точку отсчета от 0 до 17,6. На практике для этого следует воспользоваться таблицей случайных чисел и выбрать трехзначное число от 001 до 176, а впоследствии вставить десятичную запятую. Предположим, вы выбрали 042; в этом случае вашей случайной точкой отсчета будет 4,2. Тогда вашей первой выборочной ПВЕ будет та, для которой совокупное измерение размера будет иметь наименьшее значение, равное или превышающее 4,2<sup>22</sup>;
- Шаг 9: добавьте 4,2 к  $I$ , другими словами,  $4,2 + 17,6 = 21,8$ ; в этом случае вашей следующей ПВЕ в выборке окажется та, совокупный показатель которой будет иметь наименьшее значение, равное или превышающее 21,8;

---

<sup>22</sup> Киш рекомендует прибегать к округлению в тех случаях, когда интервал выборки выражен нецелым числом. См. Kish, L. (1965) Survey Sampling, Wiley, New York, page 116.

- Шаг 10: добавьте 21,8 к I, другими словами  $21,8 + 17,6 = 39,4$ ; следующей ПВЕ в выборке окажется та, совокупный показатель которой будет иметь наименьшее значение, равное или превышающее 39,4;
- Шаг 11: продолжайте проводить эти операции по всем городским СУ, после чего переходите к сельским районам, до тех пор пока не будут выбраны все 300 ПВЕ.

Эта процедура подробнее показана в таблице 4.12.

Две выборочных ПВЕ, которые показаны на этом примере, относятся к СУ 003 общины 01 и СУ 002 общины 03; обе они расположены в округе 01 и провинции 01. В случае первого СУ измерением его размера является 3 – это означает, что необходимо будет создать три сегмента, каждый численностью около 540 человек (1630 разделить на 3), после чего один из трех сегментов будет произвольно выбран для составления списка и подвыборки домашних хозяйств. Во втором выборочном СУ будут созданы два сегмента, каждый из примерно 590 человек, прежде чем один из них будет произвольно выбран.

Данный пример показывает многочисленные преимущества метода неявной стратификации. Во-первых, она крайне проста и требует лишь географической сортировки основы счетных участков до начала систематизированного отбора выборки с помощью ВПР. Во-вторых, она обеспечивает автоматическое создание выборки ПВЕ, которая пропорционально разделяется на городские и сельские районы и по провинциям (или другим территориальным единицам). Например, если 10 процентов вашего населения проживает в провинции 12, то 10 процентов вашей выборки следует отбирать именно из этой провинции. В-третьих, ее легко производить на компьютере.

Таблица 4.12

**Иллюстрация систематизированного определения выборки с помощью ВПР  
и неявной стратификации – вариант определения выборки 2**

Городской район Провинция 01 Округ 01 Община 01	Численность населения	Измерение размера (число сегментов с населением в 500 человек)	Общая сумма
СУ 001	1 470	3	3
СУ 002	562	1	4
СУ 003	1 630	3	Выбрано 7
СУ 004	1 006	2	9
Община 02			
СУ 001	412	1	10
СУ 002	1 537	3	13
СУ 003	1 312	3	16
СУ 004	397	1	17
Община 03			
СУ 001	1 540	3	20
СУ 002	1 181	2	Выбрано 22
СУ 003	1 025	2	24
Округ 02			
Община 01			
СУ 001	567	1	25
СУ 002	1 111	2	27
СУ 003	409	1	28
*			
*			
и т. д.			
Сельский район Провинция 12 Округ 05 Община 05			
СУ 001	512	1	5 280
СУ 002	493	1	5 281

После того, как ПВЕ, согласно варианту 2, будут выбраны, необходимо провести сегментацию в тех ПВЕ, где измерение размера (количество сегментов) составляет 2 или более, после чего произвольно выбирается один сегмент. После этого потребуется составить новый список домашних хозяйств по выбранным сегментам, а также по ПВЕ, состоящим из одного сегмента. Заключительный шаг процедуры отбора в рамках варианта 2 состоит в отборе выборочных домашних хозяйств внутри выбранных сегментов. Данная процедура показана на примере, который приводится в таблице 4.13.

**Таблица 4.13**  
**Выбор домашних хозяйств – вариант 2**

Предположим, что размер вашего стандартного сегмента равен 500 человек. Обозначим число домашних хозяйств, составляющих размер кластера, желательный для целей обследования, как  $\bar{n}$ .

1. Подсчитайте среднее число домашних хозяйств в одном сегменте, разделив 500 на средний размер домашнего хозяйства в вашей стране. Выразим его как  $s_n$ .
2. Разделите  $s_n$  на  $\bar{n}$ . Это ваш интервал выборки  $I$  для отбора домашних хозяйств внутри каждого сегмента выборки.  
(Обратите внимание, что если размер вашего стандартного сегмента – не 500, вам, разумеется, следует использовать это значение.)

*Пояснение:*

Предположим, что средний размер домашнего хозяйства в вашей стране составляет 5,5. В этом случае  $s_n$  равно  $500/5,5$  или 90,9. Предположим, вы хотите, чтобы размер кластера  $\bar{n}$  равнялся 25. Разделите 90,9 на 25 (с точностью до одной десятой) =  $90,9/25$  или 3,6. Затем выберите домашние хозяйства в каждом сегменте через интервал 3,6, начиная со случайного номера, расположенного между 01 и 36 (вставив десятичную запятую после выбора числа).

### ПРОЦЕДУРЫ ОПРЕДЕЛЕНИЯ ВЫБОРКИ С ПОМОЩЬЮ ВПР – ВАРИАНТ 3

В случае использования варианта 3 – модифицированной схемы сегмента, которая описывается в настоящей главе, – вместо варианта 2, неявная стратификация проводится таким же образом, *хотя измерение размера является иным*. При варианте 3, если мы, к примеру, предположим, что размер нашего сегмента будет равен 20 домашним хозяйствам (в среднем), то измерение размера следует рассчитывать путем деления количества домашних хозяйств, полученного в ходе переписи, на 20 с округлением до ближайшего целого числа. Обратите внимание, что при варианте 3 во втором столбце таблицы 4.14 следует указывать количество домашних хозяйств, а не численность населения. Вам придется определить интервал выборки  $I$ , разделив общую сумму, предположим, 26 425 на желаемое число ПВЕ, равное, допустим, 300. Тем самым вы получите следующий результат:  $26\,425/300 = 88,1$ . Если случайно выбранной отправной точкой будет 19,4, первыми двумя выбранными ПВЕ, как это показано в таблице 4.14, окажутся те, которые будут соответствовать наименьшим общим показателям, превышающим, соответственно, значения 19,4 и 107,5 ( $88,0 + 19,4$ ). Ими являются СУ 002 в общине 01 и СУ 002 в общине 03 провинции 01 округа 01.

**Таблица 4.14**  
**Иллюстрация систематизированного определения выборки с помощью ВПР**  
**и неявной стратификации – вариант 3 выборки**

Городской район Провинция 01 Округ 01 Община 01	Численность населения	Измерение размера (число сегментов с населением в 500 человек)	Общая сумма
СУ 001	290	14	14
СУ 002	120	6	Выбрано 20
СУ 003	325	16	36
СУ 004	200	10	46
Община 02			
СУ 001	81	4	50
СУ 002	307	15	65
СУ 003	261	13	78
СУ 004	80	4	82
Община 03			
СУ 001	308	15	97
СУ 002	236	12	Выбрано 109
СУ 003	205	10	119
*			
*			
*			
и т. д.			
Сельский район Провинция 12 Округ 05 Община 05			
СУ 001	102	5	26 400
СУ 002	99	5	26 405

Помните, что при варианте 3 измерение размера равно числу сегментов заранее оговоренного размера, которые должны быть созданы (в нашем примере – 20). Таким образом, для отобранных выборочных ПВЕ в первой ПВЕ следует создать 6 сегментов с приблизительным размером 20 домашних хозяйств каждый, а во втором – 12 таких сегментов. Опять-таки в данном случае один из сегментов следует затем случайно выбрать из каждой ПВЕ выборки, и для целей обследования провести опрос всех домашних хозяйств, относящихся к этому сегменту, даже если фактическое число домашних хозяйств в сегменте будет отличаться от его ожидаемого размера.

Процедуры создания сегментов для варианта 2 и варианта 3 подробно описываются в главе 6.

## СТРАНОВЫЕ ПРИМЕРЫ ИЗ РАУНДА МИКС, ПРОВЕДЕННОГО В 2000 ГОДУ

В заключительном подразделе настоящей главы мы рассмотрим схемы выборки, которые использовались в трех странах, участвовавших в раунде МИКС в 2000 году. Каждый пример иллюстрирует отдельный аспект процесса составления выборки, о которых рассказывалось в настоящей главе. Схемы выборки показаны на примере Ливана, Папуа-Новой Гвинеи и Анголы.

### ЛИВАН

Мы начнем с Ливана и дадим достаточно подробное описание ситуации, для того чтобы можно было провести полное сравнение схемы выборки с Руководством МИКС за 2000 год. Кроме того, мы продемонстрируем, каким образом были рассчитаны веса для иллюстрации осуществления этого процесса в рамках конкретного обследования.

#### *План выборки*

Обследование МИКС проводилось силами Центрального управления статистики (ЦУС). Главные цели обследования заключались в том, чтобы: 1) собрать достоверные данные по показателям, установленным на Всемирной встрече на высшем уровне в интересах детей, по состоянию на конец десятилетия, с тем чтобы составить представление о положении женщин и детей в Ливане; и 2) составить такие оценки на национальном уровне и по пяти субнациональным районам (территориальным единицам), которые включают в себя четыре крупнейших губернаторства (мухафазы), а также территорию Южного Ливана и Эн-Набатии. Методология определения выборки для МИКС в Ливане предусматривала составление вероятностной выборки из 1625 домашних хозяйств в каждой из 5 территориальных единиц с использованием стратифицированной кластерной схемы. Применялся двухэтапный процесс отбора, в котором в качестве основы выборки использовались результаты переписи зданий 1995/96 годов. В общей сложности для целей обследования по всей стране было отобрано 8125 домашних хозяйств.

Кластеры, отобранные на первом этапе (т. е. первичные выборочные единицы – ПВЕ), были обозначены как "илоты", которые по своей сути и структуре идентичны счетным участкам переписи. Они были выбраны со степенью вероятности, пропорциональной измерениям их размера (или ВПР), которые соответствуют показателям заселенности зданий, выявленных в ходе переписи зданий, проведенной в 1995/96 году. В каждой территориальной единице было отобрано по 65 илотов, что в общей сложности составляет 325 илотов. Илоты, отобранные в рамках первого этапа, должны были быть перечислены заново, т. е. следовало составить новый список домашних хозяйств, подготовленный для работы на местах. Очень крупные илоты были разделены на географические сегменты, один из которых произвольно выбирался для включения в список. Этот новый список являлся основой выборки для второго этапа отбора.

Из новых списков в каждом выборочном илоте для целей проведения опроса МИКС систематически выбирался фиксированный размер выборки – 25 домашних хозяйств. В некоторых выборочных илотах, в частности в тех, в составе которых насчитывалось менее 25 домашних хозяйств, обследование МИКС проводилось среди всех домашних хозяйств.

Предполагалось, что доля не ответивших на вопросы, задаваемые в ходе обследования, будет составлять порядка 10–12 процентов, в силу чего число опрошенных домашних хозяйств будет колебаться от 1430 до 1465 в каждой территориальной единице или от 7150 до 7325 по всей территории страны.

### ***Обсуждение плана выборки – концепция МИКС***

Важно отметить, что описываемый выше план выборки в целом вполне соответствует общей концепции МИКС в отношении схемы выборки и порядка ее определения, изложенной в предыдущем издании Руководства по проведению МИКС. В указанном издании, как и в нынешнем, особое внимание уделяется некоторым чертам схемы выборки, и все из них находят отражение в плане ЦУС. К их числу относятся следующие:

- применение простой методики определения выборки;
- использование данных последней переписи в качестве основы выборки;
- реализация выборки в два этапа;
- использование вероятностных методов на всех стадиях отбора;
- использование кластеров среднего размера;
- использование надлежащего размера выборки.

Методика определения выборки является простой и целенаправленной, в отличие от сложной. Благодаря этому ожидалось, что ошибки, не связанные с определением выборки, будут сведены к минимуму. Перепись зданий 1995/96 года считалась надлежащей основой для использования в качестве рамок выборки, так как а) она обеспечивала последние по времени данные и б) в процессе этой переписи были составлены карты, и поэтому илоты, отобранные для МИКС, было легко установить, а их внутренние границы – легко определить. Данный картографический аспект также весьма удобен для сегментации, которую необходимо проводить в случае крупных илотов. Возраст основы выборки означал, что естественное движение населения и изменения в нем могут снизить точность оценок, и по этой причине на втором этапе было необходимо составить новый список домашних хозяйств, с тем чтобы обновить основу выборки в выбранных районах.

На обоих этапах применялись жесткие методы вероятностного отбора. Это помогло обеспечить, чтобы полученные результаты были представительными и отражали различные целевые группы населения, включая домашние хозяйства, женщин фертильного возраста и детей в Ливане. Размер кластера в 25 домашних хозяйств является средним, и это способствовало тому, что эффект построения выборки был не настолько значительным, чтобы сделать результаты по важным показателям недостоверными.

Размер национальной выборки в 8125 домашних хозяйств соответствовал рекомендациям, изложенным в руководстве МИКС 2000 года, в отношении этого чрезвычайно важного параметра обследований. Как указывалось в данном Руководстве, размер выборки должен быть достаточно крупным, для того чтобы можно было решить различные задачи, связанные с измерениями. К их числу относится способность измерить ключевые целевые показатели с пределом ошибки в 3–5 процентных пункта в зависимости от показателя. Вторая важная задача заключалась в том,

чтобы обеспечить возможность получения достаточно точных оценок показателей по территориальным единицам, представляющим особый интерес: как указывалось в резюме, этими территориальными единицами являлись четыре крупных губернаторства – Бейрут, Горный Ливан, Бекаа и Северный Ливан, а также Южный Ливан/Эн-Набатия и, разумеется, вся страна в целом.

Как и в настоящем издании Руководства МИКС, в Руководстве МИКС 2000 года предполагалась возможность корректировать размер выборки в сторону увеличения, для того чтобы учесть долю отказывающихся отвечать. Этот фактор был должным образом принят во внимание.

### *Размер выборки*

Показателем, который был выбран для определения размера выборки в Ливане, являлась доля детей в возрасте до пяти лет, страдавших диареей в течение последних двух недель. Ожидалось, что их доля будет составлять порядка 15 процентов с пределами погрешности  $\pm 3$  процента. В результате доверительный интервал, который находится на уровне примерно 95 процентов, для целей обследования должен был дать значение в диапазоне от 12 до 18 процентов, если подлинная доля такого контингента равнялась 15 процентам.

Целевой группой для определения показателя заболевания диареей являлись дети в возрасте до пяти лет, и они составляли немногим менее 10 процентов от общей численности населения. Предполагалось, что эффект построения выборки составлял около 1,5 процента. Это был множитель, на который был увеличен размер выборки для того, чтобы уравновесить тот факт, что вместо простой случайной выборки использовалась кластерная схема, выбор которой оказывает определенное влияние на достоверность выборки.

Как уже указывалось ранее, обследование проводилось в 5 территориальных единицах. Предполагалось, что доля отказавшихся ответить должна была составлять около 12,5 процента. Эта оценка является консервативной, и должна привести к увеличению размера выборки сверх необходимого. Высказывались надежды, что фактическое число ответивших будет гораздо большим, чем запланированный показатель в 87,5 процента. Средний размер домашнего хозяйства в Ливане на тот момент составлял 4,8. Коэффициент, необходимый для достижения доверительного уровня в 95 процентов, равен 4.

Принимая во внимание все перечисленные выше параметры, количество домашних хозяйств, которое необходимо включить в выборку по каждой территориальной единице, можно установить с использованием следующей формулы расчета размера выборки (см. Дополнение 7 предыдущего издания Руководства). Ее пришлось изменить для пяти территориальных единиц, как описано ниже.

$$n = \frac{4r(1-r)(1,125)f}{e^2(p)(n_h)} = \frac{4(,15)(,85)(1,125)(1,5)}{(.03)(,03)(,1)(4,8)} = 1992$$

Для того чтобы скорректировать размер выборки для территориальных единиц, в руководстве (2000 года издания) предложены два альтернативных варианта. Первый состоит в том, чтобы умножить размер выборки  $n$ , необходимый для одной территориальной единицы, на число территориальных единиц. Умножаем 1992 на 5 и получаем 9961. Однако поскольку бюджет обследования не в состоянии выдержать столь большой размер выборки, можно воспользоваться другим подходом, который характеризуется несколько более высокими пределами ошибки и предусматривает умножение размера выборки  $n$  на коэффициент  $D$ , увеличенный до порядка 0,65, где  $D$  – число территориальных единиц. Это приблизительно тождественно извлечению кубического корня из  $D$ , возведенного в квадрат, что в нашем случае составляет кубический корень из 25 или около 2,92. Это число, умноженное на 1992, даст размер выборки по 5 территориальным единицам в объеме 5817 домашних хозяйств. Было принято решение использовать вариант, являющийся промежуточным между этими двумя альтернативными вариантами, т. е. взять в качестве основы выборки 7889 домашних хозяйств. Далее эта цифра была скорректирована в сторону увеличения, для того чтобы на каждую территориальную единицу приходилось 65 кластеров (для 7889 домашних хозяйств потребовалось бы около 63 кластеров, т. е.  $7889/25 = 315$  или 316 хозяйств, а  $315/5 = 63$  кластера в каждой территориальной единице).

Следует также обратить внимание на эффект предела ошибки при данном размере выборки. Расчетное значение  $n$  для диареи составляет 1992. Выше говорилось, что при этом значении предел ошибки будет составлять  $\pm 3$  процентных пункта. Вместе с тем размер окончательной выборки равен  $8125/5$  или 1625 на одну территориальную единицу. Таким образом, предел ошибки несколько выше. Увеличение рассчитывается как квадратный корень соотношения размеров выборки или квадратный корень  $1992/1625$ , что составляет около 1,033. Таким образом, ожидаемый предел ошибки для этого конкретного показателя на уровне провинции будет равняться 3,3 процентных пункта, если распространенность диареи будет составлять приблизительно 15 процентов.

### ***Распределение выборки***

При принятии решения о том, как следует правильно распределять выборку, необходимо учитывать, какие цели преследуют измерения. Когда приоритет отдается национальным оценкам, выборка должна распределяться между различными губернаторствами пропорционально. Например, если в губернаторстве проживают 35 процентов населения страны, то 35 процентов выборки следует отбирать именно из этого губернаторства в соответствии со схемой пропорционального распределения. Однако проблема данного подхода состоит в том, что малочисленные губернаторства (например, в которых проживает лишь 10 процентов населения) будут иметь слишком небольшой размер выборки для обеспечения достоверности оценок. Эта проблема усугубляется, если приоритет отдается измерениям на уровне губернаторств, а не на общенациональном уровне.

По сути дела, главным объектом внимания правительства и Отделения ЮНИСЕФ в Ливане является уровень губернаторства. Это обусловлено тем, что проекты, связанные с планированием программ и осуществлением политики и направленные на улучшение положения женщин и детей, должны реализовываться на местном уровне. Поэтому по каждому из губернаторств необходимо

было получить как можно более достоверные оценки. Для решения этой задачи выборку следует равномерно распределять между губернаторствами (четырьмя крупнейшими плюс объединенная южная единица).

В таблице 4.15 показано распределение выборки и другие ее параметры по губернаторствам.

**Таблица 4.15**  
**Параметры выборки – МИКС, Ливан, 2000 год**

Мухафаза (губернаторство)	Размер выборки – количество домашних хозяйств (КДХ)	Размер выборки – количество кластеров (илотов)	Размер кластера – количество КДХ выборки на 1 кластер	Общее число единиц в рамке (заселенных жилых помещений)	Интервал выборки для выбора ВПР илотов
Бейрут	1 625	65	25	101 707	1 в 1564,7
Горный Ливан	1 625	65	25	356 517	1 в 5484,9
Северный Ливан	1 625	65	25	150 018	1 в 2308,0
Бекаа	1 625	65	25	97 638	1 в 1502,1
Южный Ливан/ Эн-Набатия	1 625	65	25	149 093	1 в 2293,7
Всего по Ливану	8 125	325	25	854 973	–

#### ***Осуществление – процедуры определения выборки***

Первый этап отбора был связан с составлением систематизированной выборки выявленных в ходе переписи илотов с использованием ВПР, при котором за единицу размера бралось количество заселенных жилищ. До начала процедуры отбора илоты были отсортированы по губернаторствам и в рамках губернаторств и внутри провинций по округам (казам), большим зонам и илотам, причем илоты были распределены по географически последовательным рядам. Все эти операции производились с использованием семизначного географического идентификационного кода, фигурирующего в географической базе данных. Таким образом проделанный систематический отбор обеспечивал неявную стратификацию выборки по каждой территориальной единице или губернаторству. В результате по каждой территориальной единице с помощью компьютера были отобраны 65 илотов.

В рамках второго этапа отбора на местном уровне был составлен подробный список домашних хозяйств. Многие попавшие в выборку илоты были весьма крупными, и до составления списка их приходилось подвергать сегментации. Из выбранных 325 илотов 180 были подвергнуты сегментации, поскольку, согласно основе выборки 1995/96 года, в каждом было более 150 домашних хозяйств.

Сегментация проводилась в соответствии с таблицей 4.16.

**Таблица 4.16**  
**Количество подлежащих формированию сегментов**

Количество домашних хозяйств (согласно основе выборки)	Количество сегментов
Меньше 150	Без сегментации
150–249	2
250–349	3
350–449	4
450–549	5
550–649	6
650–749	7
И т. п.	И т. п.

Некоторые из отобранных илотов, примерно 12, согласно основе выборки, включали менее 25 домашних хозяйств. Для таких илотов составлять списки на местах не потребовалось. Вместо этого все домашние хозяйства, включенные в границы илота, были опрошены в рамках МИКС, даже если количество таких домашних хозяйств на момент проведения МИКС в 2000 году и превысило 25.

Для сегментации использовалась таблица 4.16, в которой показано, какое количество сегментов следует создать. Например, если, согласно основе, количество илотов в выборке составляло 580, создавалось 6 сегментов. При этом использовались процедуры построения сегментов, которые описываются в главе 6 Руководства МИКС (предыдущее издание). После сегментации из каждого илота, в котором она была проведена, произвольно отбирался один сегмент.

В случае наличия илотов, состоящих менее чем из 150 домашних хозяйств, а также отобранных сегментов, состоящих более чем из 150 домашних хозяйств, предпринималась еще одна поездка на места в целях составления полного списка домашних хозяйств, существовавших на то время. Затем выборка домашних хозяйств систематизировалась внутри каждого илота или сегмента с коэффициентом 1 в I, где I – интервал выборки, равный  $N_i + n$ , где  $N_i$  – фактическое количество домашних хозяйств в списке по кластеру  $i$ , а  $n$  равнялось 25, т. е. желательному размеру кластера. Значение интервала выборки I было различно для каждого отобранного илота или сегмента и рассчитывалось с точностью до одной десятой.

### ***Взвешивание и оценка***

Для подготовки оценок в рамках обследования приходилось прибегать к взвешиванию, поскольку выборка не являлась самовзвешивающейся. Другими словами, случаи, включенные в выборку, характеризовались не одинаковой вероятностью отбора. Процедура взвешивания обуславливала необходимость умножения необработанных данных выборки на вес выборки. Веса варьировались в зависимости от кластера, так как вес для домашних хозяйств, включенных в каждый кластер,

зависит от измерения размера и количества домашних хозяйств, определенных при составлении списка домашних хозяйств для данного кластера.

Веса равны *обратным величинам* вероятностей отбора. Вероятность для домашних хозяйств, отдельных лиц, детей, женщин или других целевых групп населения в любом конкретном кластере рассчитывается по следующей формуле:

$$P_{ih} = \frac{(65)(25)m_{ih}}{M_h N_{ih} s_{ih}}, \text{ где}$$

- $P_{ih}$  – вероятность выбора домашнего хозяйства (или отдельного лица) из кластера  $i$  территориальной единицы  $h$ ,
- **65** – количество кластеров в каждой территориальной единице,
- **25** – размер кластера,
- $m_{ih}$  – измерение размера кластера  $i$  территориальной единицы  $h$ , т. е. количество заселенных жилищ, согласно основе выборки,
- $N_{ih}$  – количество домашних хозяйств, занесенных в список по кластеру  $i$  территориальной единицы  $h$ ,
- $s_{ih}$  – количество сегментов, созданных в каждом илоте (для илотов, в которых сегментация не проводилась,  $s_{ih}$  равно 1),
- $M_h$  – общее измерение размера в территориальной единице  $h$ , т. е. сумма значений  $m_{ih}$  или  $\sum m_{ih}$ .

Обратите внимание, что значение  $N_{ih}$  – количество домашних хозяйств, занесенных в список по кластеру. В илотах, где сегментация не проводилась, это – весь илот, а в сегментированных илотах – это его сегментированная часть (см. примеры, ниже).

Как уже отмечалось ранее, вес  $W_{ih}$  является обратной величиной или  $1/P_{ih}$ .

Значения  $M_h$  приводятся в четвертом столбце таблицы 4.15 для каждой из территориальных единиц. Умножив 65 на 25, получаем 1625. Таким образом, веса для домашних хозяйств или отдельных лиц по этим пяти территориальным единицам приведены в таблице 4.17.

**Таблица 4.17**  
**Веса обследования по территориальным единицам**  
**(для илотов, состоящих из 25 или более домашних хозяйств)**

Губернаторство	Вес
Бейрут	$(62,6N_i s_i)/m_i$
Горный Ливан	$(219,4N_i s_i)/m_i$
Северный Ливан	$(92, N_i s_i)/m_i$
Бекаа	$(60,1N_i s_i)/m_i$
Южный Ливан/Эн-Набатия	$(91,7N_i s_i)/m_i$

**ПРИМЕР:**

Предположим, что в Бейруте измерение размера  $m_i$  в основе выборки для первого отобранного кластера выборки равно 612. Далее он дробится на 6 сегментов, что соответствует значению  $s_i$ . Предположим далее, что количество домашних хозяйств, которое приходится на этот кластер, равняется 110. Соответственно, вес для всех домашних хозяйств и отдельных лиц в выборке по данному кластеру составит  $(62,6)(110)(6)/(612)$  или 67,5.

**ПРИМЕР**

Предположим, что в Горном Ливане измерение размера  $m_i$  в основе выборки для первого отобранного кластера выборки равно 106. Ввиду этого он не сегментируется, и значение  $s_i$  равно 1. Предположим далее, что количество домашних хозяйств, которое приходится на этот кластер, равняется 98. Соответственно, вес для всех домашних хозяйств и отдельных лиц в выборке по данному кластеру составит  $(219,4)(98)(1)/(106)$  или 202,8.

Ожидалось, что веса, зависящие от *схемы выборки*, о которых говорилось выше, необходимо будет скорректировать дополнительным множителем для учета отказов от ответа. Для этого вес по каждой территориальной единице необходимо было умножить на коэффициент, отражающий соотношение количества домашних хозяйств, включенных в выборку, и количества опрошенных домашних хозяйств. Произведение веса, связанного со схемой выборки, и коэффициента отказов от ответа становится конечным весом, на который следует умножать включенное в выборку предварительно посчитанное число домашних хозяйств для получения оценок.

**ПАПУА-НОВАЯ ГВИНЕЯ**

Обследование МИКС в Папуа-Новой Гвинее проводилось силами Национального статистического управления (НСУ). Его схема иллюстрирует аспекты, связанные с обновлением и охватом основы выборки, а также выбором между существующей выборкой обследования (ОНЗ) и составлением новой выборки. Порядок расчета весов для обследования МИКС в Папуа-Новой Гвинее также носит информативный характер, но поскольку в некоторых отношениях он аналогичен тому, который применялся в Ливане, он не будет рассматриваться в настоящем разделе.

***Краткое описание схемы выборки***

Выборка составлялась в ходе двухэтапного процесса отбора с использованием в качестве основы выборки частично обновленных данных переписи населения 1991 года. В качестве первичных выборочных единиц (ПВЕ) были выбраны счетные участки или единицы переписи (ЕП). В общей сложности было отобрано 175 ПВЕ со степенью вероятности, пропорциональной измерениям их размера (ВПР), которые соответствовали подсчету домашних хозяйств в сельских районах в 1991 году и обновленным результатам подсчетов жилищ в городских районах. Из ЕП, выбранных на первом этапе, был составлен новый список; другими словами, для работы на местах был подготовлен новый список домашних хозяйств. Новый список представлял собой основу выборки для второго этапа отбора. На основе этих списков в каждой выборочной ЕП для целей МИКС

систематически выбирался фиксированный размер выборки, состоящий из 25 домашних хозяйств. В общей сложности целевая выборка включала в себя 4375 домашних хозяйств.

Число кластеров выборки было рассчитано путем деления 4375 на 25 и составило 175. Около 43 процентов кластеров были распределены в городские районы (75 кластеров), а 57 процентов – в сельские районы (100 кластеров). До начала отбора на первом этапе основа была подвергнута неявной стратификации путем разделения данных переписи населения на городские и сельские районы на первом уровне и по регионам, провинциям, округам, подокругам и ЕП на последующих уровнях.

### ***Изменение основы выборки и использование данных ОНЗ в качестве альтернативы***

Последней по времени переписью, которую можно было бы использовать в качестве основы для составления выборки в процессе проведения обследования МИКС в Папуа-Новой Гвинее, являлась перепись населения, которая была проведена в 1991 году, т. е. 9 лет назад. Соответственно, ее данные в различной степени устарели, в первую очередь в городских районах и, прежде всего, в Порт-Морсби. Во время проведения МИКС НСУ также занималось подготовкой к проведению в 2000 году переписи населения и в процессе переписи собиралось использовать те же административные единицы, ЕП, которые использовались в ходе переписи 1991 года. Велась работа по обновлению этих ЕП путем составления нового списка домашних хозяйств. К моменту начала МИКС обновление ограничивалось только городскими районами, которые для целей МИКС являлись самым важным сектором в отношении использования данных в качестве основы выборки.

Альтернативы заключались в том, чтобы: 1) отложить проведение обследования МИКС до тех пор, пока работа по обновлению в рамках переписи не будет завершена в сельских районах и 2) использовать данные предыдущей переписи в качестве основы выборки для сельских районов и обновленные данные – для городских районов. Ключевая задача для любой основы состоит в том, чтобы получить измерения размера, т. е. данные о количестве домашних хозяйств, которые можно использовать для определения коэффициента отбора и которые обладали бы *относительно высокой степенью* точности. Однако при этом важно помнить о том, что измерения размера не обязательно должны быть точными на 100 процентов, для того чтобы их можно было эффективно использовать для составления выборки на первом этапе. Поэтому предполагалось, что количество домашних хозяйств в сельских ЕП, определенное в ходе предыдущей переписи, будет отличаться от показателей 2000 года лишь на несколько процентных пунктов, и, тем самым, данные останутся достаточно точными. С другой стороны, показатели для городских ЕП, вероятно, будут существенно иными и могут в 2–3 раза и даже более отличаться от показателей, определенных девять лет назад. Это явление объясняется тем, что, хотя из каждой сельской ЕП мигрирует относительно небольшое число людей, все они мигрируют в одни и те же городские ЕП, количество которых сравнительно невелико. Поэтому гораздо важнее иметь обновленные данные о городских, нежели о сельских ЕП.

В связи с этим было принято решение использовать для выборки МИКС две основы: для сельских ЕП – перепись населения 1991 года, а для городских ЕП – обновленные списки ЕП с пересмотренными данными о жилищах, которые были подготовлены для проведения переписи населения в 2000 году.

Была также рассмотрена возможность использования в качестве альтернативной основы данных обследования в области народонаселения и здравоохранения (ОНЗ) 1996 года как альтернативы переписи населения 1991 года. Это обследование преследовало цели, аналогичные целям МИКС, причем вошедшие в выборку ПВЕ в 1996 году были обновлены. Благодаря этому данные о сельских ПВЕ в значительно большей степени соответствовали текущему моменту в отношении измерений размера, чем показатели переписи 1991 года. Вместе с тем задачи ОНЗ, связанные с определением показателей, заключались в получении данных на национальном уровне, и это означало, что выборочные ПВЕ распределялись пропорционально по всей территории страны. Ввиду этого количество городских ПВЕ, данные о которых были получены в рамках ОНЗ, было слишком мало, для того чтобы его можно было использовать в МИКС, важная цель которого состояла в том, чтобы получить достоверные данные о городском секторе как ключевой территориальной единице оценки. В связи с этим ОНЗ далее не рассматривалось в качестве альтернативы.

### ***Охват выборки***

Некоторые, преимущественно внутренние районы Папуа-Новой Гвинеи являются труднодоступными, и в них можно попасть только на вертолете. По мнению правительства, охват населения этих территорий переписью был бы сопряжен с чересчур высокими затратами. Поэтому правительство приняло решение не охватывать эти труднодоступные территории МИКС, хотя и признавалось, что состояние здоровья женщин и детей, проживающих в этих районах, может существенно отличаться от состояния здоровья жителей остальной части страны.

Для обеспечения того, чтобы обследование проводилось по правильной методике, важно было изъять исключенные районы из основы выборки до начала составления выборки. Благодаря этому полученная выборка по-прежнему будет являться вероятностной выборкой *охваченного* контингента, хотя этот контингент по своей численности и будет меньше всего населения страны. Альтернатива, связанная с первоначальным отбором районов выборки и последующей заменой в случае выбора исключенного из обследования района, считается необъективной в статистическом отношении процедурой и поэтому не использовалась. В результате был составлен список труднодоступных округов и подокругов, которые должны были быть исключены из обследования, и все ЕП, которые пришлось на эти районы, были исключены из рамки выборки.

### ***Размер и распределение выборки***

Правительство надеялось провести МИКС на основе выборки, включающей около 2500 домашних хозяйств. В руководстве МИКС 2000 года странам настоятельно рекомендуется найти такой размер выборки, который даст им возможность измерить наиболее важные показатели с пределом ошибки в 3–5 процентных пунктов (при 95-процентном доверительном интервале). Ожидалось, что на национальном уровне 2500 домашних хозяйств будет вполне достаточно, для того чтобы выполнить эти требования. Однако правительство было также заинтересовано в том, чтобы получить отдельные достоверные оценки по некоторым территориальным единицам или субнациональным районам, которые представляют большое значение для планирования политики. Этими единицами являются городские и сельские районы, а также четыре региона – Южный, Северный, Горный и Островной.

Предполагалось, в частности, что для региональных оценок размер выборки в 2500 домашних хозяйств окажется недостаточным, поскольку на каждый регион в этом случае приходилась бы выборка, состоящая всего лишь, в среднем, из 625 домашних хозяйств, что не достаточно для того, чтобы получить достоверные оценки по ключевым показателям. Кроме того, распределение населения между городскими и сельскими районами предполагает необходимость использования непропорциональной выборки, для того чтобы обеспечить достаточный размер выборки в городских районах. Это обусловлено тем, что в городских районах Папуа-Новой Гвинеи проживает лишь около 15 процентов населения страны. Поэтому, если бы выборка отбиралась пропорционально, в ней оказалось бы менее 400 городских домашних хозяйств лишь в 15 ПВЕ, а эти цифры слишком малы, для того чтобы обеспечить сколько-либо достоверные результаты.

Все эти соображения обуславливали необходимость следующих двух шагов: 1) для городского сектора необходимо завысить выборку и 2) необходимо увеличить размер выборки в целом, для того чтобы получить достоверные региональные оценки.

Такой аспект, как количество кластеров, т. е. единиц, отбираемых на первом этапе (ПВЕ), также следует учитывать в процессе составления оценок по этим территориальным единицам. При наличии 100 ПВЕ среднее их число для отдельного региона будет составлять всего лишь 25. Это число слишком мало, для того чтобы обеспечить надлежащее "распространение" выборки на региональном уровне. Для оценок по каждой территориальной единице необходимо иметь около 50 ПВЕ, а абсолютный минимум должен составлять не менее 30. Соответственно, было принято решение увеличить общий размер выборки с 2500 до 4000 домашних хозяйств. Помимо этого, было принято решение сделать 10-процентную поправку на число отказавшихся ответить и за счет этого увеличить размер выборки примерно до 4375 домашних хозяйств. В результате количество кластеров возросло со 100 до 175 ( $4\ 375 \div 25$ ).

И, наконец, было принято решение завысить размер городской выборки примерно в три раза, т. е. увеличить долю городского контингента в общей выборке с 15 до 43 процентов. Таким образом, в городской выборке насчитывалось бы 1875 домашних хозяйств в 75 ПВЕ, а в сельской выборке – 2500 домашних хозяйств в 100 ПВЕ. Завышение размеров выборки в городских районах преследовало цель получения достаточного контингента для составления достоверных оценок по городским районам. Однако даже несмотря на непропорциональные размеры выборки в городских районах, на долю сельских районов по-прежнему приходилось около 2500 домашних хозяйств, т. е. больше, чем требуется для получения достоверных оценок по этим районам.

### ***Осуществление – процедуры определения выборки***

Процедуры определения выборки, которые использовались в Папуа-Новой Гвинее, в целом весьма напоминают те, которые использовались в Ливане, однако между ними имеется одно существенное отличие. Поэтому мы остановимся только на этом отличии. Согласно Руководству МИКС (оба издания), можно воспользоваться одним из двух вариантов при отборе домашних хозяйств на втором (или третьем) этапе. Первый состоит в том, чтобы составить новый список домашних хозяйств в кластерах выборки и выбирать домашние хозяйства из новых списков.

Как указывается в Руководстве, это – предпочтительный для МИКС подход, поскольку он является несколько более надежным и несколько менее сложным по сравнению с другим вариантом.

Вместе с тем альтернативой является подход, позволяющий избежать затрат, связанных с составлением списков, которые могут быть довольно значительными, и именно поэтому в Руководстве он указывается как адекватная альтернатива. Этот подход предусматривает построение географических сегментов приблизительно равного размера в каждом кластере выборки (ЕП для Папуа-Новой Гвинеи), произвольный выбор одного из них и проведение опроса всех домашних хозяйств, входящих в выбранный сегмент для МИКС. Однако, как подразумевалось выше, эта альтернатива, хотя и реалистичная, не является столь же надежной, как вариант с составлением новых списков, поскольку домашние хозяйства в выборке "группируются в кластер", и эффект построения выборки будет в этом случае несколько более ощутимым. Во-вторых, сегментация сложнее, чем составление списков, и поэтому может дать более высокую не связанную с выборкой ошибку, особенно если сотрудники не имеют должного опыта использования географических карт, в том числе составления схем сегментов. НСУ решило выбрать оптимальный подход – составление списков, а не сегментацию.

## **АНГОЛА**

Обследование МИКС проводилось Национальным институтом статистики (НИС). Схема выборки в рамках обследования МИКС в Анголе свидетельствует о том, что обследование проводилось в два этапа, осуществляясь с интервалом в несколько месяцев, а также позволяет понять связанные с этим вопросы, касающиеся охвата населения, недостатков основы и представления результатов пользователям.

### ***Краткое описание схемы выборки***

Основные цели измерений, которые проводились в контексте МИКС, заключались в том, чтобы: 1) получить соответствующие показатели по состоянию на конец десятилетия по программе, принятой на Всемирной встрече на высшем уровне в интересах детей, для того чтобы иметь представление о положении женщин и детей в Анголе; и 2) получить такие оценки на национальном уровне и по трем экологическим зонам (территориальным единицам) – прибрежной, равнинной и внутренней частей страны.

В качестве методологии составления выборки для обследования МИКС, которое проводилось в Анголе в 2000 году, использовались процедуры вероятностного отбора, позволившие выбрать около 6660 домашних хозяйств в рамках стратифицированной кластерной схемы. Обследование проводилось в два этапа, на первом из которых использовалась основа выборки, составленная для Обследования доходов и расходов 2000–2001 годов (ОДР). На втором этапе, который начался примерно 2–3 месяца спустя, обследование проводилось в масштабах, соответствовавших МИКС, т. е. на территории, которая не была охвачена обследованием ОДР. Начало второго этапа было задержано, поскольку для него сначала потребовалось разработать материалы, необходимые для определения основы выборки.

### *Составление выборки*

Составление выборки проходило в два этапа в городских районах и в три этапа – в сельских районах. В качестве единиц первого этапа [т. е. первичных выборочных единиц (ПВЕ); на португальском они называются *Unidades Primarias* (основные единицы или ОЕ)] для городских районов были выбраны городские кварталы, состоящие из конкретных географических сегментов (обычно городских кварталов), в состав которых в среднем входит около 100 домашних хозяйств или семейных конклавов. Для сельских районов кластерами служат общины (коммуны), вторичными единицами являются деревни. Таким образом, кластеры МИКС сами являются единицами первого этапа в городских районах и при этом выступают в качестве единиц второго этапа в сельских районах.

Кластеры выбирались в так называемой *охваченной* территории страны. Она определяется как доступная территория страны, находящаяся под контролем правительства, и в целом состоит из провинциальных и/или муниципальных образований, а также некоторых сельских районов. Охват выборки обеспечивался путем первоначального изъятия из рамок выборки тех районов, которые следовало исключить из обследования ввиду проблем внутренней безопасности и/или доступа, включая районы, считающиеся заминированными. Помимо этого основное обследование не охватывало перемещенных внутри страны лиц (ПВЛ), проживавших в лагерях, хотя и был разработан план составления выборки и проведения опроса среди этого контингента в качестве самостоятельной операции.

Выбор единиц выборки для первого этапа осуществлялся с вероятностью, пропорциональной измерениям их размера (ВПР). Измерения размера (ПР) соответствовали насчитывавшемуся на тот момент количеству домов или других жилищ в городских районах и текущим оценкам численности сельского населения в сельских районах. Данные о количестве домов были получены в результате их фактической переписи на местах, в то время как оценки составлялись на основе местной информации, которая предоставлялась на уровне провинций.

В общей сложности было отобрано 408 кластеров: 300 – по итогам первого этапа с использованием основы ОДР и 108 – на базе остаточной основы, составленной в ходе второго этапа. На этих двух этапах применялись различные процедуры составления выборки. На первом этапе кластеры МИКС представляли собой систематизированный поднабор кластеров, уже отобранных для ОДР. Эти кластеры уже были сведены в новый перечень, т. е. в рамках операции ОДР на местах был подготовлен новый список домашних хозяйств. Этот перечень составил основу выборки для последнего этапа отбора. Для каждого из 300 кластеров МИКС, отобранных на первом этапе, была составлена систематизированная выборка из 15 домашних хозяйств, никак не связанная с теми домашними хозяйствами, которые были отобраны для ОДР, во избежание дублирования между этими двумя обследованиями. Для 108 кластеров, отобранных на втором этапе, было отобрано по 20 домашних хозяйств на кластер<sup>23</sup>.

---

<sup>23</sup> Размеры кластеров первого и второго этапов отличаются друг от друга, поскольку соотношение в 20 домашних хозяйств на каждый кластер соответствует рекомендованному для МИКС соотношению (см. руководство МИКС 2000 года), в то время как цифру 15 было удобнее с оперативной точки зрения выбрать из основы ОДР.

На втором этапе основу выборки следовало составить так же, как она была составлена для ОДР, за исключением того, разумеется, что ее следовало также распространить на те провинции и территории МИКС, которые не были охвачены ОДР. Таким образом основа, составленная на втором этапе, являлась источником для отбора 108 кластеров. Эти кластеры были отобраны со степенью вероятности, пропорциональной их размеру. Затем отобранные кластеры были сведены в список на месте, с тем чтобы на втором этапе можно было выбрать по 20 домашних хозяйств на каждый кластер; это обеспечивалось за счет систематизированного отбора.

Выборка, состоящая из 6660 домашних хозяйств, была пропорционально распределена среди трех упомянутых выше экологических зон или территориальных единиц. Каждая из этих территориальных единиц была достаточно крупной по показателю численности проживающего в ней населения, и это обеспечивало возможность получения достоверных результатов на уровне таких единиц путем применения процедуры пропорционального распределения выборки. Ожидалось, что доля отказавшихся отвечать будет составлять около 10 процентов, в результате чего количество фактически опрошенных домашних хозяйств в рамках обследования составляло бы около 6000.

Точно оценить долю населения, которую следовало охватить обследованием, не представлялось возможным из-за отсутствия актуальных демографических данных. Однако предполагалось, что в лагерях для лиц, перемещенных внутри страны, проживало около 10 процентов населения, а в недоступных районах – еще примерно 15–25 процентов. Таким образом, согласно оценкам, охват МИКС должен был составлять порядка 65–75 процентов населения страны (см. следующий раздел).

Рекомендуем читателям ознакомиться с таблицей 4.18, где содержится краткое описание параметров плана выборки.

### ***Основа и охват выборки***

Идеальные основы выборки на практике встречаются крайне редко, а в Анголе выбор основы, которая удовлетворяла бы сразу трем критериям – полнота, точность и актуальность – был сопряжен с особыми трудностями. Это обусловлено рядом причин. Одна из них заключается в том, что традиционной основой, используемой для целей национального выборочного обследования, является последняя по времени перепись, которой в случае Анголы была бы перепись населения 1970 года, данные которой безнадежно устарели. Поэтому необходимо было использовать другие, более современные основы, которые изначально имели бы менее высокое качество, чем основа проведенной недавно переписи (дальнейшую информацию см. ниже).

Во-вторых, как уже упоминалось ранее, не были охвачены заминированные районы, а также лагеря лиц, перемещенных внутри страны, и другие труднодоступные районы. Это нанесло критерию полноты ущерб, степень которого осталась неустановленной, хотя и предполагалось, что национальный охват составит около 65–75 процентов, как упоминалось выше. В-третьих, имеющиеся основы отличались степенью неточности неизвестного масштаба, особенно в сельских районах, где качество оценок численности населения (использовавшихся для определения вероятности отбора в ходе первого этапа) считалось колеблющимся в значительных пределах в различных провинциях.

Для МИКС 2000, проведенного в Анголе, это означало, что, *если конечные результаты обследования будут представлены как отражающие ситуацию во всей стране*, они будут носить предвзятый в неизвестной степени характер. Это обусловлено тем, что в отношении подгрупп населения, исключенных из обследования, определить степень представительности в отношении статистики невозможно, поскольку такие подгруппы, скорее всего, будут обладать несколько иными характеристиками, чем подгруппы, включенные в обследование. Еще один важный аспект состоит в том, что по итогам МИКС нельзя получить оценки, касающиеся всего населения страны, из-за того что из обследования были исключены некоторые подгруппы населения.

Тем не менее в результате опросов населения, *охваченного* обследованием МИКС, были получены действительные и полезные результаты благодаря применению продуманных методов вероятностного отбора домашних хозяйств из использовавшихся основ, хотя эти рамки не всегда были идеальными. В отношении охвата обследования основа, разработанная НИС для ОДР, сама по себе была идеальной. Она являлась действительным и надежным источником для использования в ходе первого этапа. Проблема, связанная с этой основой, состоит в том, что она охватывает лишь 7 провинций, а внутри этих провинций, как правило, охватывает лишь территорию городов. Именно поэтому для второго этапа было необходимо разработать другую основу, которая охватывала бы еще 11 провинций, а также всю территорию сельских районов, которая могла бы считаться доступной для целей проведения обследования.

На примере МИКС, проводившегося в Анголе, можно сделать один важный вывод, касающийся обнародования данных обследования. При обнародовании результатов обследования в докладах в них всегда следует, по мере возможности, давать точное и четкое определение доли населения, охваченной обследованием. В случае Анголы было важно уведомить пользователей, что в ходе проведения МИКС были получены достоверные с научной точки зрения результаты примерно по 65–75 процентам населения Анголы и что эти результаты не относятся к лицам, проживающим на территориях, не подконтрольных правительству; лицам, живущим вблизи некоторых заминированных территорий, исключенных из обследования; а также лицам, перемещенным внутри страны и проживающим в лагерях.