



Global Adult Tobacco Survey (GATS)



**Обеспечение качества:
рекомендации и документация**

**Глобальный опрос взрослого населения
о потреблении табака (GATS)
Обеспечение качества:
рекомендации и документация**

Сентябрь 2020 г.

**Глобальный опрос
взрослого населения
о потреблении табака (GATS)
Комплексный стандартный протокол**

План выборки обследования GATS

Основной вопросник с дополнительными вопросами
Повопросные разъяснения

Проведение выездных работ

Руководство по составлению плана выборки
Руководство по определению удельного веса элементов выборки

Проведение выездных работ в связи с обследованием GATS

Руководство для выездного интервьюера
Руководство для полевого супервайзера
Руководство по составлению карт и перечней

Работа с данными обследования GATS

Руководство по программированию комплекса "General Survey System"
Технические параметры для программирования основного вопросника
План проведения работы с данными
Рекомендации по проведению инструктажа

**Обеспечение качества обследования GATS:
рекомендации и документация**

**Пакет материалов для проведения анализа
и подготовки отчетности по обследованию GATS**

Шаблон информационного бюллетеня
Страновой отчет: план и руководство по составлению таблиц
Определения показателей

Выпуск и распространение данных обследования GATS Политика

опубликования данных
Распространение данных: рекомендации
по первоначальному опубликованию данных

Пример оформления библиографической ссылки для цитирования

Сотрудничающая группа по проведению Глобального обследования потребления табака взрослым населением. Глобальный опрос взрослого населения о потреблении табака (GATS). Обеспечение качества: рекомендации и документация. Атланта, Джорджия: Центры США по контролю и профилактике заболеваний, 2020.

Выражение признательности

Организации, сотрудничающие в сфере проведения обследования GATS

- Центры по контролю и профилактике заболеваний Соединенных Штатов Америки (CDC);
- Фонд CDC;
- Школа общественного здравоохранения Блумберга при Университете Джона Хопкинса;
- организация "RTI International";
- Всемирная организация здравоохранения (ВОЗ).

Финансовая поддержка

Финансовая поддержка оказана Инициативой Блумберга по сокращению потребления табака по линии фонда CDC и при содействии гранта программы "Bloomberg Philanthropies".

Уведомление об ограничении ответственности: мнения, выраженные в данном руководстве, могут не совпадать с мнениями организаций, сотрудничающих в сфере проведения обследования GATS.

Содержание

Глава	Страница
1. Введение	1-1
1.1 Обзор Глобального обследования потребления табака взрослым населением	1-2
1.2 Использование данного руководства	1-2
1.3 Обеспечение качества обследования GATS	1-3
2. План процесса по обеспечению качества	2-1
3. Обеспечение качества: этап до сбора данных	3-1
3.1 Процесс разработки и анализа вопросника	3-1
3.2 Программирование и контроль версий	3-2
3.3 Предварительная проверка (апробация)	3-5
3.4 Окончательная доработка вопросника	3-5
3.5 Передача данных и обработка данных	3-6
3.6 Процесс разработки и анализа вопросника	3-7
3.7 Подготовка контрольного файла выборки	3-7
3.8 Обучающие семинары	3-8
4. Обеспечение качества: сбор и работа с данными	4-1
4.1 Сбор данных при проведении выездных работ: настройка и техническое обслуживание	4-1
4.2 Обработка данных в Национальном центре данных	4-7
5. Обеспечение качества: этап после сбора данных	5-1
5.1 Очистка и подготовка данных для расчета удельного веса элементов выборки	5-2
5.2 Показатели качества: выборка, ошибка выборки и удельные веса	5-7
5.3 Показатели качества: охват, доля неответивших и другие ошибки, не связанные с выборкой	5-12
5.4 Официальное рассмотрение качества статистических данных	5-18
5.5 Создание файла аналитических данных	5-21
5.6 Создание файла данных для публичного использования	5-22
6. Требования к данным и документации	6-1
7. Библиография	7-1
Приложение А. Терминологический глоссарий	A-1
Приложение В. Этап до сбора данных	B-1
В.1 Адаптация вопросника GATS и процесс анализа	B-1
В.2 Процесс программирования вопросника GATS	B-2
В.3 Процесс настройки программы "GSS" на портативных устройствах	B-2
Приложение С. Этап после сбора данных	C-1

C.1	Окончательные диспозиционные коды и вычисление доли ответивших лиц	C-1
C.2	Модель калибровки постстратификационных весов с помощью корректировочных ячеек	C-4
C.3	Эффект вариации весов выборки в точности оценок обследования	C-6
C.4	Общий эффект плана выборки в точности оценок обследования и внутрикластерная однородность в пределах ПЕО соответствующих основных оценочных показателей обследования	C-7
C.5	Предел ошибки основных оценок обследования	C-9
C.6	Оценки ошибок выборки	C-12
C.7	Уровень охвата домохозяйств рамкой выборки	C-20

Иллюстрации

Номер	Страница
Иллюстрация 2-1. Диаграмма процесса обеспечения качества.....	2-1

1. Введение

Употребление табака является основной предотвратимой причиной преждевременной смертности и заболеваемости во всем мире, вместе с тем порядка 1,4 миллиарда людей в возрасте 15 лет и старше употребляют табак¹. Помимо этого, свыше 8 миллионов человек ежегодно умирают от заболеваний, связанных с потреблением табака². При сохранении текущих тенденций к концу этого столетия от потребления табака может погибнуть один миллиард человек, в то же время порядка трех четвертых всех летальных исходов придется на страны с низким и средним уровнем дохода³. Для мониторинга и борьбы с этой эпидемией необходим эффективный и систематический механизм эпиднадзора.

Глобальное обследование потребления табака взрослым населением (GATS), которое входит в структуру Глобальной системы эпиднадзора за потреблением табака (GTSS), является мировым стандартом систематического мониторинга употребления табака среди взрослых и отслеживания ключевых показателей в области борьбы против табака. Обследование GATS является национальным репрезентативным обследованием домохозяйств с участием взрослых в возрасте 15 лет и старше, которое проводится с использованием изученного и утвержденного международными экспертами стандартного основного вопросника, плана выборки, процедур сбора и обработки данных. Обследование GATS предназначено для укрепления потенциала стран в области разработки, осуществления и оценки мер борьбы против табака.

Чтобы максимально увеличить эффективность данных, собранных в результате проведения обследования GATS, было создано несколько руководств. Целью данных руководств является предоставление странам стандартных требований, а также ряда рекомендаций по разработке и проведению обследования на каждом этапе процесса обследования GATS. Кроме того, они разработаны в качестве рекомендательного документа по адаптации элементов протокола GATS с целью обеспечить максимальную пользу применения данных в странах. В интересах согласованности и сопоставимости данных между странами настоятельно рекомендуется соблюдать требования стандартного протокола.

В руководствах по проведению обследования GATS изложены систематические рекомендации, касающиеся подготовки плана и практического осуществления обследования.

¹ Всемирная организация здравоохранения. Доклад ВОЗ о глобальной табачной эпидемии, 2019 г.: предложение помощи в целях прекращения употребления табака. Женева, Швейцария: Всемирная организация здравоохранения; 2019. <https://apps.who.int/iris/handle/10665/326047>

² GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Seattle, WA: Institute for Health Metrics and Evaluation; 2018.

³ Mathers, C.D., and Loncar, D. Projections of Global Mortality and Burden of Disease from 2002 to 2030. *PLoS Medicine*, 2006, 3(11):e442.

1.1 Обзор Глобального обследования потребления табака взрослым населением

Обследование GATS предназначено для проведения оценки взрослого населения на национальном и субнациональном уровне в разных странах. Целевая группа населения охватывает неинституционализированное мужское и женское население в возрасте 15 лет и старше, CDC и указывающих конкретную страну в качестве постоянного места проживания. Выборка всех участников целевой группы населения будет проводиться в домохозяйствах, в которых они проживают постоянно.

В обследовании GATS применяется методика многоступенчатого формирования выборки в группах, выделенных по географическому принципу, с целью определить конкретные домохозяйства, с которыми интервьюерам предстоит установить контакт. Территорию страны условно разделяют на первичные единицы отбора, сегменты в составе этих единиц, а также домохозяйства в составе сегментов. Затем производится случайная выборка домохозяйств для участия в обследовании GATS.

Проводимое в ходе обследования GATS интервью строится на основе двух компонентов: *вопросник для домохозяйств* и *индивидуальный вопросник*. Заполнение данных вопросников проводится с применением электронного устройства сбора данных.

Проводимое в ходе обследования GATS интервью строится на основе двух компонентов: *вопросник для домохозяйств* и *индивидуальный вопросник*. *Вопросник для домохозяйств* (скрининг домохозяйств) и *индивидуальный вопросник* (индивидуальное собеседование) заполняются с применением электронного устройства сбора данных.

Выездные интервьюеры прибывают по каждому адресу в выборке и с использованием *вопросника для домохозяйств* проводят опрос одного взрослого человека, проживающего в настоящее время в домохозяйстве. Цель *вопросника для домохозяйств* состоит в том, чтобы определить, отвечает ли выбранное домохозяйство критериям включения в обследование GATS, и составить список или реестр всех членов домохозяйства, которые могут участвовать в обследовании. После формирования списка членов домохозяйства, которые могут участвовать в обследовании, случайным образом выбирают одного человека для заполнения *индивидуального вопросника*. В *индивидуальном вопроснике* содержатся вопросы общего характера; вопросы о потреблении курительного табака; бездымного табака; отказе от потребления табака; пассивном курении; финансовой составляющей; источниках информации; и знаниях, отношении и восприятии различных аспектов, связанных с потреблением табака.

1.2 Использование данного руководства

В этом пособии предоставлены руководящие принципы по оценке и обеспечению качества данных, собранных при проведении GATS. Оно является дополнением к существующим руководствам GATS и не охватывает процедуры, которые уже описаны в других документах. Вместо этого в данном руководстве излагаются элементы процесса обеспечения качества, который должен иметь место в ходе проведения обследования GATS. Соблюдение этих руководящих принципов обеспечения качества является чрезвычайно важным для успеха обследования. Данное руководство предназначено для использования всеми участниками проведения обследования GATS и процесса обеспечения качества, в том числе ИТ-персоналом и статистиками,

ответственными за выборку и взвешивание. Терминологический глоссарий представлен в *приложении А*.

Процесс обеспечения качества обследования GATS, описанный в этом документе, хронологически разделен на три этапа: до сбора данных, в процессе сбора и работы с данными, после сбора данных. Деятельность по обеспечению качества различна на разных этапах, и руководящие принципы изложены для каждого этапа отдельно. В начале *главы 2* представлен план процесса и описан общий объем деятельности по обеспечению качества.

Глава 3. В описании этапа до сбора данных изложена процедура обеспечения качества, которая должна быть реализована в период планирования и анализа странового вопросника, внесения вопросника в программируемые устройства, применяемые в ходе сбора данных, составления выборки и создания контрольного файла выборки.

Глава 4. Описание этапа сбора данных и работы с данными представлено в трех разделах. Первый раздел посвящен деятельности по обеспечению качества, которая проводится в ходе выездных работ, включая настройку и обслуживание системы анализа случаев. Кроме того, он охватывает подразделы, касающиеся работы с данными, и содержит стандарты регулярной отчетности во время выездных работ. Во втором разделе описаны действия по работе с данными и обеспечению качества, проводимые после окончания интервью и загрузки страновых данных на центральный сайт. Третий раздел посвящен агрегированию данных и проблемам контроля качества, которые имеют место на уровне стран. Кроме того, в этом разделе описана подготовка файла с необработанными данными.

Глава 5. В описании этапа после сбора данных изложены действия по обеспечению качества, которые должны быть проведены после выездных работ и подготовки файла с необработанными данными. В первом разделе описаны процедуры очистки и подготовки данных для вычислений по взвешиванию выборки. Во втором разделе описывается оценка качества выборки, ошибки выборки и взвешивание выборки, тогда как третий раздел посвящен оценке качества охвата выборки, доли не ответивших и других ошибок, не связанных с проведением выборки. Наконец, в четвертом разделе описано создание конечного файла с аналитическими данными.

В этих главах изложены руководящие принципы по обеспечению качества и рекомендации по лучшей практике. Подробное описание методологии / документации для каждого раздела главы имеется в соответствующих добавлениях.

1.3 Обеспечение качества обследования GATS

Обеспечение качества – это процесс, состоящий из систематических мероприятий, направленных на обеспечение, оценку и подтверждение качества данных, собранных в ходе обследования (Viemer и Lyberg, 2003). Ранее понятием качества данных обозначали точность данных, однако в настоящее время этот термин имеет более общее значение. Данные высокого качества определяются как данные, которые соответствуют своему прямому назначению. Имеется в виду не только точность, но и своевременность, доступность и сопоставимость, что делает качество многомерным понятием. Набор данных точен в той степени, в какой он не содержит ошибок. Он является своевременным, если имеется в наличии в необходимый момент. Доступность данных определяется их относительным удобством или сложностью использования. Качественные

данные сопоставимы в случае, если их характеристики одинаковы для всех единиц: единиц сравнения отдельных лиц, интервьюеров, ПЕО или государств. Стандартизация повышает сопоставимость данных обследования, например использование стандартизированных вопросов обеспечивает единообразие вопросов, задаваемых различными интервьюерами. Кроме того, стандартизация процессов обеспечения качества гарантирует, что ошибки данных будут исправлены одинаковым способом.

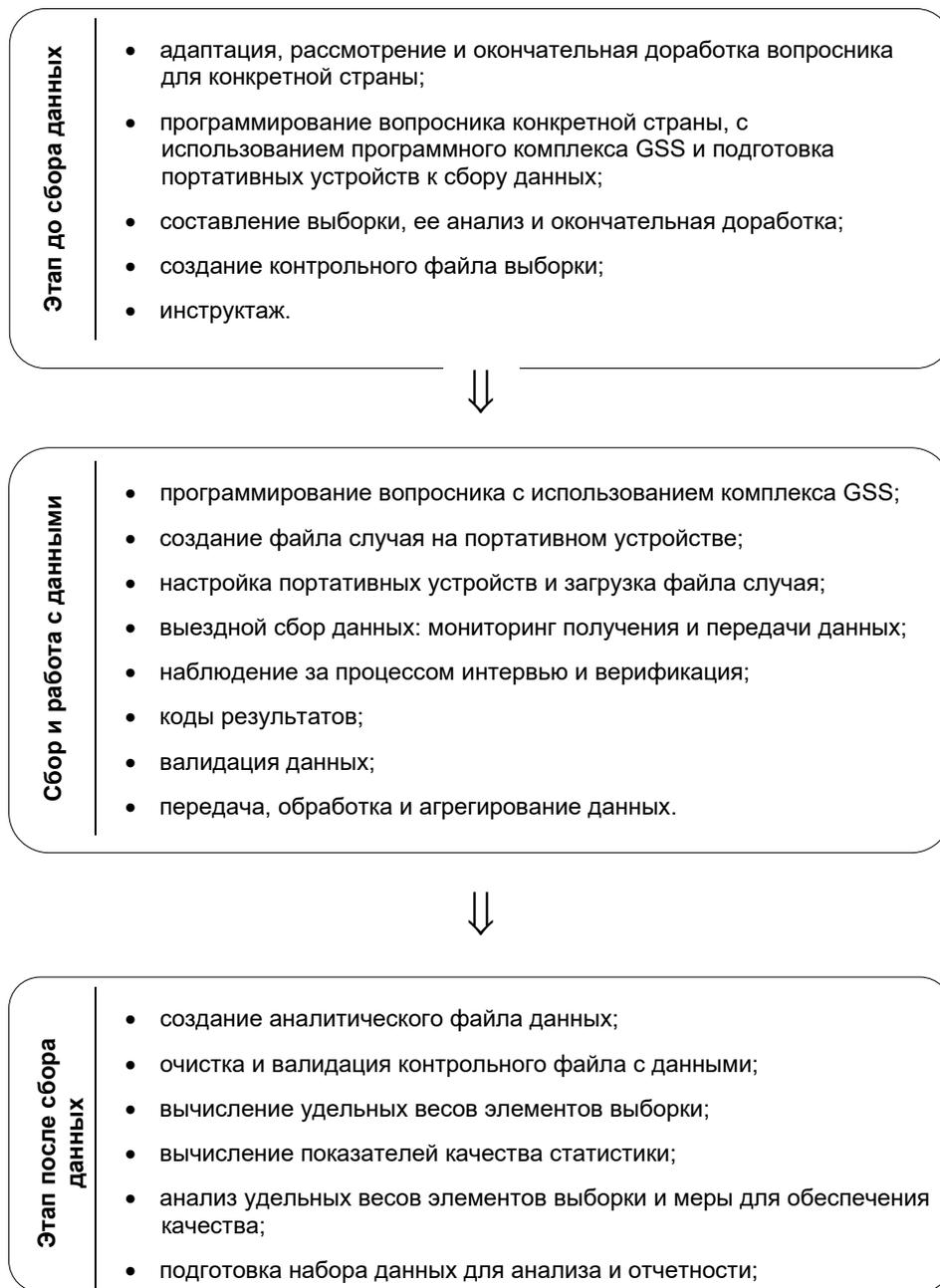
Многочисленные параметры качества часто ограничивают друг друга, так что достижение максимального качества является вопросом нахождения компромисса и баланса. Например, данные высокого качества должны быть точными, но вместе с тем должны быть своевременными. В случае, если достижение очень высокой точности займет продолжительное время, данные могут потерять актуальность.

Руководящие принципы, описанные в этом документе, представляют собой стандартизированные процедуры для обеспечения качества данных, собранных в результате обследования GATS. Странам рекомендуется включить дополнительные меры для обеспечения качества в дополнение к описанным в этом документе.

2. План процесса по обеспечению качества

На приведенной ниже иллюстрации представлена диаграмма действий по обеспечению качества в рамках общего временного графика и последовательности непересекающихся шагов. Эти шаги распределены на три этапа: до сбора данных, в ходе сбора данных и после сбора данных.

Иллюстрация 2-1. Диаграмма процесса обеспечения качества



3. Обеспечение качества: этап до сбора данных

Этап до сбора данных включает в себя ряд задач, которые необходимо выполнить для подготовки к сбору полного массива данных обследования. Эти задачи охватывают разработку вопросника, программирование, проведение предварительной проверки, разработку выборки и подготовку контрольного файла выборки, а также инструктаж. В этой главе описан механизм обеспечения качества и руководящие принципы по стандартизации, которые должны соблюдаться в этих различных областях. Данная глава посвящена предметному анализу следующих аспектов:

- 3.1 Процесс разработки и рассмотрения вопросника;
 - 3.1.1 рекомендации по адаптации вопросника.
- 3.2 Программирование и контроль версий;
 - 3.2.1 процесс программирования;
 - 3.2.2 проверка вопросника;
 - 3.2.3 настройка портативного устройства.
- 3.3 Предварительный опрос.
- 3.4 Окончательная доработка вопросника.
- 3.5 Передача данных и обработка данных.
- 3.1 Процесс разработки и рассмотрения вопросника.
- 3.7 Подготовка контрольного файла выборки.
- 3.8 Обучающие семинары.

3.1 Процесс разработки и анализа вопросника

В рамках обследования GATS предусмотрен стандартизированный процесс для стран-участниц, направленный на разработку адаптированных для стран вопросников. Комитет по рассмотрению вопросника GATS – это группа, состоящая из экспертов в области борьбы против табака и в сфере разработки вопросников из развитых и развивающихся стран, которая рассматривает и утверждает все вопросники обследования GATS. Данный комитет ведет тесное сотрудничество с каждой страной в целях адаптации вопросника обследования GATS в соответствии с ситуацией в каждой стране при одновременном сохранении стандартных основных вопросов обследования GATS (см. подробные сведения в документе «Основной вопросник обследования GATS с дополнительными вопросами»), обеспечивающих возможность межстранового сравнения. Конкретные аспекты адаптации вопросника и процесса его рассмотрения представлены в приложении В.1.

В целях обеспечения качества и сопоставимости данных, полученных в разных странах, Комитет по рассмотрению вопросника обследования GATS проводит анализ и утверждение вопросника.

3.1.1 Руководство по адаптации вопросника

Для упрощения процесса анализа и обеспечения соблюдения стандартов качества при адаптации «Основного вопросника обследования GATS» Комитет по рассмотрению вопросника рекомендует странам придерживаться ряда конкретных положений:

1. Выделить все элементы «Основного вопросника обследования GATS», которые были адаптированы (для удобства работы). Имеются в виду пункты вопросов, категории ответов, дополнительные вопросы и вопросы, добавленные странами.
2. Зачеркивать (для удобства) основные вопросы, которые страна планирует удалить.
3. Сохранять нумерацию и порядок основных и дополнительных вопросов для обеспечения сравнимости данных в разных странах (это не всегда бывает возможно или не обеспечивает оптимальный режим работы, однако, по возможности, должно быть выполнено).
4. Необходимо, чтобы добавленные страной вопросы обозначались двойными буквами в зависимости от раздела (например, AA10, BB17, EE4). Это позволит не нарушать текущую нумерацию из-за добавления странами новых вопросов.
5. Изменение инструкций, касающихся перехода от одного вопроса к другому, допускается только в том случае, если это необходимо для включения дополнительных вопросов или включения вопросов, добавленных странами.
6. В целях обеспечения единообразия сравнений между странами не следует проводить пересмотр основных стандартных вопросов (за исключением специфичных для стран категорий). О любых исключениях необходимо информировать Комитет по рассмотрению вопросника).
7. Следует ограничить количество дополнительных вопросов, с тем чтобы сохранить разумный объем вопросника.
8. Для упрощения программирования вопросника каждой из стран необходимо ограничить сложность дополнительных вопросов.
9. Следует предоставить обоснование внесенных изменений. Включение обоснований помогает Комитету по рассмотрению вопросника осуществлять процесс рассмотрения и утверждения национальных вопросников.

3.2 Программирование и контроль версий

В рамках обследования GATS используется режим сбора данных путем личного интервью с помощью компьютера (CAPI) с применением портативного КПК в качестве электронного инструмента. Эти портативные устройства оснащены программным обеспечением (GSS)¹. Так как каждая страна адаптирует вопросник к собственным нуждам и использует в работе собственный язык с собственной

Для обеспечения высокого качества эффективности программирования вопросника GATS необходимо выполнение стандартных процедур и проверки.

¹ Программный комплекс "General Survey System" (GSS) был разработан для использования на системе Android, а также проверен и отлажен с использованием планшетных компьютеров на системе Android.

письменностью, а также может применять различное оборудование (например, различные модели портативного оборудования), в каждой стране всегда имеют место уникальные особенности. Учитывая эти сложности, имеется ряд шагов и рекомендаций общего характера, которые необходимо соблюдать.

Существует последовательность шагов (изложены в приложении В.2), реализация которых требует заблаговременного проведения принимающей страной интенсивной работы в духе сотрудничества (желательно, за 6 – 8 недель до начала инструктажа по апробации). Для выполнения данных шагов необходимо в период вводного семинара или вскоре после него определить состав и доступность штата сотрудников принимающей стороны, в частности специалистов в области информационных технологий/программистов, специалистов по проведению обследований и менеджеров обследования.

Перед началом подготовительных работ для данных сотрудников должен быть проведен инструктаж в очной форме либо в форме вебинаров, и ожидается их продолжительное участие в подготовке к апробации GATS и полномасштабному проведению обследования.

В подразделах ниже рассмотрены конкретные вопросы, связанные с процессом подготовки вопросника и обследования, на которые необходимо обратить внимание в интересах повышения качества.

3.2.1 Процедуры для программирования

Для обеспечения неизменного качества и эффективности подготовки к проведению обследования GATS следует придерживаться следующих процедур:

- Ответственность за проверку точности формулировок на языке страны (переводов) на основе странового вопросника несут: персонал обследования/выездной персонал/координаторы проведения обследования (но не ИТ-персонал). В круг ведения ИТ-персонала входит взаимодействие с персоналом обследования/выездным персоналом/координаторами проведения обследования по вопросу ввода текста при помощи ПО для разработки вопросников GSS.
- Перед проведением обучения выездного персонала (по конкретной версии) координатор организации-исполнителя должен утвердить *вопросник для домохозяйств* (HQ) и *индивидуальный вопросник* (IQ) (в письменном виде).
- Изменения в тексте: при корректировке формулировок вопросника данные изменения должны быть отражены в версиях на всех языках, которые будут применяться.
- Изменение логики программирования (диапазоны, проверка пригодности данных, и т.д.) утвержденного вопросника требует повторного проведения анализа Комитетом по рассмотрению вопросника.
- Изменение содержания или формулировки вопроса либо ответа требует повторного проведения анализа Комитетом по рассмотрению вопросника.
- Контроль версий файлов DB3 (файлы базы данных вопросника обследования): принимающая страна является владельцем файлов DB3, и ИТ-персонал/персонал

обследования страны несет ответственность за ведение контрольного журнала изменений в вопросниках и программных спецификациях ПО GSS.

- После подведения итогов предварительной проверки и начала полномасштабного проведения обследования страна должна выполнить отправку окончательных файлов "CMSDB", "Survey0" и "Survey1" в формате DB3.

3.2.2 Проверка программы, содержащей вопросник

Имеется ряд шагов для обеспечения качества, позволяющих проверить программу вопросника, которые необходимо соблюдать и документировать:

- Программисты должны обновлять версию программы вопросника после каждого пересмотра. Кроме того, данную информацию необходимо подтвердить перед началом пересмотра вопросника. Архив устаревших файлов DB3 для истории и резервного копирования хранится программным обеспечением GSS.
- Полный перечень спецификаций для программирования GSS на каждом языке проверяется путем анализа каждого вопроса на портативном устройстве и проверки точного соответствия текста каждого вопроса, ответов и специальных символов тексту вопроса, ответам и специальным символам в документе «Технические параметры для программирования в ходе обследования GATS». В случае выявления каких-либо сложностей их четко отражают в спецификациях, устраняют в рамках одного цикла и затем проводят полную проверочную итерацию. Этот итеративный процесс повторяют до тех пор, пока работа портативного устройства с каждым вопросом не будет сочтена удовлетворительной. Отклонения от данной процедуры не допускаются. Данная работа должна проводиться совместно с лингвистом, содействующим проведению обследования. Эту задачу не следует поручать программистам либо персоналу, который не обладает необходимым опытом в области лингвистики. После анализа внесенных изменений/обновлений страновой координатор или представитель должен заверить подписью итоговую версию. Перед утверждением спецификации должны быть проанализированы на предмет того, что каждый вопрос был включен в очередность вопросника и проверен в прямом и обратном порядке (подробнее см. документ «Руководство по обучению управлению данными для проведения обследования GATS», модуль 2.1).
- Всякий раз перед выездом на места необходимо проводить полный анализ переходов от одного вопроса к другому, а также англоязычного текста. Программное обеспечение вопросника не следует изменять в ходе сбора данных на местах за исключением случаев крайней необходимости.
- Для каждого языка количество начальных и конечных HTML-тегов, символов «(», «)», «[», «]», «{» и «}» должно быть одинаковым. В противном случае проводят повторную проверку и исправление.

3.2.3 Настройка портативного устройства

Для загрузки ПО GSS в портативные устройства предусмотрены официальные процедуры обеспечения качества:

- Перед загрузкой на всех портативных устройствах должна быть завершена первоначальная настройка и, в применимых случаях, должны быть удалены предыдущие версии программного обеспечения GATS и соответствующие папки.
- На портативные устройства прикрепляют ярлыки с перечнем действий, необходимых для обеспечения качества (для установки материалов GATS на портативные устройства). Лицо, завершающее контроль качества загрузки на портативное устройство, должно подтвердить выполнение каждого действия.

В приложении В.3 предоставлена подробная информация о пошаговом процессе загрузки портативных устройств (см. подробные сведения в модуле 2.2 «Руководства по обучению управлению данными для проведения обследования GATS»).

3.3 Предварительная проверка (апробация)

Имеется шесть задач проведения предварительной проверки в стране до начала полного цикла обследования. Выполнение следующих задач будет способствовать тому, что сбор данных в рамках полного обследования будет проведен на высоком уровне:

1. Предоставление сотрудникам в стране помощи в программировании, тестировании и проведении опросов с использованием портативных устройств по мере необходимости.
2. Анализ вопросника GATS для проведения предварительной проверки, включая формулировку вопросов и ответов, указания для обследования, программирование логики пропусков, продолжительность и время интервью.
3. Оказание содействия персоналу в стране в использовании портативного оборудования и программного обеспечения по мере необходимости.
4. Обеспечение наблюдения за проведением апробации в одном городском и одном сельском районе в случае, если возможно и применимо.
5. Проверка во время апробации плана сбора, передачи и обработки данных, который впоследствии будет адаптирован для проведения полного цикла обследования.
6. Совместное с партнерами обсуждение сильных и слабых сторон, а также рекомендацией по результатам подготовки и проведения апробации обследования GATS.

Предварительная проверка перед обследованием GATS дает возможность тестирования вопросника и процедур сбора и управления данными в целях обеспечения высокого качества полномасштабного исследования.

3.4 Окончательная доработка вопросника

Для окончательной доработки вопросника GATS в целях проведения полного обследования необходимо предпринять следующие шаги:

- **Анализ результатов предварительной проверки.** Все сотрудники (выездные интервьюеры, начальники выездных работ, персонал проекта) должны сообщить о функционировании вопросника в целом, а также о конкретных вопросах обследования, которые необходимо пересмотреть.

- **Анализ данных предварительной проверки.** Анализ данных предварительной проверки может внести ценный вклад в осуществление изменений после предварительной проверки. Некоторые конкретные вопросы для рассмотрения:
 - Отсутствие ответа на конкретные пункты: выявление вопросов с более высокой, чем ожидалось, долей ответов «неизвестно» или «отказ от ответа».
 - Частоты: изучить частоту распределения ответов на основные вопросы, с тем чтобы определить, обладают ли данные соответствующими характеристиками. Необычные отклонения могут указывать на проблемы разработки вопросника или другие проблемы (например, проблема инструктажа).
 - Допустимые диапазоны: провести анализ целесообразности допустимых диапазонов, внесенных в портативные устройства, и внести изменения согласно ситуации для проведения полного цикла обследования.
 - Категории ответов: определить, уместны ли адаптации вопросников для конкретных стран (например, изменение перечней вопросов и категорий ответов). Например, для изменения перечня торговых марок в вопроснике для проведения полного цикла обследования необходимо определить частоту ответов о марках сигарет.
- **Пересмотр данных для полного проведения обследования.** Пересмотренная версия, которая будет подготовлена после предварительной проверки, должна быть представлена в Комитет по рассмотрению вопросника, в том числе с обоснованием (на основе результатов предварительной проверки) изменений. При подготовке вопросника для проведения полного цикла обследования (в том числе изменений в программируемые параметры и т. п.) должны быть соблюдены те же руководящие принципы обеспечения качества, которые описаны в разделе 4.1 и 4.2 (если применимо).

3.5 Передача данных и обработка данных

Предварительная проверка позволяет всесторонне проверить процесс обработки данных, который будет реализован в ходе полного цикла проведения обследования. На основании опыта предварительной проверки и сделанных рекомендаций процесс передачи и обработки данных будет доработан для целей полномасштабного обследования. Поэтому каждый этап процесса обработки данных должен проводиться в соответствии с планом обработки данных в контексте всего обследования. Например:

- Во время проведения предварительной проверки (апробации) необходимо отработать функционирование механизма обратной связи с предоставлением отчетов.
- В ходе апробации необходимо проверить функционирование способов передачи данных, например, с использованием wi-fi, SIM-карт, облачного хранилища или FTP, и при необходимости должны быть внесены коррективы с учетом уроков, извлеченных из предварительной проверки.

3.6 Процесс разработки и анализа вопросника

В рамках обследования GATS осуществляется стандартизированный процесс для стран-участниц, направленный на определение плана выборки. Комитет по рассмотрению выборки обследования GATS – группа, состоящая из экспертов по методологии выборки из развитых и развивающихся стран – рассматривает и дорабатывает состав выборки для стран-участниц. Комитет по рассмотрению выборки и координаторы CDC работают в тесном взаимодействии с каждой страной в целях адаптации состава выборки обследования GATS при сохранении стандартов GATS для обеспечения сопоставимости в целях межстранового анализа. Конкретные данные о составе выборки для обследования GATS представлены в документе «Рекомендации по проведению инструктажа в сфере работы с данными GATS».

В ходе определения выборки для обследования GATS может быть необходимым составление карт и перечней, с тем чтобы получить основу выборки, которая опирается на районирование по условиям вероятности. Подробные сведения о процедурах и рекомендациях обеспечения качества при составлении карт и перечней изложены в документе «Руководство по составлению карт и перечней для обследования GATS».

3.7 Подготовка контрольного файла выборки

После проведения выборки для обследования GATS необходимо подготовить файл для составления основной выборки. Контрольный файл выборки представляет собой массив данных, который содержит код случая для каждого домохозяйства в списке выборки наряду с информацией, необходимой для вычисления удельного веса элементов выборки и анализа сложных данных обследования.

Контрольный файл выборки включает в себя идентификаторы выборки и географические показатели. Идентификаторы выборки, такие как страта, первичная единица отбора (ПЕО), вторичная единица отбора (ВЕО/сегмент), тип места проживания (город/сельская местность) и предварительное распределение респондентов домохозяйства по критерию пола в случае, если проводилась рандомизация по данному критерию. Географические идентификаторы, такие как информация об адресе, включая коды страны, региона, провинции/штата, округа/района и города/поселка городского типа/села.

Контрольный файл выборки включает уникальный идентификатор случая, идентификаторы выборки и географические идентификаторы домохозяйства.

По завершении формирования контрольного файла выборки, его необходимо упорядочить и перекодировать в текстовый файл с применением ПО Excel и/или MS Access. Текстовый файл должен использоваться для внесения случая в портативные устройства для сбора данных в применимых ситуациях. Кроме того, при работе с ПО GSS необходимо заполнить данные на экранах "Case Grid" (таблица случаев) и "Case Info" (информация о случаях). В следующей главе приведена дополнительная подробная информация об этом процессе.

3.8 Обучающие семинары

Инструктаж является важнейшим механизмом обеспечения качества при проведении обследования GATS. Обучающие семинары, проводимые для стран, охватывают ознакомление с обследованием GATS, а также инструктаж по предварительной проверке и полному циклу обследования. Обучающие семинары являются частью протокола GATS и стандартов обеспечения данных высокого качества для каждой страны.

Обучающие семинары GATS проводятся для коллектива страновой группы обследования на всех уровнях, от интервьюеров до начальников работ, персонала ИТ-служб и центральной группы, осуществляющей надзор за полным циклом процесса от его начала до его завершения. Национальные и международные партнеры содействуют проведению инструктажа по обследованию GATS на всех уровнях. Инструктаж всех сотрудников гарантирует, что всем лицам, участвующим в процессе обследования, известны их роли в обеспечении данных надлежащего качества. Инструктаж является непрерывным процессом, проводимым до начала обследования, а также в ходе реализации обследования с подведением итогов в конце каждого инструктажа.

Целями инструктажа по проведению обследования GATS являются:

- Обеспечить стандартизированное применение протокола обследования, стандартов и материалов.
- Обеспечить четкое понимание обоснования обследования и протокола обследования.
- Обеспечить стандартное применение системы сбора, обработки, передачи и агрегирования электронных данных обследования GATS.
- Обеспечить привлечение и мотивирование интервьюеров.
- Подготовить практические предложения.
- Повысить общее качество получаемых данных.

3.8.1 Ознакомительный семинар по обследованию GATS

Ознакомительный семинар является первым этапом в обеспечении качества обследования GATS. Он охватывает углубленный анализ стандартного протокола GATS, включая вопросник, план выборки, электронный сбор и работу с данными, а также их распространение. Страны осуществляют обмен опытом обследований, которые отражают их технический потенциал для проведения GATS в соответствующей стране, а также для понимания необходимого объема технической помощи учреждению со стороны партнеров в целях достижения желаемых результатов. В ходе ознакомительных семинаров рассматриваются рекомендации по разработке предложений, что позволяет странам адаптировать протокол к конкретной ситуации на местах. Помимо этого, ориентационная сессия позволяет ознакомиться со всеми этапами разработки, подачи, рассмотрения и финансирования предложения, что имеет решающее значение для инициирования и проведения обследования.

3.8.2 Предварительная проверка (апробация)

Цели проведения предварительной проверки были подробно изложены в разделе 3.3. Инструктаж по проведению предварительного тестирования охватывает различные элементы, в том числе подготовку сотрудников ИТ-службы, выездных интервьюеров/начальников выездных работ, выездную практику и разбор результатов. Предварительная проверка позволяет персоналу проекта получить опыт подготовки выездных сотрудников. В ходе предварительной проверки

ИТ-персонал и выездные сотрудники знакомятся с программными и аппаратными средствами, которые необходимы для содействия проведению обследования GATS. Обучение гарантирует надлежащий уровень подготовки персонала, а также ознакомление с программными и аппаратными средствами, поддерживающими проведение обследования GATS.

3.8.3 Инструктаж по полному циклу проведения обследования

Инструктаж по полному циклу проведения обследования является важнейшим механизмом обеспечения качества в рамках обследования GATS. Его целью является подготовка интервьюеров и начальников работ в области получения полезных и надежных данных выездного обследования. В ходе инструктажа рассматривается обоснование обследования, которое естественным образом проистекает из истории борьбы против табака, а также из деятельности глобальной системы эпиднадзора за табаком, методы проведения интервью, выявление отобранных домохозяйств, согласие, содержание вопросника и практические компоненты, которые включают имитационные упражнения и наблюдение на местах. Инструкторы совместно с участниками изучают вопросники в бумажном и электронном виде на планшетах, с тем чтобы повысить степень осведомленности об обследовании. Кроме того, имитационные упражнения и выездная практика помогают обучающимся в ознакомлении с вопросником и сущностью проблем, возникновение которых можно ожидать в ходе работы с «трудными» респондентами, а также в преодолении этих проблем. Кроме того, обучающимся предоставляются учебные материалы для получения справочной информации в ходе обучения и выездной работы. К числу этих материалов относятся вопросник, повопросные разъяснения и руководство интервьюера.

4. Обеспечение качества: сбор и работа с данными

В этой главе описаны процессы и процедуры обеспечения качества данных при проведении сбора и обработки данных обследования GATS. Процесс сбора и обработки данных охватывает настройку программного комплекса GSS для проведения обследования GATS, использование программного обеспечения во время выездного сбора данных, передачи собранных данных, агрегирования данных и создания контрольного файла объединенной базы данных. Он также охватывает мониторинг, ведение отчетности и оказание технической поддержки во время выездного сбора данных.

В любой стране сбор, агрегирование и обеспечение качества данных должно осуществляться на двух уровнях: 1) на местах, 2) в Национальном центре данных (НЦД). В этой главе описаны процессы и процедуры обеспечения качества, применимые на всех двух уровнях, в частности:

- 4.1 Сбор данных при проведении выездных работ: настройка и техническое обслуживание;
 - 4.1.1 файл случая и анализ случаев;
 - 4.1.2 электронные устройства для сбора данных и программное обеспечение для проведения обследования;
 - 4.1.3 процедуры обеспечения качества в ходе выездного сбора данных.
- 4.2 Обработка данных в Национальном центре данных;
 - 4.2.1 агрегирование данных;
 - 4.2.2 контроль качества собранных данных;
 - 4.2.3 связь и техническая поддержка.

4.1 Сбор данных при проведении выездных работ: настройка и техническое обслуживание

В рамках стандартного протокола обследования GATS рекомендуется внедрение и соблюдение определенных систематических процедур и их соблюдение сотрудниками ИТ-служб и выездным персоналом в целях получения точных и полных данных во время всего периода сбора данных. Эти процедуры описаны ниже.

4.1.1 Файлы случаев и анализ случаев

1. Рекомендуется проводить подготовку и обработку файлов случаев в Национальном центре данных (НЦД) с использованием ПО GSS в применимых случаях согласно национальному плану обработки данных (необходимо заметить, что файлы случаев являются компонентом контрольного файла выборки).
2. В ходе создания файла случая и настройки портативного устройства:
 - Необходимо указать всех выездных интервьюеров (ВИ) и закрепить за ними портативные устройства. Назначенные идентификаторы случаев ("CaseID") должны быть уникальными и конкретными для каждого ВИ и связаны с переменными FIID в файле случая. На момент загрузки конкретные идентификаторы случаев ("CaseID") не должны закрепляться более чем за одним идентификатором ВИ ("FIID"). Это позволит

гарантировать отсутствие дублирования, обеспечить закрепление разумного количества случаев за ВИ и упростить идентификацию случаев.

- Назначение всех идентификаторов ВИ ("FIID") должно производиться в соответствии с правилами GATS, согласно которым присваивается цифровой идентификатор из 6 цифр.
 - Назначение всех идентификаторов случаев ("CaseID") должно производиться в соответствии с правилами GATS, согласно которым присваивается цифровой идентификатор из 6 цифр.
3. Файл случая должен содержать информацию о задачах по сбору данных, которые закреплены за каждым портативным устройством/ВИ.
- Необходимо удостовериться в том, что за всеми ВИ, указанными в таблице идентификаторов ВИ ("FIID") закреплены назначения в файле случая. Перед настройкой портативных устройств данный файл должен быть проверен. При загрузке на портативные устройства в системе анализа случаев должно отобразиться ориентировочное количество случаев. В противном случае сверьте идентификатор ВИ ("FIID") в портативном устройстве с идентификатором выездного интервьюера в файле случая. Необходимо заметить, что для каждого идентификатора случая ("CaseID"), будет создано две записи, которые загружаются в систему анализа случаев (HQ (вопросник для домохозяйств) = 00 и IQ (индивидуальный вопросник) = 01).
4. Каждая таблица случаев портативного устройства должна содержать только те случаи, которые были закреплены и предназначены для обработки конкретным выездным интервьюером. Как правило, за одним ВИ закрепляют с последующей загрузкой в портативное устройство не более 300 идентификаторов случаев ("CaseID"). Если загружено большее количество случаев, чем рекомендуется, это может сказаться на параметрах производительности, таких как загрузка приложения и возможность оперативной идентификации случаев в ходе выездных работ.
5. Перед загрузкой случая выполните на компьютере НЦД частотный анализ значений полей в файле случая. Требования к значениям полей случая в системе анализа случаев см. в документе «Руководство по программированию комплекса "General Survey System"».
- Поле "CASEID" (идентификатор случая) необходимо проверить на предмет дублирования, а также соблюдения стандартных для обследования GATS 6-значных цифровых обозначений.
 - Необходимо убедиться, что в поле "PROJECTNAME" (название проекта) указано название страны, так как это позволит обеспечить наличие уникального идентификатора страны в файле случая (может применяться запрос на обновление поля "projectname" в шаблоне файла случая в БД MS Access).
 - Необходимо убедиться в том, что значение переменной "CreateDate" (дата создания) соответствует дате, когда был создан файл.
 - Для обеспечения полноты и правильности сравните контрольный файл выборки с файлом случая на портативном устройстве (необходимо заметить, что файл случая является компонентом контрольного файла выборки).

В документе «Руководство по программированию комплекса "General Survey System"» содержится подробная информация о создании файла случая, описании полей и таблице инвентаризации для выездного интервьюера.

4.1.2 Электронное устройство для сбора данных и программное обеспечение для проведения вопросника

1. **Настройки, доступные для выездного интервьюера:** на каждом портативном устройстве должен быть нанесен уникальный идентификатор ВИ ("FIID") в соответствии с правилами GATS, и тот же идентификатор ВИ ("FIID") должен быть установлен и должен отображаться в меню настроек системы анализа случаев. Национальный центр данных должен вести учет этих сведений, а также данных об имени выездного интервьюера и серийном номере портативного устройства в табличной форме.
 - Дополнительная проверка настроек для идентификатора ВИ ("FIID"): до начала сбора данных убедитесь в том, что задание, указанное в таблице, значение поля идентификатора ВИ ("FIID") в меню "Settings" (настройки) и маркировка портативного устройства совпадают.
2. **Версии программного обеспечения: система анализа случаев, вопросник для домохозяйств и индивидуальный вопросник.** Необходимо обеспечить загрузку в портативные устройства одобренных вопросников в окончательной версии.
 - Необходимо убедиться в том, что на каждом портативном устройстве имеется загруженная и функционирующая версия каждого вопросника: для этого в верхней части экрана следует выбрать пункт "Information" (информация) и удостовериться в актуальности версии, указанной в разделе "Survey Information" (информация об обследовании).
 - По возможности все портативные устройства необходимо подготовить одновременно в целях обеспечения единообразия версий, процесса подготовки и контроля качества.
3. **Загрузка случаев.** На момент начала сбора данных/выездных работ в системе анализа случаев на каждом портативном устройстве должны отображаться только случаи, указанные в файле, который закреплен за конкретным идентификатором ВИ ("FIID"). До начала выездных работ необходимо удалить учебные случаи, с тем чтобы исключить возможность проведения работы по тренировочным случаям и экспорт этих случаев. Аналитики должны быть подготовлены к разграничению учебных случаев и рабочих данных в ситуациях, когда выездные интервьюеры нарушают указанное правило.

В «Руководстве по программированию комплекса "General Survey System"» содержатся подробные указания относительно настройки портативных устройств, а также создания и установки соответствующих файлов.

4.1.3 Процедуры обеспечения качества выездного сбора данных

В этом разделе представлен обзор рекомендованных процедур обеспечения качества, которые могут быть реализованы в процессе выездного сбора данных. Просьба обращаться за подробными сведениями к документам «Руководство для выездного интервьюера обследования GATS» и «Руководство для полевого супервайзера обследования GATS».

Роль полевого супервайзера

Начальники выездных работ (НР) выполняют важную задачу по обеспечению взаимодействия ВИ и управляющего персонала обследования GATS. Планируется, что они будут контролировать проведение и эффективность обследования и сообщать о любых вопросах, возникающих в ходе работы на местах, которые могут повлиять на своевременное завершение обследования GATS. Им поручено обеспечение соответствия данных, собираемых подчиненными, протоколу сбора данных обследования GATS, а также самым высоким стандартам качества. Для понимания ВИ важности, которую придает группа проведения обследования GATS качеству данных, НР должен сделать контроль качества неотъемлемой частью еженедельного цикла работы.

Первым шагом НР к обеспечению качества данных является информирование ВИ об ожидаемом качестве данных. Ожидается, что начальники работ будут поддерживать постоянную обратную связь с ВИ по вопросам качества, включая как положительные, так и отрицательные аспекты, и будут подчеркивать ту важность, которую группа проведения обследования GATS придает *качеству*.

Контроль генерирования данных

Контроль генерирования данных охватывает все мероприятия, необходимые для успешного достижения целей обследования GATS в части доли ответивших лиц. Эти цели охватывают налаживание взаимодействия с каждым назначенным домохозяйством; установление связи, определение местонахождения и приглашение к участию; а также успешное выполнение всех этих процедур в соответствии со всеми требованиями обследования GATS. НР и аналитики в стране должны регулярно контролировать время, затраченное на проведение интервью. Кроме того, контроль генерирования данных в период их сбора необходим для своевременного завершения обследования GATS и его соответствия установленному бюджету.

С помощью ПО GSS могут быть созданы указанные ниже типы отчетов, и они должны использоваться специалистами по проведению обследования в странах или ИТ-специалистами в целях контроля различных аспектов генерирования данных и в идеале должны использоваться совместно с начальником выездных работ на регулярной основе. Приемлемое место и время использования данных отчетов будет определяться внедренной в каждой из стран моделью работы с данными.

- *Статус случаев, учитываемых по вопроснику для домохозяйств, и случаев, учитываемых по индивидуальному вопроснику.* В этих отчетах приводится перечень последних кодов результата по вопросникам для домохозяйств и индивидуальным вопросникам, с которыми работал каждый из интервьюеров. Данный отчет можно просмотреть на национальном уровне и направить НР и/или ВИ.
- *Отчет об исключениях.* Этот отчет будет содействовать персоналу обследования GATS в ходе контроля за количеством повторных посещений по любому незавершенному случаю, а также позволит понять причины повторных посещений.
- *Отчет о частотном распределении при работе по индивидуальному вопроснику.* Данный отчет будет содействовать ИТ-персоналу в мониторинге ответов, содержащихся в вопроснике, в разбивке по каждому интервьюеру или ПЕО, в целях оценки качества интервью обследования на основе таких показателей качества, как:

- потенциальные проблемы, связанные с выбором тактики интервью или «недостающими» данными;
- доля ответов «неизвестно» или «отказ от ответа», превышающая ожидаемые уровни;
- вопросы, в которых респонденты отдавали предпочтение ответу «другое», нежели другим предложенным вариантам, чаще ожидаемого;
- выбор ответа «другое, укажите», который допускает двоякое толкование и затруднителен для кодирования, или дублирующий один из ранее закодированных вариантов.

Контроль передачи данных

Регулярная передача данных от ВИ в НЦД играет важнейшую роль для управления обследованием GATS и раннего выявления и решения возможных проблем. НР несут ответственность за обеспечение передачи данных ВИ согласно регулярному графику (как правило, ежедневно). ВИ могут взаимодействовать с ИТ-персоналом в странах для определения возможных проблем передачи данных.

Наблюдение

Наблюдение за ВИ является важнейшим этапом усилий НР по обеспечению контроля качества и может содействовать повышению качества работы ВИ. Кроме того, наблюдение за работой ВИ будет способствовать пресечению любых попыток упрощенного проведения интервью или иных нарушений протокола проведения интервью для обследования GATS. По меньшей мере ожидается, что НР будут наблюдать за вверенными ВИ в течение первых нескольких дней выездной работы, а затем реже, исходя из рекомендаций организации-исполнителя. В период этих наблюдений НР сопровождает ВИ во время посещения выбранных домохозяйств, с тем чтобы убедиться в правильности учета результатов визита ВИ в системе анализа случаев GSS, а также в том верно ли ВИ проводит скрининг домохозяйств и последующее интервью. Подобные визиты на ранних этапах работы будут крайне полезными для того, чтобы обеспечить своевременное выявление ошибок при заполнении вопросников или использовании кодов результатов, а также для организации дополнительного инструктажа для выездных интервьюеров в случае необходимости и подготовку методических рекомендаций.

Кроме того, НР может принять решение о сопровождении ВИ в наиболее сложных случаях (например, в домохозяйства, которые ранее отказались от участия).

Проверка

Одним из способов проверки качества данных, собранных сотрудниками, проводящими интервью, является проверочное короткое интервью с домохозяйствами, которые уже прошли скрининг и были опрошены. Проведение коротких проверочных интервью позволит НР подтвердить, что ВИ предприняли следующее:

- Верная идентификация домохозяйства и проведение скрининга. Иногда ВИ может выбрать иное домохозяйство, чем то, которое вошло в выборку. В этом случае НР должен направить выездного интервьюера в нужное домохозяйство для проведения скрининга и интервью в данном домохозяйстве.

- Правильное указание адреса, пола и отношения членов домохозяйства к курению. Если возраст указан неверно (например, ВИ включает в перечень жителей младше 15 лет в дополнение к лицам в возрасте 15 лет и старше), эта ошибка повлияет на идентификацию тех лиц, которые имеют право участвовать в обследовании, и может привести к неверному выбору ВИ члена домохозяйства для интервью.
- Проведение интервью согласно *индивидуальному вопроснику* с выбранным членом домохозяйства.

Точное количество проверочных интервью будет определяться совместно с управляющим персоналом организации-исполнителя; вместе с тем обычно предполагается, что НР должен провести проверочные интервью со случайно выбранными 10% всех домохозяйств, назначенных каждому ВИ. Проверочное интервью может проходить в форме (1) краткой проверки, в ходе которой выбранному респонденту задается несколько вопросов с целью удостовериться в том, что его недавно опрашивали по темам, связанным с курением, и оценить проведенное интервью; или (2) повторный опрос в полном объеме по *вопроснику для домохозяйств* и, если возможно, по *индивидуальному вопроснику*, при условии, что данная процедура предусмотрена на уровне страны. Принимая во внимание, что респондент может воспринять повторное интервью как обременительное, не следует проводить интервью по *индивидуальному вопроснику* в полном объеме. Кроме того, проверочные интервью могут проводиться как в письменной форме, так и с использованием портативного устройства с введенными в него данными о домохозяйствах, участвующих в проверке. Портативное устройство ВИ никогда не должно использоваться для проведения повторного интервью, вместо этого следует использовать другое портативное устройство.

В дальнейшем НР сравнит ответы, полученные при проведении повторного интервью, с ответами, записанными ВИ. В случае, если будут выявлены расхождения, и будет установлено, что ВИ ошибочно проведены собеседования с иным членом домохозяйства, НР должен повторно направить ВИ в данное домохозяйство для проведения интервью с конкретным членом домохозяйства и вновь сравнить ответы. Во всех случаях независимо от наличия или отсутствия ошибок, выявленных в процессе проверки, информация о проверочных интервью должна быть передана в НЦД наряду со всеми другими материалами.

Присвоение кодов окончательного результата незавершенным интервью

До окончания работы по вопроснику НР должны утвердить окончательное присвоение кода незавершенного интервью. (В «Руководстве для полевого супервайзера обследования GATS» представлены дополнительные сведения о мерах, которые могут быть предприняты для окончания незавершенных интервью.) Если в конкретном случае не предусмотрено дополнительных мер, ВИ должен внести в систему анализа случаев соответствующий код незавершенного интервью (перечень и описание кодов результата выездных работ см. в «Руководстве для выездного интервьюера обследования GATS»).

По окончании сбора данных в определенной области (например, ПЕО, регион, штат и т. д.), всем случаям, по которым проводилась работа, должен быть присвоен код окончательного результата. Программа автоматически присвоит код окончательного результата завершенным

По завершении сбора данных всем случаям должен быть присвоен код окончательного результата.

вопросникам для домохозяйств и индивидуальным вопросникам. ВИ могут присваивать коды окончательного результата незавершенного интервью при проведении интервью *по вопроснику для домохозяйств и индивидуальному вопроснику* в случаях, которые не окончились завершением интервью (например, отказы или отсутствие дома), с разрешения своего непосредственного руководителя.

4.2 Обработка данных в Национальном центре данных

В этом разделе описаны процедуры по сбору и обработке данных на уровне НЦД: объединение/агрегирование файлов, аспекты контроля качества, в том числе отчетность и контроль, связь и техническая поддержка, а также создание файлов с необработанными данными.

4.2.1 Агрегирование данных

При поступлении данных в пункт агрегирования (облачный сервер, FTP-сервер и т. п.) файлы должны пройти анализ, обработку и контроль качества. Необходимо документирование и архивация файлов с данными.

Во время обработки программное обеспечение сортирует и обрабатывает файлы в выбранной папке (как правило, C:\GATSAndroid2\IncommingDB3s\...) в алфавитном порядке, вне зависимости от даты. Крайне важно, чтобы агрегирование файлов проводилось в порядке их экспорта. Благодаря использованию при агрегировании всех экспортируемых файлов принятой номенклатуры, FI#_ГГГГ_ММ_ДД.db3, обеспечивается автоматическая сортировка в алфавитном порядке и правильная обработка данных. Экспортируемые файлы являются накопительными и, таким образом, самый последний файл содержит все данные, в том числе обновления и изменения, которые имелись в предыдущих файлах.

Процесс агрегирования не позволяет использовать дважды использовать один и тот же идентификатор случая ("CaseID").

Обработка при агрегировании

- В начале сбора данных в папке C:\GATSAndroid2\DATA под именем "master.db3zero" имеется контрольный файл агрегирования. Это необходимо в начале сбора данных для проведения агрегирования с контрольного файла, не содержащего никаких данных. Необходимо скопировать файл "master.db3zero" и переименовать последующий файл в "master.db3" при необходимости повторной инициализации.
- Файл "Node.id" должен оставаться в папке C:\GATSAndroid2\
 - До начала первого агрегирования данных на конкретном компьютере имя файла "node" необходимо изменить на уникальное и информативное имя.
 - После начала агрегирования файл "Node.ID" нельзя изменять.
- В странах, применяющих шифрование данных, позднейшие версии полученных файлов db3 должны быть расшифрованы с использованием функции дешифровки в ПО GSS: Data Aggregation > Database Decryptor. В случае использования шифрования данных эту операцию необходимо проделать до начала агрегирования.

- При условии получения и обработки файлов в определенном порядке (по дате экспорта), следует агрегировать позднейшие полученные файлы db3 с агрегированными ранее в целях обеспечения эффективности. Хотя допустимо объединение данных с пустым контрольным файлом при каждом запуске агрегирования, следует отметить, что это займет больше времени, так как в этом случае должны быть обработаны все экспортированные файлы db3.
- Все полученные экспортированные файлы необходимо сохранить путем архивирования во вложенной папке в C:\GATSAndroid2\. Не следует проводить их повторное агрегирование.

Во время и после агрегирования необходимо проверить несколько аспектов:

- Наименования всех полученных файлов сделаны по дате экспорта, что исключает их неупорядоченную обработку. Если файлы получены не по порядку, (например, файлы с более ранними данными получены позже), последние данные для конкретного идентификатора случая ("CaseID") могут быть заменены на более ранние, возможно, неполными. Чтобы убедиться в актуальности обрабатываемых данных при агрегировании с непустым файлом "master.db3", необходимо выполнить сверку с архивированными файлами db3. (Следует всегда агрегировать файлы в соответствии с датой или использовать наиболее актуальные файлы.)
- Правильность и полнота агрегирования данных: были ли учтены и включены в процесс агрегирования все файлы db3?
- Проведено ли агрегирование наиболее актуального файла db3 каждым ВИ?
- Анализ данных: должна быть выполнена проверка и валидация данных. Для выявления возможных проблем, схем или несогласованных данных аналитик должен проверить даты событий, а также заполненные вопросники. Кроме того, в ходе этой работы будут сгенерированы частоты по каждому вопросу. Сведения о просмотре файлов DB3 можно найти в документе «План проведения работы с данными GATS».

Для подтверждения значений необходимо проанализировать сводный отчет об агрегировании и отчет о статусе, которые были получены после агрегирования.

В Отчете об агрегировании данных предоставлена информация о результатах агрегирования. В каждой строке указаны результаты для единиц агрегирования (как правило, ВИ) и определенная таблица баз данных. Таблица представлена в колонке "Aggtable", число строк в таблице исходных данных показано в колонке "AggRowsTotal", а число строк, вставленных из этой таблицы, показано в последней колонке "AggRowsInserted".

Отчет о состоянии скрининга домохозяйств содержит сводную информацию о статусе всех случаев в файле с агрегированными данными. Общее количество случаев для каждой единицы агрегирования (как правило, ВИ) показано в колонке "Cases" (случаи). Затем данное общее количество разделяют в зависимости от статуса. Содержание столбцов с информацией о статусе (выполненные интервью, не начатые, ожидающие завершения и завершенные) определяется наиболее актуальным кодом результата для конкретного случая.

- С помощью утилит объединения данных в инструментах разработчика могут быть созданы неструктурированные файлы данных или файлы CSV на основе агрегированных данных или любого файла db3.
- С тем чтобы ознакомиться с содержанием файлов и особенностями работы на местах, рекомендуется проводить анализ данных о событиях (имеется в файлах DU и DuEvt).
- В файле DU для каждого идентификатора случая содержится только одна запись (идентификаторы случаев в вопросниках для домохозяйств и индивидуальных вопросниках различаются и обозначены как -00 и -01 соответственно) и самый актуальный код результата по данному случаю. Файл DuEvt содержит историю событий для конкретного идентификатора случая ("CaseID"), которая указывается ВИ, и может содержать более одной записи на идентификатор случая ("CaseID").

Генерирование файлов с данными для анализа

- Данные вопросника (таблица ответов), полученные из форм вопросника для домохозяйств и индивидуального вопросника, хранятся в одной таблице. Для каждого вопроса предусмотрено по одной строке (QID), в связи с чем до начала анализа необходима реструктуризация этих данных. Данные об ответах необходимо транспонировать для упрощения анализа и простоты использования в работе статистического ПО. В результате этого данные вопросника для домохозяйств и индивидуального вопросника войдут в структуру новой записи на основе идентификатора случая ("CaseID"). Программное обеспечение GSS выполняет транспозицию данных об ответах (вопросник для домохозяйств и индивидуальный вопросник), которые основаны на db3 файлах обследования, с уточнениями на основе файлов уровня отдельных идентификаторов случая ("CaseID"), неструктурированных файлов.
- Данные о событиях GATS хранятся в таблице DUEVT, в одной строке на каждый результат по каждому идентификатору случая ("CaseID"). Данные GATS DU (жилищная единица) хранятся в таблице DU, в одной строке на каждый идентификатор случая ("CaseID"). Кроме того, данные DU содержат копию самого последнего кода результата по каждому случаю.

В таблице ниже обобщены источники этих данных из системы портативного устройства.

Источник табличных данных	Таблица	Строки
Данные вопросника	Ответы	Одна строка на вопрос для каждого идентификатора случая ("CaseID") и каждой формы
Данные о событии	DUEVT	Одна строка на результат для каждого идентификатора случая ("CaseID") и каждой формы
Жилищная единица	DU	Одна строка на одну жилищную единицу
Редактирование или изменение адреса	Журнал адресов	Одна строка на одно изменение адреса
Примечания ВИ	Примечания	Одна строка на примечание определенного типа (на уровне идентификатора случая ("CaseID") либо на один вопрос (QID))

4.2.2 Контроль качества собранных данных

В НЦД должны быть реализованы процедуры по обеспечению контроля качества для регулярного анализа, мониторинга, учета и резервного копирования данных по мере их поступления и обработки. В НЦД должен быть установлен по крайней мере ежедневный график отчетности и мониторинга, а также график, который предусматривает регулярный процесс резервного копирования.

Стандартный протокол GATS состоит из списка переменных показателей GATS и предлагаемой отчетности на уровне страны. Подробные сведения и информация об основных переменных показателях представлены в документе «Определения показателей для обследования GATS». Страны могут добавлять/адаптировать переменные показатели соответственно окончательно утвержденному вопроснику.

В процессе сбора данных рекомендуется анализ основных переменных показателей на уровне исходных данных.

Отчеты, подготавливаемые в стране: рекомендуется наладить в странах подготовку отчетов о событиях на уровне отдельных случаев. Мониторинг обследования (например, ход выездных работ, сбора данных и их передачи) является основным видом постоянно координируемой деятельности.

- Во время периода сбора данных, Национальный центр данных должен представлять еженедельный отчет о доле ответивших, генерируемый с помощью ПО GSS. Должны быть указаны сведения о завершенных интервью и доле ответивших лиц. Примеры выездной работы и отчетов о доле ответивших лиц в текущий момент представлены в документе «Руководство по программированию комплекса "General Survey System"».
- НЦД рекомендуется применять отчеты, полученные с помощью ПО GSS, и, по мере необходимости, вносить отчеты в системы Access^{1,2}, SPSS^{1,3}, SAS^{1,4}, STATA^{1,5} или другие программные продукты, которые будут содействовать анализу данных на предмет несоответствий, отклонений от нормы и неполноты. Для создания отчетов должны использоваться транспонированные агрегированные данные.

Несоответствия в данных могут быть обнаружены путем проверки достоверности переменных внутри и между вопросником для домохозяйств и индивидуальным вопросником. Для получения информации о проверке достоверности просьба обратиться к главе 5, посвященной обеспечению качества после сбора данных.

Процедуры резервного копирования: в зависимости от частоты агрегирования должно производиться регулярное, например еженедельное резервное копирование данных.

- Отдельные файлы db3 (уровень ВИ) подлежат архивации.

¹ Использование торговых наименований предназначено только для идентификации и не подразумевает одобрения со стороны Министерства здравоохранения и социального обеспечения или Центров по контролю и профилактике заболеваний.

² Microsoft Office Access© (Microsoft Corporation, Redmond, Washington).

³ SPSS© (SPSS Inc., Chicago, Illinois).

⁴ SAS© (SAS Institute Inc., Cary, North Carolina).

⁵ STATA© (Stata Corp., College Station, Texas).

- Кроме того, заархивированные данные должны пройти процедуру резервного копирования на какой-либо другой носитель, например сетевой или флеш-накопитель, и в двух разных физических местах должно храниться по одной копии.
- После окончания сбора данных должна проводиться сверка полученных данных с контрольным файлом для составления выборки, а также экспортированных файлов, полученных на местах.

4.2.3 Связь и техническая поддержка

Информирование со стороны НЦД во время сбора данных должно охватывать еженедельную отправку отчетов о положении дел в ВОЗ и CDC, а также информирование о любых и/или всех технических проблемах. Следует сообщать по электронной почте обо всех решенных или нерешенных технических проблемах, связанных с портативными устройствами или программным обеспечением, как можно скорее после их возникновения. Это будет способствовать оказанию своевременной технической поддержки, решению проблемы и обмену информацией.

5. Обеспечение качества: этап после сбора данных

Этап после сбора данных включает в себя ряд задач, которые должны быть выполнены в целях подготовки файла аналитических данных к проведению анализа, и относится к стадии, на которой все данные обследования собраны и объединены. В том числе имеется в виду подготовка данных для расчета веса выборки; оценка качества выборки, ошибки выборки и весов; а также измерение качества охвата структуры выборки, доли не ответивших и ошибок, не связанных с выборкой. Помимо этого, имеется в виду подготовка файла аналитических данных. В этой главе описаны применимые руководящие принципы и процедуры обеспечения качества, которые рекомендуются по каждому из этих видов деятельности, в частности:

- 5.1 Очистка и подготовка данных для расчета удельного веса элементов выборки;
 - 5.1.1 создание контрольного файла базы данных;
 - 5.1.2 удаление конфиденциальных идентификационных переменных;
 - 5.1.3 очистка и валидация объединенного файла данных;
 - 5.1.4 назначение окончательных диспозиционных кодов.
- 5.2 Показатели качества: выборка, ошибка выборки и удельные веса;
 - 5.2.1 модель калибровки постстратификационных весов с помощью корректировочных ячеек;
 - 5.2.2 мультипликативный эффект отклонений весов выборки в точности оценок обследования;
 - 5.2.3 общий эффект плана выборки в точности оценок обследования и внутрикластерная однородность в пределах ПЕО соответствующих основных оценочных показателей обследования;
 - 5.2.4 предел ошибки основных оценочных показателей обследования.
- 5.3 Показатели качества: охват, доля не ответивших и другие ошибки, не связанные с выборкой;
 - 5.3.1 уровень охвата домохозяйств основной выборки;
 - 5.3.2 модели досрочного прекращения интервью респондентами;
 - 5.3.3 модели формирования доли ответов домохозяйств по стратам первого этапа выборки;
 - 5.3.4 модели формирования доли ответивших на уровне отдельных лиц для переменных, используемых при корректировке на отсутствие ответа;
 - 5.3.5 модели формирования доли отказавшихся на уровне отдельных лиц для переменных, используемых при корректировке на отсутствие ответа;
 - 5.3.6 доля отсутствующих ответов на конкретные пункты для переменных показателей информационного бюллетеня.
- 5.4 Официальное рассмотрение качества статистических данных;
 - 5.4.1 документация по определению удельного веса элементов выборки.
- 5.5 Создание аналитического файла данных.
- 5.6 Создание файла данных для публичного использования.

5.1 Очистка и подготовка данных для расчета удельного веса элементов выборки

Этот раздел содержит рекомендации по слиянию файлов данных, валидации переменных и шаблонов пропуска данных, а также по созданию диспозиционных кодов с использованием пакета статистического программного обеспечения¹.

5.1.1 Создание контрольного файла базы данных

Функция "Generate Analysis Files" (создать файлы для анализа) позволяет объединить данные обследования в единый файл со значениями, разделенными запятыми, и формирует программы ввода данных в SAS, SPSS и STATA, которые обеспечивают чтение нового объединенного файла базы данных. Это может быть сделано с использованием ПО GSS.

Файлы, используемые утилитой "Generate Analysis Files", включают итоговый агрегированный файл "Master db3", который сформирован путем агрегирования файлов db3, полученных от всех интервьюеров, и двух баз данных вопросников ("Survey0.db3" и "Survey1.db3").

При использовании утилиты "Generate Analysis Files" необходимо убедиться в использовании актуальных месторасположений из агрегированного файла и баз данных вопросника обследования. Для внесения изменений необходимо обновить месторасположение или наименование файла в текстовом поле или выбрать кнопку «...», после чего система позволит выбрать существующий файл. Контрольный файл базы данных будет представлять собой файл со значениями, разделенными запятыми, в формате Unicode (UTF-8). Утилита также сгенерирует код ввода для чтения данных в SAS, SPSS и STATA.

Для просмотра файла CSV необходим текстовый редактор с поддержкой Unicode, например Wordpad. Файлы SAS, SPSS и STATA также имеют формат Unicode для обеспечения правильного отображения меток и форматов. Возможно, будет необходимо обновление программ SAS, SPSS и STATA перед их запуском, с тем чтобы убедиться в установке специфичных для страны параметров. В результате этого процесса будет создан контрольный массив данных. Созданный контрольный файл базы данных должен быть использован для удаления любых конфиденциальных идентификационных переменных показателей, а также должен быть предоставлен ИТ-координатору в стране.

5.1.2 Удаление конфиденциальных идентификационных переменных

Из созданного контрольного массива данных через раздел "Master Sample Selection" (определение контрольной выборки) необходимо удалить поля, содержащие конфиденциальные показатели.

5.1.3 Очистка и валидация объединенного файла данных

Убедитесь, что переменные имеют адекватные значения, схема пропуска была корректной и проведите проверку на предмет наличия пустых полей. Хотя многие из этих проверок качества данных осуществляются во время проведения интервью программным обеспечением портативного устройства, важно провести повторную проверку на случай невыявленных ошибок при программировании ПО портативных устройств.

¹ Координационный центр данных обследования GATS в CDC оказывает техническое содействие в использовании следующих пакетов статистических программ: SAS, SPSS и STATA.

1. Убедитесь в том, что схемы пропуска данных функционировали так, как предусмотрено вопросником. Исходные шаблоны пропуска данных, созданные на основе итогового утвержденного странового вопросника, должны быть проверены по всем основным переменным, и также рекомендуется проверить данные шаблоны по остальным переменным.
 - В случае, если респондент не ответил на вопрос из-за схемы пропуска, поле для данного вопроса останется незаполненным.
 - Например, если респондент отказался отвечать и на вопрос A02a (месяц рождения респондента) и на вопрос A02b (год рождения респондента), респондент должен ответить на вопрос A03 (возраст респондента), однако в случае, если предоставлены оба ответа как на вопрос A02a, так и на вопрос A02b, вопрос A03 необходимо пропустить.
 - Например, если в вопросе B01 респондент указывает, что он/она курит «ежедневно» (значение 1), он/она должен затем перейти к вопросу B04, и ответы на вопросы B02 и B03 должны отсутствовать.
 - В программе SAS пустое поле для числовой переменной равно «.» (для символьной переменной это «»). В SAS данная схема пропуска может быть проверена с использованием кода, указанного ниже. Любые записи, которые не соответствуют схеме пропуска, будут выведены в массив данных SAS.


```
IF B01 = 1 & (B02 NE . OR B03 NE .) THEN FLAG=1;
ELSE IF B01 = 2 & (B02 = . OR B03 = .) THEN FLAG =1;
ELSE IF B01 = 3 & (B02 NE . OR B03 = .) THEN FLAG=1;
ELSE FLAG=0;
IF FLAG=1 THEN OUTPUT; *Значение, равное «1» для FLAG, означает, что схема пропуска не функционировала, тогда как значение, равное «0», свидетельствует о функционировании схемы.
```
2. Необходимо убедиться, что не заполнены только те поля, которые были пропущены в связи с функционированием схемы пропуска (например, значение поля не применимо для респондента и должно быть пустым). Любые другие пустые значения недопустимы.
 - Поля, которые были пропущены в связи с функционированием схемы пропуска, будут необходимо преобразовать в новые значения. Таким образом, они будут отличаться от пустых полей, которые не связаны с функционированием схемы пропуска (недействительные пустые поля). Так как не допускается перезапись поверх исходных значений переменных, для завершения этого шага следует создать новый набор данных.
 - В случае, если вопрос был неприменим (пропущен правильно), для недостающего значения в SAS необходимо использовать код «.S». Код SAS для схемы пропуска, приведенный выше, может быть изменен следующим образом:


```
IF B01 = 1 & (B02 = . OR B03 = .) THEN B02 = .S & B03 = .S;
ELSE IF B01 = 3 & B02 = . THEN B02 = .S;
```
 - После того как все значения схемы пропуска заменены, иных пустых полей оставаться не должно.

В случае, если остаются другие пустые поля, эти записи необходимо вывести в файл ошибок.

3. Проверьте каждую переменную, с тем чтобы убедиться в отсутствии недействительных значений. Используйте справочник данных по стране (книгу кодов) для вывода действительных значений по каждой переменной в массиве данных.
 - Убедитесь, что значения ответов на каждый вопрос являются действительными.
 - Например: единственными действительными ответами на вопрос B01 являются: 1 («ежедневно»), 2 («не каждый день»), 3 («совсем нет»), -9 («отказ») или код недостающего значения, который применяется в конкретном программном обеспечении (например, «.S» при использовании SAS).
 - Например: для некоторых переменных существуют диапазоны значений, такие, что значения меньше или больше этого диапазона недопустимы. Так, для переменной B04, самое низкое возможное значение должно составлять 0, и самое высокое возможное значение должно составлять 98. Кроме того, существует вариант значения -9, если был дан ответ «неизвестно» или «отказ от ответа». Код пропущенного значения, специфичный для использованного программного обеспечения (например, «.S» при использовании ПО SAS) также будет действительным значением.
 - *В ПО SAS записи с недействительными ответами для B01 могут быть выделены и выведены в файл ошибок с использованием следующего кода:*

```
IF B01 NOT IN(1,2,3,-7,-9) THEN OUTPUT.
```
 - *Вместо этого по предпочтению аналитика для каждой переменной может быть выполнено простое определение частот, что выявит наличие значений за пределами диапазона. Любые записи с недействительными значениями впоследствии могут быть определены и выведены в файл ошибок.*
 - Проверка возраста респондента
 - Перед проверкой правильности указанного возраста респондента данный возраст необходимо предварительно вычислить и включить в массив данных. Для расчета следует использовать дату проведения интервью по индивидуальному вопроснику (IQ_EVENTDATE) и месяц (A02a), а также год (A02b) рождения респондента, с предполагаемым днем рождения 15 числа данного месяца. В случае ответа «неизвестно» или «отказ от ответа» на вопрос A02a или A02b переходят к вопросу A03 и просят респондента указать возраст (если на вопрос A02a или A02b нет ответа, следует получить ответ только на вопрос A03). Если месяц или год рождения не сообщается, для определения возраста респондента следует использовать данные ответа на вопрос A03. Если месяц и год сообщены, следует подсчитать возраст путем вычитания даты рождения из даты проведения интервью. Убедитесь, что возраст респондента является допустимым (действительный диапазон составляет от 15 до 125).

- В ПО SAS возраст может быть подсчитан и затем проверен с помощью следующего кода:

```
IF A02A <=12 AND A02B NE -7 AND A02B NE -9 THEN DO;
  BIRTHDATE = MDY(A02A, 15, A02B);
  AGE = FLOOR(YRDIF(BIRTHDATE, IQ_EVENTDATE, 'ACTUAL'));
  END;
  ELSE AGE = A03.
```

4. Другие способы проверки действительности данных

- Необходимо выполнить проверку на предмет того, что, если любая из следующих переменных не были пропущена, их величины меньше возраста респондента: B04, B05, B08, B09, B11, B12, B13, C04, C05, C08, C09, C11, C12 и C13 (в зависимости от размерности переменной данные значения сначала может быть необходимо пересчитать в годы).

В каждом случае необходимо определить окончательные диспозиционные коды. Итоговые диспозиционные коды используются для вычисления и учета доли ответивших лиц и мер для обеспечения качества.

5.1.4 Назначение окончательных диспозиционных кодов

Объединенный файл данных будет содержать коды окончательного результата, назначенные ВИ по всем случаям в выборке, по которым проводилась работа во время сбора данных (см. описание кодов результатов выездной работы в документе «Руководство Выездного интервьюера обследования GATS»). Каждый случай в интервью для домохозяйств и индивидуальном интервью должен иметь один код окончательного результата (см. раздел 4.1.3 о назначении окончательного кода незавершенным интервью во время сбора данных).

С использованием кодов окончательного результата должны быть назначены окончательные диспозиционные коды. Окончательные диспозиционные коды впоследствии используются для вычисления и учета доли ответивших лиц и мер для обеспечения качества. Информацию о назначении окончательных диспозиционных кодов на основании кодов окончательного результата, сведения о вычислении доли ответивших лиц, а также шаблоны учета этих результатов по итогам сбора данных см. в приложении С.1.

Ниже представлены важные руководящие указания для данного процесса, которые необходимо соблюдать.

1. На основании кода окончательного результата по итогам выездных работ с использованием вопросника для домохозяйств за каждым случаем должен быть закреплен один окончательный диспозиционный код по вопроснику для домохозяйств. Если интервью для домохозяйства было завершено и для проведения индивидуального интервью был выбран один человек (окончательный диспозиционный код 1), за вопросником для домохозяйства также необходимо закрепить окончательный диспозиционный код. Если окончательный диспозиционный код вопросника для домохозяйств отличается от 1 (например, не был выбран ни один человек, отказ от ответа, интервью не завершено), окончательный диспозиционный код для вопросника для домохозяйств назначен не будет, так как интервью по индивидуальному вопроснику не проводилось.

2. В индивидуальном вопроснике имеются основные вопросы, которые определяют правильность данного интервью. Если по любому из этих вопросов указан ответ «неизвестно» или «отказ от ответа», интервью не считается правильным. Таким образом, для всех случаев, в которых проводилось интервью по индивидуальному вопроснику с кодом окончательного результата 400 (интервью *по индивидуальному вопроснику* завершено), необходимо соблюдать указанные ниже правила, приведенные для того, чтобы определить необходимый диспозиционный код.
 - Если в ответах на вопросы B01, B02 или B03 (основные вопросы о распространенности курения) указано «неизвестно» или «отказ от ответа», следует указать окончательный диспозиционный код «16» (выбранный респондент некомпетентен). Это не должно быть обычным явлением (кроме того, данное правило может применяться для вопросов C01, C02 и C03 (основные вопросы о распространенности бездымного табака) только в случае, если употребление бездымного табака в стране широко распространено).
 - Всем другим случаям интервью по индивидуальному вопроснику с кодом окончательного результата 400 (*интервью по индивидуальному вопроснику* завершено) должен быть присвоен окончательный диспозиционный код «11» (интервью по индивидуальному вопроснику завершено). Это будет происходить в большинстве случаев.
3. Для определения необходимого диспозиционного кода во всех случаях, в которых проводилось интервью по индивидуальному вопроснику с кодом окончательного результата 402 (интервью по *индивидуальному вопроснику* завершено частично), необходимо соблюдать приведенные ниже правила.
 - Если индивидуальный вопросник был пройден хотя бы до вопроса E01, и ни на один из основных вопросов (описанных выше в пункте 2) не было дано ответа «неизвестно» или «отказ от ответа», необходимо назначить окончательный диспозиционный код «11» (*интервью по индивидуальному вопроснику* завершено).
 - Если интервью было прервано до вопроса E01, должен быть назначен окончательный диспозиционный код «12» (не завершено). (Случаи с окончательным диспозиционным кодом «12» (не завершено) учитываются как случаи отсутствия ответов на вопросы обследования GATS, и эти данные не включаются в анализ).
 - Если индивидуальный вопросник был пройден хотя бы до вопроса E01, и на любой из основных вопросов (описанных выше в пункте 2) был дан ответ "неизвестно" или "отказ от ответа", должен быть назначен окончательный диспозиционный код «16» (выбранный респондент некомпетентен) (как описано выше в пункте 2, эти респонденты рассматриваются как некомпетентные в случае, если они не могут предоставить допустимых ответов на основные вопросы и, таким образом, данные их интервью не включаются в анализ).
4. В окончательный набор аналитических данных должны быть включены только случаи с окончательным диспозиционным кодом «11». (Таким образом, правильное назначение диспозиционных кодов имеет важное значение.)

5. Для сверки всех кодов окончательного результата с диспозиционными кодами и выявления возможных ошибок классификации необходимо использовать перекрестные таблицы. В случае, если два кода не совпадают так, как ожидалось, это указывает на проблемы программного кода, используемого для формирования диспозиционных кодов.

5.2 Показатели качества: выборка, ошибка выборки и удельные веса

В этом разделе описаны расчеты для непосредственного определения качества оценок выборки GATS и определения влияния кластеров выборки и неравных весов выборки на эти оценки. Кроме того, включены рекомендации по проверке точности рассчитанного удельного веса элементов.

5.2.1 Модель калибровки постстратификационных весов с помощью корректировочных ячеек

В ходе заключительного этапа подсчета удельного веса элементов выборки проводится калибровка весов относительно численности населения по известным коррелятам основных показателей результата исследования, называемых калибровочными переменными (например, пол, образование, возраст, проживание в городской/сельской местности и регион согласно «Руководству по определению удельного веса элементов выборки обследования GATS»).

Общие сведения и указания для вычисления постстратификационных корректировок приведены ниже в этом разделе. Подробные сведения по данной теме представлены в приложении С.2.

1. **Общие сведения.** Калибровка позволяет корректировать дисбаланс выборки, который не затронут поправкой на неполученные ответы. Ко всем позициям заблаговременно подготовленных корректировочных ячеек применяются разные корректировки. Необходимо увеличить удельный вес подгрупп населения, которые остаются недостаточно представленными, и снизить вес подгрупп, которые по-прежнему представлены избыточно. Чем больше значения калибровки отличаются от 1,00 (в большую или меньшую сторону), тем более выраженными будут потенциальные последствия дисбаланса выборки (помимо вызванных внесением поправок на неполученные ответы) в плане ошибки оценочных величин.
2. **Проведение постстратификационных корректировок.** С процедурной точки зрения калибровка в виде постстратификации охватывает формирование «корректировочных ячеек» путем перекрестной классификации величин коррелятов. Значение «постстратификационной корректировки» в каждой из этих корректировочных ячеек будет менее 1 в случае, если подвыборка данной категории чрезмерно представлена в выборке после поправок на проведение выборки и долю неполученных ответов, и данное значение будет более 1 в тех категориях, в которых подвыборка представлена недостаточно.
3. **Отчетность о постстратификационной корректировке.** Необходимо создать таблицу, в которой перечислены все корректировочные ячейки, с указанием для каждой из них того, каким образом ячейка определяется в терминах категорических переменных, используемых для калибровки. Для каждой ячейки необходимо указать расчетную величину постстратификационной корректировки и сравнить ее размер с 1,00. Оптимальная ситуация, при которой все значения постстратификационной корректировки близки к 1, некоторые – несколько больше 1 и остальные – несколько меньше 1,00.

5.2.2 Мультипликативный эффект отклонений весов выборки в точности оценок обследования

«Рекомендации по проведению инструктажа в сфере работы с данными GATS» предусматривает такой план выборки, в котором вероятности выбора (и, таким образом, веса выборки) будут несколько различаться в связи с использованием расчетных величин размеров кластеров, корректировок весов выборки и в связи с равным распределением размеров выборки между регионами, когда требуются региональные оценки, соответствующие стандартам точности GATS. В Руководстве по определению удельного веса элементов выборки для обследования GATS описано, каким образом необходимо вычислять эти веса. После очистки данных вопросника и присвоения окончательных значений удельного веса элементам выборки данные выборки обследования GATS будут готовы для анализа и выполнения этой задачи.

Общие сведения и указания для вычисления данного эффекта приведены ниже в этом разделе. Подробные сведения по этой теме представлены в приложении С.3.

- 1. Общие сведения.** Вариация удельных весов может приводить к повышению значения ошибки выборки в оценках обследования и, таким образом, вызвать большую вариацию и стандартную ошибку данных оценок. В частности, многократное увеличение дисперсии оценок обследования, определяемое как $Meff_w$, зависит от того, насколько изменчивыми являются удельные веса серии наблюдений выборки, которые используются для получения оценок.

Чем более переменны значения удельных весов, тем больше величина $Meff_w$.

- 2. Оценка $Meff_w$** – Для того, чтобы вычислить $Meff_w$, необходимы среднее арифметическое и дисперсия весов для серии наблюдений выборки.
 - Значение $Meff_w$ для оценок, полученных из этого набора наблюдений выборки, определяют путем вычисления соотношения дисперсии и квадрата среднего арифметического с последующим добавлением единицы к этому соотношению. В качестве формулы для вычислений используется уравнение (1), представленное в *добавлении С.3*.
- 3. Учет $Meff_w$** – Так как $Meff_w$ применяется ко всем оценкам, полученным на основе конкретной серии наблюдений выборки, которые определяют категорию учета оценок обследования (например, общая численность населения, население сельских районов, образование, возраст, пол и т. п.),
 - необходимо вычисление и учет для всех основных подгрупп населения, по которым будут вычисляться и учитываться оценочные величины обследования GATS. Это может быть сделано в виде таблицы со списком подгрупп отчетности и соответствующих значений $Meff_w$.
- 4. Интерпретация $Meff_w$.** Надлежащая интерпретация значения $Meff_w$, вычисленного для определенной категории учета:

Предпочтительно значение $Meff_w < 2,00$.

- «вариация удельных весов элементов выборки привела к повышению величины дисперсии всех оценок (для конкретной категории учета) на величину (M_{eff_w})».
- В случае, если $M_{eff_w} < 2,00$ принятие мер для уменьшения вариации удельных весов не требуется.

5. **Пример.** Например, при оценках в отношении населения сельских районов $M_{eff_w} < 1,62$

«вариация удельных весов элементов выборки привела к повышению величины дисперсии всех оценок для жителей сельских районов на величину 1,62».

- Значение $M_{eff_w} = 1,62$ рассматривается как достаточно низкое для данных оценок.

5.2.3 Общий эффект плана выборки в точности оценок обследования и внутрикластерная однородность в пределах ПЕО соответствующих основных оценочных показателей обследования

«Рекомендации по проведению инструктажа в сфере работы с данными GATS» предусматривает такой план выборки, при котором выбираются кластеры респондентов обследования, тогда как порядок вычисления удельного веса элементов выборки описан в «Руководстве по определению удельного веса элементов выборки обследования GATS». После очистки данных вопросников и вычисления окончательных весов элементов выборки, данные выборки GATS будут готовы к анализу перед опубликованием выводов.

Общие сведения и указания для вычисления общего эффекта плана выборки в оценке обследования и внутрикластерной однородности в рамках ПЕО приведены ниже в этом разделе. Подробная информация по этой теме представлена в приложении С.4.

1. **Общие сведения:** дисперсия оценочной величины ($\hat{\theta}$) сложной выборки, отнесенная к дисперсии сопоставимой оценочной величины, полученной на основе простой случайной выборки аналогичного размера, называется общим «эффектом плана выборки» для $\hat{\theta}$ или $Deff(\hat{\theta})$. Существует два мультипликативных компонента $Deff(\hat{\theta})$: эффект плана выборки, обусловленный кластерной выборкой, $Deff_{cs}(\hat{\theta})$, и мультипликативный эффект различий весов выборки, M_{eff_w} (см. раздел 5.2.2). Кроме того, $Deff_{cs}(\hat{\theta})$ непосредственно связан с количественными характеристиками однородности внутри ПЕО, ρ_{θ} , и средним размером выборочного кластера.
2. **Оценка $Deff(\hat{\theta})$ и ρ_{θ}** – Оценка $Deff(\hat{\theta})$ может быть получена непосредственно при использовании определенных статистических программных пакетов (например, SUDAAN) или вычислена на основе оценочных величин и дисперсии в случае, если оценка выражена в форме пропорции или доли. Так как $Deff(\hat{\theta})$ зависит от $Deff_{cs}(\hat{\theta})$, ρ_{θ} и среднего размера кластера выборки, оценка ρ_{θ} может быть получена путем решения уравнения для данного параметра с использованием предоставленных оценок $Deff(\hat{\theta})$ и других перечисленных показателей.
 - Подробные формулы для вычисления $Deff(\hat{\theta})$ в ситуациях, когда ПО для статистического анализа этого не позволяет, а также для оценки ρ_{θ} в любых условиях

изложены в уравнениях (1) и (2) в приложении С.4. Полученные оценочные величины обозначаются как $d_{eff}(\hat{\theta})$ и $\hat{\rho}_{\theta}$ соответственно. Авторами используется $d_{eff_{CS}}(\hat{\theta})$ для обозначения оценочной величины $Deff_{CS}(\hat{\theta})$

- Так как значения $d_{eff}(\hat{\theta})$ и $\hat{\rho}_{\theta}$ многочисленны, их необходимо обобщить путем указания медианного, минимального и максимального значений.
3. **Учет $Deff(\hat{\theta})$ и ρ_{θ}** – Необходимо отразить оценочные значения $Deff(\hat{\theta})$ и ρ_{θ} для всех основных единиц измерения (см. определения в приложении С.4).
4. **Интерпретация $Deff(\hat{\theta})$ и ρ_{θ}** – Надлежащая интерпретация ориентировочных значений $d_{eff}(\hat{\theta})$ и ρ_{θ} :
- «Дисперсия полученного в обследовании ориентировочного значения (популяционной характеристики θ) с учетом схемы выборки обследования GATS в ($d_{eff}(\hat{\theta})$) раз выше, чем следовало бы ожидать при получении аналогичного количества респондентов методом простой случайной выборки».
 - Предпочтительны значения эффекта плана выборки в оценке или $Deff(\hat{\theta})$ на уровне менее 2,50.
 - Величина полной вариации в показателях обследования (связано с θ), которая не была учтена внутри ПЕО: $\hat{\rho}_{\theta}$.
 - В редких случаях $\hat{\rho}_{\theta} > 0,15$; или отрицательное значение; чаще $\hat{\rho}_{\theta} < 0,05$ либо $\hat{\rho}_{\theta} < 0,01$.
5. **Пример** – В ходе анализа для оценки текущей распространенности курения, P , на основе выборки, в которой средний размер кластера выборки составил 20, получено: $d_{eff}(\hat{\theta}) = 1,92$, $Meff_w = 1,36$.

В редких случаях
 $\hat{\rho}_{\theta} > 0,15$;
 $\hat{\rho}_{\theta} < 0,05$;
 $\hat{\rho}_{\theta} < 0,01$.
чаще
либо

«Дисперсия полученного в обследовании ориентировочного значения текущей распространенности курения с учетом схемы выборки обследования GATS в 1,92 раза выше, чем следовало бы ожидать при получении аналогичного количества респондентов методом простой случайной выборки».

- Вычисленное значение $\hat{\rho}_{\theta} = 0,022$ является типичным.

5.2.4 Предел ошибки основных оценочных показателей обследования

Существует несколько способов учета статистической точности оценок обследования. Одним из них является оценка предела ошибки. В ходе обследования GATS рекомендуется учитывать ожидаемую ошибку (МОЕ) наряду с оценками основных показателей обследования.

В «Руководстве по составлению плана выборки обследования GATS» описано три основных характеристики выборок GATS, которые будут влиять на статистическое качество оценок и выводов, сделанных на основе данных обследования. К этим характеристикам относятся применение стратификации, выбор кластеров членов целевой группы населения и выбор

представителей населения с неравной вероятностью (что требует применения весов выборки при анализе). Для учета этих характеристик при анализе данных обследования GATS могут использоваться различные пакеты аналитического программного обеспечения.

Общие сведения и указания по вычислению представлены в этом разделе. Подробные сведения по данным темам представлены в приложении С.5.

1. **Предел ошибки специфичен для каждого показателя.** Каждый показатель, полученный на основе данных обследования GATS, характеризуется специфичным пределом ошибки. По определению предел ошибки является ориентировочной половинной шириной доверительного интервала оценки основного показателя обследования. Другим способом интерпретации предела ошибки является близость данной оценки к реальному показателю обследования в популяции.
2. **Основные показатели обследования.** Для различных поведенческих аспектов/темпов потребления табака должны быть получены оценки изменения за период от одного раунда до другого, а также специфичные для конкретного раунда обследования GATS оценки. Эти оценки могут быть учтены для населения в целом, а также по различным отчетным подгруппам населения, таким как подгруппы, выделенные по критерию пола, проживания в городской/сельской местности и по возрастным категориям. В случае, если общий размер региональных выборок составляет не менее 8000, могут быть учтены оценочные величины на региональном уровне.

Следующие три уровня были расценены как наиболее важные для планирования:

- *показатель распространенности потребления табака* - процентная доля лиц, которые пользуются табачной продукцией в настоящее время;
 - *показатель пассивного курения* - процентная доля лиц, подверженных воздействию дыма при курении окружающих; и
 - *показатель отказа от табака* - процент нынешних курильщиков, которые предприняли попытку отказа от курения.
3. **Оценка предела ошибки** – как правило, оценка предела ошибки статистическим ПО не производится, однако информация для данных вычислений зачастую доступна. Для вычисления и интерпретации MOE необходимы три элемента: (i) оценочная величина показателя обследования, (ii) оценочная величина стандартной ошибки (или дисперсии) для данного показателя и (iii) конкретный показатель, связанный с желаемой статистической достоверностью значения MOE .

Уровень достоверности обычно основывается на значении (Z) стандартного нормального распределения.

Например, для уровня достоверности 95% может быть использована величина $Z = 1,96$.

Предел ошибки вычисляют как произведение показателя желаемой достоверности и стандартной ошибки оценки. Фактическая формула для вычислений представляет собой уравнение (1) в приложении С.5.

Рекомендуется вычисление предела ошибки для всех оценок, касающихся основных показателей обследования.

4. **Учет** МОЕ – рекомендуется вычисление МОЕ для всех оценок, касающихся основных показателей обследования. Сюда входят общенациональные оценки этих показателей, а также оценки этих показателей для всех важных подгрупп отчетности (например, по полу, условиям проживания в городской/сельской местности, региону и возрасту).
- Основные оценки обследования и соответствующие им значения МОЕ должны быть представлены совместно. Информация о представлении оценок ошибок выборки и пример шаблонов таблиц с основными показателями обследования GATS и рекомендуемыми значениями отчетности представлены в приложении С.6.
5. **Интерпретация.** При учете наряду с оценочной величиной МОЕ показывает, насколько оценка близка к реальному значению данного показателя в популяции.

При использовании $Z=1,96$ для вычисления МОЕ оценка обследования и значение предела ошибки в паре могут быть интерпретированы следующим образом:

«имеется 95% степень достоверности того, что оценка (ЗНАЧЕНИЕ ОЦЕНКИ) находится в пределах (ЗНАЧЕНИЕ МОЕ) соответствующего значения среди населения».

6. **Пример.** Пусть учтенная величина оценки по обследованию GATS составляет 43,4%, а стандартная ошибка 1,5611%, что было вычислено в соответствии с фактическим планом выборки в стране.

Значение МОЕ для этой оценки составит $3,1\%=(1,96)(1,5611\%)$.

Интерпретация:

«имеется 95% степень достоверности того, что оценка 43,4% находится в пределах 3,1% соответствующего значения среди населения».

5.3 Показатели качества: охват, доля не ответивших и другие ошибки, не связанные с выборкой

5.3.1 Уровень охвата домохозяйств рамкой выборки

Как описано в «Руководстве по составлению выборки для обследования GATS», предпоследний этап составления выборки охватывает отбор домохозяйств из числа всех перечисленных в ПЕО или сегментах. Таким образом, качество выборки домохозяйств и, в конечном счете, оценок GATS зависит от полноты этих списков домохозяйств, которые могут быть получены различными способами.

Общие сведения и указания по вычислению охвата домохозяйств рамкой выборки приведены ниже в этом разделе. Подробная информация по этой теме представлена в приложении С.7.

1. **Общие сведения.** Основной частью процесса составления выборки в обследовании домохозяйств является формирование перечня домохозяйств на стадии выборки, в котором выбираются адреса. Могут применяться различные подходы к формированию перечня домохозяйств (например, составление обычного перечня домохозяйств выездными работниками, модифицированные перечни избирателей). Обычно в итоге получают неполный перечень домохозяйств и, таким образом, неполный охват конечной

выборки домохозяйств. Так как эффект ошибки, вызванной неполнотой перечня, зависит от: (i) совокупной разницы между основными показателями обследования по домохозяйствам, охваченным и не охваченным данной рамкой, так и от (ii) степени неполноты охвата (или 100 процентов минус уровень охвата домохозяйств) в выборке, целесообразно оценить уровень охвата домохозяйств рамкой выборки.

- Уровень охвата рамкой выборки определяется как количество домохозяйств, учтенных в данной рамке выборки, отнесенное к количеству домохозяйств в популяции.
 - Хотя они необходимы для описания охвата домохозяйств рамкой выборки, данные об общем количестве домохозяйств всего населения из недавней переписи населения часто отсутствуют. Если они имеются, должен быть вычислен уровень охвата, однако это не является принципиально важным элементом данных для официального рассмотрения статистического качества показателей GATS.
2. **Оценка уровня охвата домохозяйств рамкой выборки.** Числитель для данного показателя может быть вычислен с помощью суммы удельных весов до корректировки (базисных) для всех выбранных домохозяйств. Знаменатель должен быть получен из актуального и заслуживающего доверия источника (например, недавняя перепись населения или другое недавнее обследование, размер выборки которого больше таковой в GATS).
 3. **Учет уровня охвата домохозяйств рамкой выборки.** Вычисленный уровень охвата рамкой выборки должен быть отражен в Страновом отчете обследования GATS с описанием плана выборки или с отдельной оценкой качества данных.
 4. **Интерпретация уровня охвата домохозяйств рамкой выборки.** Уровень охвата домохозяйств рамкой выборки показывает степень такого охвата, но не величину ошибки, так как последнее зависит от суммы изначальной и совокупной разницы (в основных показателях обследования) между охваченными и неохваченными домохозяйствами. Приемлемым будет считаться уровень охвата домохозяйств свыше 95%.
 5. **Пример.** Оптимальная из имеющихся внешних оценок числа домохозяйств в целевой группе населения страны GATS составляет 36 436 103, тогда как взвешенная сумма 13 000 выбранных домохозяйств для обследования GATS составляет 34 788 543. Приемлемым будет считаться итоговый уровень охвата домохозяйств, который

$$\text{составляет } 0,955 = \frac{34,788,543}{36,436,103}$$

Приемлемым будет считаться уровень охвата домохозяйств свыше 95%.

5.3.2 Модели показателей досрочного прекращения интервью респондентами

Некоторые интервью завершаются не полностью. Респондент может посчитать участие в интервью дополнительным бременем и в связи с этим отказаться от дальнейшего общения с интервьюером, или, вероятно, что респонденту будет необходимо окончить интервью раньше по уважительным причинам, и он (она) не сможет или не захочет вносить изменения в свой график для завершения интервью. В любом случае частично завершенное интервью является показателем потери интереса со стороны респондента, что может рассматриваться как

характеристика качества работы интервьюера или отношения респондента к теме обследования, что, в конечном счете, скажется на качестве данных.

- **Источник(и) данных:** для данных расчетов необходимо использовать файл данных с окончательными диспозиционными кодами для домохозяйств/отдельных лиц (см. приложение С.1). Диспозиционный код «11» индивидуального вопросника означает, что респондент закончил по крайней мере разделы А–D индивидуального интервью. Диспозиционный код «12» индивидуального вопросника указывает, что интервью было закончено до завершения работы по разделам А–D.
- **Метод/вычисление:** укажите, что «респондент» обследования – это любой выбранный человек, за которым закреплен окончательный диспозиционный код индивидуального вопросника «11» или «12». Кроме того, необходимо указать, что уровень досрочного прекращения интервью респондентами (COR) составит:

$$COR = \frac{R_c}{R} = \frac{\text{Количество респондентов обследования GATS, которые завершили интервью частично}}{\text{Количество респондентов обследования GATS}}$$

или

$$= \frac{\text{Диспозиционный код 12}}{\text{Диспозиционные коды (11 + 12)}}$$

- **Использование:** значения COR должны быть вычислены непосредственно по выборке в качестве заключительного этапа обеспечения качества после завершения сбора данных. Более того, на протяжении всего периода сбора данных НР должны постоянно проводить работу по выявлению и принятию мер в отношении неэффективных ВИ (с использованием в ходе вычислений соответствующих кодов окончательного результата по итогам выездных работ). В том числе для этого будет необходимо вычисление COR по отдельным интервьюерам или группе интервьюеров и выявление тех интервьюеров/групп интервьюеров, у которых эти уровни неприемлемо высоки (например, уровни превышающие среднее значение плюс два стандартных отклонения уровней среди всех остальных интервьюеров/групп). Кроме того, значения COR могут быть вычислены по: (1) неделе сбора данных, во время которой проводились интервью, (2) возрасту респондента, (3) полу респондента и (4) критерию проживания респондента в городской/сельской местности.
- **Интерпретация:** как правило, предпочтителен как можно более низкий уровень COR. Снижение величины COR на протяжении периода сбора данных может означать меньшую мотивацию респондентов к предоставлению полезных данных для обследования GATS к концу периода сбора данных. Хотя уровни досрочного прекращения интервью респондентами являются полезным показателем проведения сбора данных и общего качества обследования, их вычисление для официального рассмотрения статистического качества показателей GATS не имеет решающего значения.

5.3.3 Модели формирования доли ответов домохозяйств по стратам первого этапа выборки

Доля ответов домохозяйств вычисляется как отношение взвешенного количества домохозяйств, которые представили полный перечень данных, ко взвешенному количеству охваченных обследованием домохозяйств (см. формулу 19 в «Руководстве по определению удельного веса элементов выборки обследования GATS»). Домохозяйства, которые предоставили частичные данные, рассматриваются как отказавшиеся или прервавшие интервью и не включаются в числитель при вычислении доли ответивших домохозяйств. Опрос по *вопроснику для домохозяйств* считается завершенным в случае, если ему присвоен окончательный диспозиционный код «1» (см. приложение С.1). При условии учета только охваченных обследованием домохозяйств доля ответов на уровне домохозяйств вычисляется следующим образом:

$$\text{Доля ответивших на уровне домохозяйств} = \frac{[1]}{[1] + [3] + [4] + [5] + [6] + [9]} \times 100 \quad ,$$

где

- 1= Вопросник для домохозяйства завершен, выбран один человек;
- 2= Вопросник для домохозяйства завершен, не выбран ни один человек;
- 3= Вопросник для домохозяйства завершен частично, реестр не заполнен (интервью не завершено);
- 4= Вопросник для домохозяйства не заполнен, невозможно установить подходящего по критериям респондента для скрининга;
- 5= Никого нет дома;
- 6= Отказ домохозяйства;
- 9 = Другая причина отсутствия ответов домохозяйства.

Необходимо заметить, что окончательный диспозиционный код «2» для домохозяйства исключен как из числителя, так и из знаменателя формулы для подсчета доли ответивших лиц на уровне домохозяйств, так как эти домохозяйства считаются не соответствующими критериям участия. Эта доля ответивших соответствует уровню RR1 согласно определению Американской ассоциации изучения общественного мнения (AAPOR) (<https://www.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx>).

Страны должны предоставить полную документацию о вычислении доли ответов домохозяйств. Доли ответов домохозяйств должны быть представлены по всей стране в целом в разбивке по стратам первого этапа выборки (то есть проживание в городской/сельской местности и регионе), в случае, если размеры выборки достаточны для региональной оценки. Дальнейшая разбивка, при необходимости, должна быть сделана с помощью перекрестной классификации по критериям проживания в городской/сельской местности и регионе.

5.3.4 Модели формирования доли ответивших на уровне отдельных лиц для переменных, используемых при корректировке на отсутствие ответа

Доля ответов на уровне отдельных лиц вычисляется как соотношение взвешенного количества частично или полностью завершённых интервью GATS к взвешенному количеству лиц, отвечающих критериям участия в выборке, выбранных из реестра всех членов домохозяйства.

В частности, окончательные диспозиционные коды на уровне отдельных лиц следующие:

- 11= индивидуальный вопросник заполнен;
- 12= незавершённое интервью;
- 13= выбранный человек, как было установлено впоследствии, не отвечает критериям участия в обследовании GATS;
- 14 = выбранного респондента нет дома;
- 15 = отказ выбранного респондента;
- 16= выбранный респондент некомпетентен;
- 17 = другие причины отсутствия ответа от респондента.

Один из способов вычисления доли ответивших на уровне отдельных лиц выглядит следующим образом.

$$\text{Доля ответивших на уровне отдельных лиц} = \frac{[11]}{[11] + [12] + [14] + [15] + [16] + [17]} \times 100$$

При вычислении доли ответивших на уровне отдельных лиц, указанной выше, предполагается, что все выбранные лица, соответствие которых критериям участия неизвестно (т. е. окончательный диспозиционный код «14»), имеют право на участие в обследовании GATS. Это может привести к недооценке доли ответивших на уровне отдельных лиц в случае, если интервьюеры будут часто выбирать лиц, в отношении которых после начала интервью будет выявлено несоответствие критериям участия в обследовании. Странам рекомендуется оценивать долю выбранных из реестра респондентов, которые в действительности отвечают критериям участия в обследовании GATS, (e) с применением взвешенных частот диспозиционных кодов:

$$e_{\text{person-level}} = \frac{[11] + [12] + [15] + [16] + [17]}{[11] + [12] + [13] + [15] + [16] + [17]}$$

Если данная доля для выборки в целом составит менее 0,90, страны должны внести поправки на неизвестный компонент в доле ответивших лиц путем умножения количества лиц, соответствие которых критериям участия неизвестно, (окончательный диспозиционный код «14») на эту долю (e). Следующая формула соответствует доле ответивших лиц RR3 по AAPOR (2004).

$$\text{Доля ответивших на уровне отдельных лиц} = \frac{[11]}{[11] + [12] + [e_{\text{person-level}}[14]] + [15] + [16] + [17]} \times 100$$

Данная доля ответивших соответствует уровню RR3 по определению AAPOR, как указано выше.

Страны должны предоставить полную документацию о вычислении доли ответов.

Доли ответивших на уровне отдельных лиц должны быть представлены по всей стране в целом, а также по более мелким административным единицам районов в случае, если размеры выборки достаточны для региональной оценки. Дальнейшая разбивка в рамках региона (если применимо) должна быть сделана по тем переменным, которые использовались для определения корректировки на неполучение ответов. В эту категорию могут входить параметры проживания в городской/сельской местности и возраста, указанного в перечне (15–24, 25–34, 35–44, 45–54, 55+), пола (мужской, женский) и отношения к курению в данный момент (курящий, некурящий).

В «Руководстве для полевого супервайзера обследования GATS» изложены обязанности НР по рассмотрению и принятию мер при низкой доле ответивших лиц в ходе работы группы интервьюеров.

5.3.5 Модели формирования доли отказавшихся на уровне отдельных лиц для переменных, используемых при корректировке на отсутствие ответа

Доля отказавшихся вычисляется как отношение количества выбранных респондентов, отвечающих критериям участия, которые отказались от интервью по *индивидуальному вопросу* обследования GATS (окончательный диспозиционный код «15»), к общему взвешенному количеству завершенных (11) и незавершенных (12) индивидуальных вопросников GATS в сумме с количеством не ответивших респондентов, соответствие которых критериям участия неизвестно (14), количеством отказавшихся (15), некомпетентных респондентов (16) и других респондентов, которые не ответили по иным причинам (17). Порядок вычислений с использованием диспозиционных кодов индивидуального уровня:

$$\text{Доля отказавшихся на уровне домохозяйств} = \frac{[15]}{[11] + [12] + [14] + [15] + [16] + [17]} \times 100$$

Эта доля отказавшихся от ответа соответствует уровню REF1 по определению Американской ассоциации изучения общественного мнения (AAPOR).

Доли отказа на уровне отдельных лиц должны быть вычислены непосредственно по выборке в качестве заключительного этапа обеспечения качества после окончания сбора данных. Более того, на протяжении всего периода сбора данных НР должны постоянно проводить работу по выявлению и принятию мер в отношении неэффективных ВИ. В том числе для этого будет необходимо вычисление доли отказавшихся лиц по отдельным интервьюерам или группе интервьюеров и выявление тех интервьюеров/групп интервьюеров, у которых эти уровни неприемлемо высоки (например, уровни превышающие среднее значение плюс два стандартных отклонения уровней среди всех остальных интервьюеров/групп). Кроме того, эти доли отказавшихся могут быть вычислены по: (1) неделе сбора данных, во время которой проводились интервью, (2) возрасту респондента, (3) полу респондента и (4) критерию проживания респондента в городской/сельской местности.

Полная документация о том, каким образом высчитывались уровни отказа, должна быть представлена странами. Доли отказавшихся на уровне отдельных лиц должны быть представлены по всей стране в целом, а также по более мелким административным единицам районов в случае, если размеры выборки достаточны для региональной оценки. Дальнейшая разбивка в рамках региона (если применимо) должна быть сделана по тем переменным, которые использовались для

определения корректировки на неполучение ответов. В эту категорию могут входить параметры проживания в городской/сельской местности и возраста, указанного в перечне (15–24, 25–34, 35–44, 45–54, 55+), пола (мужской, женский) и отношения к курению в данный момент (курящий, некурящий).

5.3.6 Доля отсутствующих ответов на конкретные пункты для переменных показателей информационного бюллетеня

После завершения обработки данных обследования по пунктам вопросника, соответствующим всем показателям, которые будут включены в Страновой бюллетень GATS (см. «Определения показателей обследования GATS» для получения дополнительной информации) должен быть рассчитан и задокументирован уровень неполучения ответов по вопроснику GATS (то есть процентная доля респондентов, которые не смогли предоставить полезные данные по конкретному пункту вопросника, среди всех респондентов, которые должны были ответить на конкретный вопрос). Доли отсутствия ответов на конкретные пункты вычисляются как отношение доли респондентов, от которых не было получено действительных ответов на заданный вопрос (“MISSING^x” по пункту x), к общему количеству респондентов для данного уровня (I) за вычетом количества респондентов, которые пропустили пункт x по уважительной причине (V^x):

$$NRRI^x = \frac{MISSING^x}{(I - V^x)} \times 100$$

Общее количество x не ответивших на конкретном уровне (“MISSING^x”) должно быть получено на основании невзвешенной частоты респондентов, данных об ответе которых на пункт x не имеется, после надлежащей очистки данных для обеспечения правильности соблюдения схемы пропуска. Общее количество респондентов конкретного уровня будет получено исходя из общей невзвешенной частоты ответов мужчин или женщин на вопрос A01 (переменная без ожидаемых пустых полей). Общее количество респондентов с действительным пропуском пункта x может быть получено как частота ответов «.S» на пункт x (см. раздел 5.1.3). Многие показатели Странового бюллетеня обследования GATS являются комбинированными, полученными на основе ответов на несколько вопросов. Отсутствие ответа по любому вопросу, используемому для получения комбинированного показателя, приведет к отсутствию ответа по комбинированному показателю. Уровни неполучения ответов по вопросу ниже 5% считаются низкими. Аналогичным образом, как описано для мер по обеспечению качества в разделах 5.3.2 и 5.3.5, НР должны постоянно предпринимать усилия для предупреждения избыточной доли неполучения ответов по конкретным вопросам у отдельных интервьюеров или групп интервьюеров в качестве средства улучшения качества сбора данных во время проведения выездных работ.

5.4 Официальное рассмотрение качества статистических данных

Разработан научный протокол для проведения Комитетом по рассмотрению выборки официального анализа показателей качества статистических данных в каждой стране GATS на этапе после обработки данных обследования и вычисления удельного веса элементов выборки и до проведения анализа данных и получения результатов. Цель данного анализа состоит в подтверждении соответствия удельного веса элементов выборки вычислениям в соответствии со спецификациями «Руководства по определению удельного веса элементов выборки обследования GATS», в рассмотрении различных показателей, указывающих на высокое качество данных

обследования, и подтверждении полноценного документирования основных аспектов отбора проб и проведения анализа. При подготовке к этому анализу будет оказываться техническая помощь.

Далее описаны конкретные показатели качества и документация, которая будет необходима для проведения этого анализа.

5.4.1 Документация по определению удельного веса элементов выборки

Каждый этап процесса определения удельного веса элементов выборки и корректировки данных должен быть полностью описан в документе, подготовленном национальными статистиками, непосредственно отвечающими за вычисление и корректировку статистических весов выборки обследования GATS. Этот документ должен включать в себя описание с подробным изложением процесса взвешивания и корректировки, а также наглядные примеры того, как именно были вычислены все компоненты окончательных весов для 5 – 10 реальных респондентов выборки. В частности, каждый из перечисленных ниже пунктов необходимо подготовить для проверки лицами, которые будут проводить анализ¹.

- Необходимо подвести краткий итог плана выборки обследования по этапам и описать порядок расчета вероятности выбора для каждого этапа и базисных весов.
- Необходимо подробно описать применение стратегий работы в случае отсутствия ответов и стратегий калибровки, уделяя особое внимание следующему: (i) порядок определения классов взвешивания для поправок на отсутствие ответов на уровне домохозяйств и отдельных лиц, (ii) использование конкретных корректировочных ячеек для калибровки весов и (iii) конкретные числовые характеристики населения, по которым выполнялась калибровка удельных весов. Для получения рекомендаций по составлению этого описания обратитесь к «Руководству по определению удельного веса элементов выборки обследования GATS».
- Выберите 5 – 10 реальных респондентов, с тем чтобы продемонстрировать процесс, который использовался для данного вычисления весов. Респонденты должны быть выбраны по различным стратам, ПЕО и гендерным группам. В случае, если использовались компоненты взвешивания, основанные на полуоткрытых интервалах, в число указанных 5 – 10 примеров необходимо включить респондента, относящегося к одному из этих географических кластеров. По завершении данного этапа необходимо привести вычисление окончательно скорректированного веса по каждому примеру. В частности, для каждого из этих респондентов необходимо:
 - определить компоненты каждой условной и безусловной вероятности (с каждого этапа отбора проб), которые использовались для получения БАЗИСНОГО УДЕЛЬНОГО ВЕСА.
 - изложить и детально описать каждый компонент при вычислении КОРРЕКТИРОВКИ НА ОТСУТСТВИЕ ОТВЕТОВ при вычислениях на уровне домохозяйств и отдельных лиц. По завершении данного этапа необходимо

¹ Следует заметить, многие из пунктов данного списка также должны быть включены в документацию по каждому страновому отчету обследования GATS, в связи с чем проекты текста этой документации (содержащие окончательные варианты требуемых числовых результатов) могут быть отправлены лицам, которые проводят официальный анализ качества статистических данных на уровне страны.

провести корректировку базисных весов на отсутствие ответа на уровне домохозяйств и отдельных лиц.

- Необходимо представить и подробно описать каждый компонент метода КАЛИБРОВКИ ВЕСА для получения окончательно взвешенного распределения выборки, для достижения соответствия выбранному населению.
- Предоставьте следующие результаты расчетов, выполненных на основе формул, приведенных в разделах 5.2 и 5.3 данного руководства:
 - постстратификационная калибровка для каждой корректировочной ячейки, используемой для калибровки (5.2.1);
 - мультипликативный эффект переменных весов выборки (*Meff*) по всем подгруппам населения, по которым будут предоставлены оценки в страновом отчете обследования GATS (5.2.2);
 - индивидуальные и сводные показатели (то есть минимальный, максимальный и медианный) общего эффекта плана выборки и показателя однородности в пределах ПЕО для всех оценок обследования, представленных в страновом бюллетене обследования GATS (5.2.3);
 - индивидуальные и сводные показатели (то есть минимальный, максимальный и медианный) предела ошибки (*MOE*) или других сопоставимых показателей точности для всех оценок обследования в страновом бюллетене обследования GATS (5.2.4);
 - индивидуальные и сводные показатели (то есть минимальный, максимальный и медианный) доли ответов домохозяйств по корректировочным ячейкам, используемым для проведения корректировки на неполученные ответы на уровне домохозяйств (5.3.3);
 - доли ответов на уровне отдельных лиц в разбивке по корректировочным ячейкам, использованным для корректировки на неполученные ответы на уровне отдельных лиц (5.3.4);
 - доли отказов на уровне отдельных лиц в разбивке по корректировочным ячейкам, использованным для корректировки на неполученные ответы на уровне отдельных лиц (5.3.5);
 - индивидуальные и сводные показатели (то есть минимальный, максимальный и медианный) доли неполученных ответов по конкретным вопросам обследования, использованных для определения уровней распространенности употребления табака в страновом бюллетене обследования GATS (5.3.6).

В «Руководстве по определению удельного веса элементов выборки обследования GATS» (глава 4) представлено несколько предлагаемых методов для обеспечения более высокого качества веса выборки. Будет проведена проверка следующих параметров и, следовательно, данные этой проверки будут необходимы лицам, проводящим анализ качества статистической работы:

- средний размер базисных весов, отнесенный к среднему размеру веса (*весов*) скорректированного на неполучение ответов, приблизительно равен окончательной общей доле ответов на уровне отдельных лиц.
- Взвешенное распределение выборки (с использованием окончательно скорректированных весов) по всем корректировочным ячейкам должно точно совпадать с соответствующим распределением внешней оценки численности населения, по которой производилась калибровка весов.
- Сумма окончательно откорректированных весов *по всей выборке* должна быть равна общей численности населения (в возрасте 15 лет и старше), согласно источнику, использованному для калибровки. Если таким источником являлась последняя перепись населения, эта сумма весов должна совпадать с численностью целевой группы населения по данным последней переписи.
- Большая часть постстратификационных корректировок (по корректировочным ячейкам) должна быть несколько больше или несколько меньше 1.
- Значение *Meff* составляет $\leq 2,00$ для каждой отчетной подгруппы странового отчета обследования GATS.

5.5 Создание файла аналитических данных

После взвешивания выборки и завершения всех проверок по обеспечению качества необходимо создать новый файл, содержащий только случаи с окончательным диспозиционным кодом «11» на уровне отдельных лиц (*индивидуальный вопросник* завершен). «Респондентами» для *индивидуального вопросника* обследования GATS будут считаться только случаи с окончательным диспозиционным кодом «11» на уровне отдельных лиц.

Следует заметить, что отсутствует необходимость создания окончательного набора аналитических данных до проведения проверок по обеспечению качества с использованием предела ошибки, эффекта плана выборки или основных оценок обследования, вычисленных посредством процедур SAS или SPSS для сложных выборок, так как случаи с окончательными диспозиционными кодами для индивидуального вопросника, отличающимися от «11», не будут обладать значениями для весовых переменных (то есть у них будут отсутствовать значения), и они будут по умолчанию автоматически пропущены программным обеспечением, если не указано иное.

Этот новый файл называют файлом аналитических данных, который должен использоваться при проведении анализа данных для получения оценок обследования (то есть показателей), в том числе оценок для странового бюллетеня и странового отчета.

5.6 Создание файла данных для публичного использования

Набор данных для публичного использования – это файл данных, который CDC направляет населению. В него входят только завершённые случаи интервью, содержащие все переменные с ответами, и стандартные основные показатели обследования GATS. Кроме того, в файле данных содержатся географические переменные и переменные выборки, такие как вес страты, кластера и результатов окончательного обследования.

Набор данных для публичного использования подлежит обнародованию в целях обеспечения доступа внешних исследователей к сведениям GATS. На основании политики опубликования данных обследования GATS, данный материал будет обнародован через один год после опубликования странами национальных версий странового бюллетеня обследования GATS.

6. Требования к данным и документации

В этой главе описаны файлы данных и документация, которые должны быть отправлены из стран и помещены в архив Координационного центра данных обследования GATS.

Файлы данных и документация, запрашиваемые от страны:

1. запрограммированные файлы DB3 в окончательной версии: "CMSDB.db3", "Survey0.db3" и "Survey1.db3";
2. контрольный файл DB3, созданный ПО для агрегирования на основе "CMSDB.db3", "Survey0.db3" и "Survey1.db3";
3. контрольный файл базы данных (с использованием переменной на английском языке), созданный с применением утилиты "Generate Analysis Files" (создание файлов для анализа), содержащий транспонированные неструктурированные файлы для таблицы ответов с действительными ответами на *вопросник для домохозяйств и индивидуальный вопросник*, а также контрольный файл для составления выборки (идентификатор случая ("Case ID") и идентификатор выборки);
4. файлы синтаксиса SAS, SPSS и STATA;
5. компоненты весов выборки и документация (см. раздел 5.4);
6. набор взвешенных данных: окончательный очищенный неструктурированный файл(-ы) с весами и спецификациями для взвешивания.

7. Библиография

- Biemer, Paul P. and Lars E. Lyberg. 2003. Introduction to Survey Quality. John Wiley and Sons, Inc., Hoboken, NJ.
- Encyclopedia of Survey Research Methods*, (2007). Lavrakas, P.J. (Ed.), Sage Publishing, Inc. Los Angeles, CA.
- Gabler, S., Haeder, S., and Lahiri, P. (1999). "A model based justification of Kish's formula for design effects for weighting and clustering," *Survey Methodology*, 25 (1): 105-106.
- Сотрудничающая группа по проведению Глобального обследования потребления табака взрослым населением. Глобальное обследование потребления табака взрослым населением (GATS): технические параметры для программирования основного вопросника. Атланта, Джорджия: Центры США по контролю и профилактике заболеваний, 2019 г.
- Сотрудничающая группа по проведению Глобального обследования потребления табака взрослым населением. Глобальное обследование потребления табака взрослым населением (GATS): основной вопросник с дополнительными вопросами. Атланта, Джорджия: Центры США по контролю и профилактике заболеваний, 2019 г.
- Сотрудничающая группа по проведению Глобального обследования потребления табака взрослым населением. Глобальное обследование потребления табака взрослым населением (GATS): план проведения работы с данными. Атланта, Джорджия: Центры США по контролю и профилактике заболеваний, 2019 г.
- Сотрудничающая группа по проведению Глобального обследования потребления табака взрослым населением. Глобальное обследование потребления табака взрослым населением (GATS): рекомендации по проведению инструктажа в сфере работы с данными. Атланта, Джорджия: Центры США по контролю и профилактике заболеваний, 2019 г.
- Сотрудничающая группа по проведению Глобального обследования потребления табака взрослым населением. Глобальное обследование потребления табака взрослым населением (GATS): руководство для выездного интервьюера. Атланта, Джорджия: Центры США по контролю и профилактике заболеваний, 2019 г.
- Сотрудничающая группа по проведению Глобального обследования потребления табака взрослым населением. Глобальное обследование потребления табака взрослым населением (GATS): руководство для полевого супервайзера. Атланта, Джорджия: Центры США по контролю и профилактике заболеваний, 2019 г.
- Сотрудничающая группа по проведению Глобального обследования потребления табака взрослым населением. Глобальное обследование потребления табака взрослым населением (GATS): руководство по составлению карт и перечней. Атланта, Джорджия: Центры США по контролю и профилактике заболеваний, 2019 г.
- Сотрудничающая группа по проведению Глобального обследования потребления табака взрослым населением. Глобальное обследование потребления табака взрослым населением (GATS): руководство по программированию комплекса "General Survey System". Атланта, Джорджия: Центры США по контролю и профилактике заболеваний, 2019 г.
- Сотрудничающая группа по проведению Глобального обследования потребления табака взрослым населением. Глобальное обследование потребления табака взрослым населением (GATS): повопросные разъяснения. Атланта, Джорджия: Центры США по контролю и профилактике заболеваний, 2019 г.
- Сотрудничающая группа по проведению Глобального обследования потребления табака взрослым населением. Глобальное обследование потребления табака взрослым населением (GATS): руководство по составлению плана выборки. Атланта, Джорджия: Центры США по контролю и профилактике заболеваний, 2019 г.

Сотрудничающая группа по проведению Глобального обследования потребления табака взрослым населением. Глобальное обследование потребления табака взрослым населением (GATS): руководство по определению удельного веса элементов выборки. Атланта, Джорджия: Центры США по контролю и профилактике заболеваний, 2019 г.

Internal Reference for SAS Special Missing Values:

<http://support.sas.com/documentation/cdl/en/lrcon/61722/HTML/default/a000992455.htm>

Kish, L. (1965) *Survey Sampling*, Wiley and Sons, New York.

Kalton, G. and Flores-Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics*, 19(2), 81-97.

Potter, F. J. (1988). Survey of Procedures to Control Extreme Sampling Weights. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 453-458.

Приложение А. Терминологический глоссарий

Файл случая: список выборки и информация, которая загружается на портативные устройства для сбора данных.

Основной вопросник: стандартные основные вопросы, используемые при проведении обследования GATS. Страны-участницы адаптируют основной вопросник к собственным потребностям, следуя определенным рекомендациям.

Файлы DB3: программирование ПО "GSS" для проведения обследования проводится с использованием файлов DB3 (SQLite).

Система "General Survey System" (GSS): программное обеспечение, загружаемое на портативные устройства для проведения обследования GATS.

Технические параметры для программирования ПО GSS: в дополнение к бумажной версии вопросника, которая используется в стране для создания адаптированной версии документа, указания по программированию ПО GSS разрабатываются для каждой адаптированной версии вопросника и используются для программирования вопросника для портативных устройств.

Настройка портативного устройства: процесс копирования сборки программного обеспечения сборки на портативное устройство, как правило, выполняется одновременно на всех устройствах перед началом инструктажа выездных интервьюеров.

Контрольный файл для составления выборки: набор данных, содержащий код идентификатор случая ("CaseID") для каждого домохозяйства в выборке, а также информацию, которая необходима для вычислений удельного веса элементов выборки и анализа комплексных данных обследования.

Обеспечение качества: процесс, состоящий из систематических мероприятий, направленных на обеспечение, оценку и подтверждение качества данных, собранных в ходе обследования.

Комитет по рассмотрению вопросника: группа, состоящая из международных специалистов в области борьбы против табака и составления вопросников. Комитет анализирует и утверждает все адаптированные для страны вопросники с целью обеспечения качества, стандартизации и сопоставимости.

Комитет по рассмотрению выборки: группа, состоящая из международных специалистов в области работы с выборками. Комитет анализирует и утверждает составы выборок, взвешивание выборок и меры для обеспечения качества данных для стран участниц.

Приложение В. Этап до сбора данных

В.1 Адаптация вопросника GATS и процесс анализа

Адаптация вопросника: страны GATS будут работать совместно с CDC над разработкой адаптированного для страны вопросника GATS, что включает следующее.

- **Адаптация основного вопросника обследования GATS.** Страны будут адаптировать основной вопросник для конкретной ситуации на местах. В том числе имеется в виду внесение изменений в конкретные пункты вопросников (например, типы табака) и варианты ответа (например, перечень марок сигарет), которые подходят для конкретной страны.
- **Включение дополнительных вопросов.** В обследовании GATS предусмотрен перечень дополнительных вопросов, которые страны могут выбрать для включения в вопросник.
- **Включение дополнительных вопросов, разработанных страной.** Страны могут включать в обследование собственные вопросы, которые отсутствуют в основном вопроснике и перечне дополнительных вопросов.
- **Удаление нецелесообразных вопросов.** Страны могут удалить из перечня основных вопросов те, которые не относятся к ситуации на местах (например, употребление бездымного табака не распространено). Должно быть представлено должное обоснование.

Процесс рассмотрения вопросника GATS состоит из следующих 10 этапов.

1. **Адаптация проекта вопросника для страны.** Страны будут взаимодействовать с CDC по линии создания проекта вопросника конкретной страны для предоставления его в Комитет по рассмотрению вопросника.
2. **Форматирование.** CDC рассмотрит представленный проект вопросника и отформатирует его по необходимости (например, отредактирует указания по пропуску вопросов и выделит их для анализа Комитетом во рассмотрению вопросника).
3. **Экспертный анализ Комитетом по рассмотрению вопросника.** Члены Комитета по рассмотрению вопросника анализируют вопросник и предоставляют комментарии.
4. **Компиляция комментариев рецензентов.** CDC обобщает комментарии Комитета по рассмотрению вопросника и возвращает материалы в страну для изучения.
5. **Коммуникация (принятие решения).** Комитет по рассмотрению вопросника и страна совместно работают над разрешением всех нерешенных вопросов (по электронной почте и/или конференц-связи).
6. **Окончательная доработка адаптированного вопросника.** После решения всех вопросов адаптированный для страны вопросник считается доработанным, и Комитет по рассмотрению вопросника официально его утверждает.
7. **Уведомление о принятии страной.** Страна подтверждает окончательное принятие вопросника, утвержденного Комитетом по рассмотрению вопросника.
8. **Перевод.** Страны должны выполнить перевод утвержденного вопросника GATS на соответствующий язык(-и) и предоставить обратный перевод вопросника на английский язык. (Примечание: это может проводиться одновременно с этапом 1, на котором страны

сначала переводят основной вопросник на национальный язык и адаптируют переведенную версию, а впоследствии выполняют обратный перевод на английский язык для предоставления в Комитет по рассмотрению вопросника.)

9. **Проверка итоговой переведенной версии и ее обратного перевода.** Комитет по рассмотрению вопросника выполняет анализ переведенной версии и ее обратного перевода для проверки адекватности перевода (то есть того, что смысл вопросов остался неизменным). (Примечание: в случае, если перевод выполняют одновременно с этапом 1, и версию в обратном переводе первоначально представляют для анализа Комитетом по рассмотрению вопросника, данная ситуация рассматривается как часть процесса анализа Комитетом по рассмотрению вопросника.)
10. **Программирование.** Подготавливают параметры программирования для адаптированного на национальном уровне вопросника и программируют адаптированное для страны обследование на портативном устройстве.

В.2 Процесс программирования вопросника GATS

Далее приведены основные шаги и рекомендации по срокам программирования вопросника GATS.

1. Страновой вопросник обследования GATS и предложение по предварительной проверке должны быть утверждены за 6 – 8 недель до проведения инструктажа по предварительной проверке.
2. Программирование вопросника на английском языке начинают за 6 недель до проведения инструктажа по предварительной проверке.
3. Доставку и наладку портативных устройств завершают за 4 недели до начала инструктажа.
4. Перевод программных меню ПО "GSS" и оповещений завершают за 1 неделю до проведения инструктажа по предварительной проверке.
5. Перевод текста вопросников (*вопросник для домохозяйств и индивидуальный вопросник*) на все языки страны, проводящей обследование, должен быть завершен и загружен в базы данных "survey0" и "survey1" за 2 недели до начала инструктажа по предварительной проверке.
6. Страна, проводящая обследование, должна официально утвердить *вопросник для домохозяйств и индивидуальный вопросник* за 1 неделю до проведения инструктажа по предварительной проверке.
7. Необходимо начать контроль версий вопросников с официально утвержденной версии и с этого момента поддерживать строгий контроль.

В.3 Процесс настройки ПО GSS на портативных устройствах

В этом разделе описаны меры для обеспечения качества, которые должны быть реализованы на различных этапах процесса, с тем чтобы убедиться, что на каждом портативном устройстве, задействованном в выездных работах, установлены верные и одинаковые программные и аппаратные параметры.

- 1. Создание контрольной версии прошивки.** Контрольная версия прошивки должна быть создана на одной машине, на которой нет вирусов и установлено антивирусное программное обеспечение. После создания одной контрольной версии прошивки ее необходимо загрузить на два различных портативных устройства и затем провести полный процесс итеративного тестирования на обоих устройствах, отвечая на каждый вопрос вопросника. При выявлении любых ошибок на любом портативном устройстве необходимо внести исправления в программируемые параметры – как описано в разделе 3.3.2, – и данный процесс необходимо проводить повторно до достижения 100% правильности результатов. Данный процесс необходимо проводить в присутствии координатора/контактного лица и совместно с ведущим ИТ-специалистом.
- 2. Наклеить стикеры, свидетельствующие о прохождении контроля качества (шаблон).** На всех портативных устройствах должен быть наклеен стикер с идентификатором ВИ ("FIID"), серийным номером и пустым контрольным перечнем для отметок о проведении контроля качества, который размещают на задней крышке портативного устройства. Ниже приведен пример.

701010		SN A10	
	<u>КК1</u>		<u>КК2</u>
<input type="checkbox"/> Исх. конфиг	<input type="checkbox"/> Дата	<input type="checkbox"/> Дата	
<input type="checkbox"/> APK закружен	<input type="checkbox"/> FIID	<input type="checkbox"/> FIID	
<input type="checkbox"/> Передача	HQ ____	<input type="checkbox"/> Удал. учебн. случаи	
	IQ ____	<input type="checkbox"/> Батарея ____	
	<input type="checkbox"/> Батарея ____	Передачи	
Нач ____	Передача	Случаи ____	
	КК1 ____	КК2 ____	

- 3. Подготовка портативных устройств.** После копирования окончательной версии прошивки все портативные устройства должны быть настроены согласно соответствующему пошаговому процессу (см. процесс подготовки портативных устройств, который описан в «Руководстве по программированию комплекса "General Survey System"»). Необходимо вновь провести распределить портативные устройства по различным коробкам (перед настройкой и после настройки).
- 4. Проведение первого этапа контроля качества (КК1).** Портативные устройства необходимо поровну распределить между сотрудниками службы контроля качества из команды, которая проводила настройку портативных устройств или контролировала этот процесс. На каждом портативном устройстве необходимо выполнить каждый этап КК1. КК1 должен быть строгим процессом проверки качества, и на каждом портативном устройстве необходимо проверить правильность перечисленных ниже пунктов с последующим указанием контрольного перечня на стикере:
 - 1) дата и время;
 - 2) идентификатор ВИ ("FIID") (проверить наклейку с идентификатором вопросника для домохозяйств);

- 3) версия вопросника для домохозяйств;
- 4) версия индивидуального вопросника;
- 5) питание: время работы батареи.

Может быть необходима дополнительная проверка других параметров в зависимости от условий в стране, например количество случаев, портативные устройства для интервьюеров мужского/женского пола и т.д. При выявлении любых проблем необходимо переместить портативные устройства в другую коробку и вновь провести процесс аппаратного сброса, и вся процедура на данных портативных устройствах должна быть проведена повторно начиная с этапа 1.

5. **Проведение второго этапа контроля качества (КК2).** Процесс КК2 проводится после того как все портативные устройства прошли КК1, о чем сделана запись на стикерах. Портативные устройства должны быть перераспределены таким образом, чтобы первый и второй этапы контроля качества проводились разными людьми. В ходе КК2 должен быть реализован тот же процесс, что и описанный выше для КК1; дополнительным шагом в КК2 является проверка всех стикеров на предмет внесения отметок и выполненной проверки. Проведение КК2 является важным, так как известно, что в ходе его проведения удается выявить небольшое количество ошибок. Все портативные устройства с проблемами, выявленными во время проведения КК2, должны пройти весь процесс заново, начиная с аппаратного сброса на первом этапе.
6. **Подписание акта руководителем службы ИТ и Страновым координатором.** После того как каждое портативное устройство пройдет КК1 и КК2 необходимо, чтобы руководитель службы ИТ и Страновой координатор обследования GATS провели документирование и подписание краткого обобщающего документа, в котором будет указано количество портативных устройств, версия вопросника для домохозяйств и индивидуального вопросника, дата, время, а также сотрудники, проводившие подготовку портативных устройств и контроль качества. Официальное подписание должно стать подтверждением того, что портативные устройства подготовлены для использования в ходе инструктажа по полному циклу проведению обследования / выездным работам, в дальнейшем не будет внесено изменений в ПО, и устройства не будут использоваться в иных целях до окончания проведения обследования GATS в стране.

Приложение С. Этап после сбора данных

С.1 Окончательные диспозиционные коды и вычисление доли ответивших лиц

После тщательного анализа доступной литературы по системам эпидемиологического надзора, национальных обследований домохозяйств, связанных и не связанных с потреблением табака, в рамках которых для сбора данных в домохозяйствах использовался либо телефонный опрос, либо личный опрос, для обследования GATS был разработан перечень окончательных диспозиционных кодов расположения и вычисления доли ответивших лиц. Для этой цели были рассмотрены несколько обследований, включая обследования Бразильского института географии и статистики (IBGE), Национальное обследование демографии и репродуктивного и сексуального здоровья 2004 г. (ENDSSR-2004), Исследования репродуктивного здоровья в Грузии (RHS), Демографические и медико-санитарные обследования (DHS), Национальное обследование состояния здоровья семьи (NFHS), Система эпидемиологического надзора за поведенческими факторами риска (BRFSS), Система эпидемиологического надзора за поведенческими факторами риска молодежи (YRBSS) и исследования Американской ассоциации по изучению общественного мнения (AAPOR).

Отчетность по окончательным диспозиционным кодам и долям ответивших лиц описана в следующих таблицах этого раздела. В таблицах С-1 и С-2 представлены окончательные диспозиционные коды для вопросников для домохозяйств и индивидуальных вопросников, которые должны быть присвоены различным кодам окончательного результата из вопросников для домохозяйств и индивидуальных вопросников. (Подробные сведения информация о кодах результата изложена в Руководстве выездного интервьюера обследования GATS.) В таблице С-3 представлен шаблон вычислений и учета доли ответивших лиц и диспозиционных кодов в страновом отчете обследования GATS.

Таблица С-1: окончательные диспозиционные коды и коды результата вопросника для домохозяйств

Окончательные диспозиционные коды	Коды окончательного результата	Название и описание
1	200	Вопросник для домохозяйств завершен, выбран один человек - вопросник для домохозяйств считается завершенным в случае, если перечень заполнен
2	201	Вопросник для домохозяйств завершен, не выбран ни один человек -нет лиц в возрасте от 15 лет и старше -нет лиц, которые считают данное домохозяйство своим обычным местом проживания -Указано, что от данного домохозяйства должны участвовать мужчины, однако мужчины в нем не проживают, либо указано, что от данного домохозяйства должны участвовать женщины, однако женщины в нем не проживают (если была использована рандомизация по полу)
3	202	Вопросник для домохозяйств завершен частично, завершить перечень невозможно -незавершенное интервью

Окончательные диспозиционные коды	Коды окончательного результата	Название и описание
4	203	Вопросник для домохозяйств не завершен, невозможно установить подходящего по критериям респондента для скрининга -дома отсутствуют лица в возрасте от 18 лет или старше -имеющееся лицо в возрасте от 18 лет или старше некомпетентно -в данном домохозяйстве могут проживать либо не проживать лица, имеющие право участвовать в обследовании
5	209	Никого нет дома
6	204	Отказ домохозяйства от участия
7	205	Нежилой дом
8	206	Выбранный адрес не является домохозяйством
9	208	Отсутствие ответа домохозяйства по другой причине

Таблица С-2: окончательные диспозиционные коды и коды результата индивидуального вопросника

Окончательные диспозиционные коды	Коды окончательного результата	Название и описание
11	400, 402*	Индивидуальный вопросник заполнен -вопросник заполнен до вопроса E01 -ни на один из основных вопросов (например, B01/B02/B03 и C01/C02/C03, если применимо) был получен ответ «неизвестно» или «отказ от ответа»
12	402*	Не завершен -интервью по вопроснику прекращено до вопроса E01
13	403	Впоследствии было установлено, что выбранный респондент не имеет права участвовать в обследовании -возраст < 15 лет -респондент не считает данное домохозяйство своим постоянным местом жительства -неверно выбран пол (если использована рандомизация по критерию пола)
14	409	Выбранного респондента нет дома
15	404	Отказ выбранного респондента
16	407	Выбранный респондент недееспособен
17	408	Отсутствие ответов в индивидуальном вопроснике по другим причинам

* Случаям с кодом окончательного результата «402» (индивидуальный вопросник заполнен частично) будет присвоен либо код «11» (индивидуальный вопросник завершен), либо «12» (не завершен). См. правила, описанные в *разделе 5.1.7*.

Таблица С-3: таблица странового отчета GATS 3.1—шаблона для учета доли ответивших лиц

	Проживание				Всего	
	Городское население		Сельское население		Количество	Процент
	Количество	Процент	Количество	Процент		
Выбранное домохозяйство						
Завершено (HC)						
Завершено – нет лиц, отвечающих критериям участия (HCNE)						
Не завершено (HINC)						
Нет респондентов для скрининга (HNS)						
Никого нет дома (HNN)						
Отказ (HR)						
Не заселен (HUO)						
Адрес не жилой (HAND)						
Другое ¹ (HO)						
Всего выбрано домохозяйств		100		100		100
<i>Доля ответивших домохозяйств (HRR) (%)²</i>						
Выбранный человек						
Завершено (PC)						
Не завершено (PINC)						
Неправомочен (PNE)						
Нет дома (PNH)						
Отказ (PR)						
Недееспособный (PI)						
Другое ¹ (PO)						
Общее количество лиц, вошедших в выборку		100		100		100
<i>Доля ответивших лиц на уровне отдельных респондентов (PRR) (%)³</i>						
<i>Совокупная доля ответивших лиц (TRR) (%)⁴</i>						
¹ «Другое» охватывает любые другие не указанные результаты.						
² Доля ответов домохозяйств (HRR) вычисляется как:						
$\frac{HC * 100}{HC + HINC + HNS + HNN + HR + HO}$						
³ Доля ответов отдельных респондентов (PRR) вычисляется как:						
$\frac{PC * 100}{PC + PINC + PNH + PR + PI + PO}$						
⁴ Совокупная доля ответивших лиц (TRR) вычисляется как:						
$\frac{(HRR * PRR)}{100}$						
Примечания.						
— Незавершенное интервью по вопроснику для домохозяйств (то есть нет возможности заполнить перечень) рассматривается как отсутствие респондента для обследования GATS. Таким образом, эти случаи (HINC) не были включены в числитель при определении доли ответивших домохозяйств.						
— Общее количество выбранных респондентов должно быть равным количеству завершенных [HC] интервью для домохозяйств.						
— Завершенные интервью респондентов [PC] включают в себя респондентов, ответивших по крайней мере на вопрос E01 и предоставивших действительные ответы на вопросы B01/B02/B03 (и C01/C02/C03, если применимо). Респонденты, не отвечающие этим критериям, расцениваются как не завершившие интервью (PINC) и отказавшиеся от ответа респонденты GATS, которые, таким образом, не были включены в числитель при вычислении уровня ответов отдельных респондентов.						

С.2 Модель калибровки постстратификационных весов с помощью корректировочных ячеек

Общие сведения

В ходе заключительного этапа подсчета удельного веса элементов выборки проводится калибровка весов относительно численности населения по известным коррелятам основных показателей результата исследования, называемых калибровочными переменными (например, пол, образование, возраст, проживание в городской/сельской местности и регион, согласно «Руководству по определению удельного веса элементов выборки обследования GATS»). С точки зрения процедуры калибровка в виде постстратификации включает формирование «корректировочных ячеек» путем перекрестной классификации величин коррелятов. Значение постстратификационной корректировки в каждой из этих корректировочных ячеек составит менее 1,00 в случае, если подвыборка в этой категории была чрезмерно представлена в данной выборке после поправки на произведение выборки и неполучение ответов, и составит более 1,00 в тех категориях, в которых подвыборка была представлена недостаточно. Оценка шаблонов постстратификационных корректировок по корректировочным ячейкам помогает выявить дисбаланс, сохраняющийся после устранения дисбаланса выборки в связи с неполучением ответов.

Источник(и) данных

Для этих расчетов следует использовать файл данных, содержащий окончательный набор весов выборки с поправкой на неполучение ответов, а также всех компонентов для вычисления окончательного веса каждого респондента.

Программное обеспечение для вычислений

Какого-либо специализированного оборудования не требуется, так как необходимые вычисления могут быть легко запрограммированы в любой программной среде.

Вычисление

Этапы процедуры вычисления постстратификационных корректировок описаны в разделе 3.3 Руководства по определению удельного веса элементов выборки обследования GATS. Используя окончательно вычисленные результаты этих расчетов, необходимо представить набор постстратификационных корректировок в виде таблицы сопряженности по коррелированным переменным, которые служили для определения корректировочных ячеек (то есть калибровочные переменные). Предпочтительна ситуация, при которой все значения постстратификационной корректировки близки к 1,00, некоторые – несколько больше 1,00 и остальные – несколько меньше 1,00.

Интерпретация

Существенные отклонения от 1,00 в этих постстратификационных корректировках указывают на серьезные проблемы, связанные с дисбалансом выборки за счет дифференциального охвата выборки домохозяйств, не выявленные проблемы вычисления весов выборки либо масштабные демографические изменения в целевой группе населения между датой переписи населения, по которой была проведена калибровка выборки, и датой проведения выборки и сбора данных по обследованию GATS.

Таблица С-4: гипотетический пример таблицы корректировок постстратификационной калибровки весов в разбивке по региону, проживанию в городской/сельской местности, полу и возрасту.

Регион	Город/ Сельская местность	Пол	Возраст	Корректировка
1	Городское население	Мужчины	15 – 24	
			25 – 44	
		Женщины	45 – 59	
			60+	
	Сельское население	Мужчины	15 – 24	
			25 – 44	
Женщины		45 – 59		
		60+		
2	Городское население	Мужчины	15 – 24	
			25 – 44	
		Женщины	45 – 59	
			60+	
	Сельское население	Мужчины	15 – 24	
			25 – 44	
Женщины		45 – 59		
		60+		
3	Городское население	Мужчины	15 – 24	
			25 – 44	
		Женщины	45 – 59	
			60+	
	Сельское население	Мужчины	15 – 24	
			25 – 44	
Женщины		45 – 59		
		60+		

С.3 Эффект вариации весов выборки в точности оценок обследования

Общие сведения

Вариация весов выборки зачастую приводит к увеличению дисперсии (то есть к снижению точности) оценок обследования. Такой вид вариации станет важной проблемой при определении плана выборки в странах, в которых отмечается переизбыток выборки по критерию проживания в городской/сельской местности или конкретном регионе. Kish (1965, раздел 11.7) предложил простую, но давно применяемую модель оценки мультипликативного увеличения дисперсии в случае, если размер весов не коррелирует с показателем обследования, используемым для оценки простой характеристики населения (θ). Для данного мультипликативного эффекта будет использоваться обозначение $Meff_w$.

$Meff_w$ не специфичен для какой-либо оценки, однако он является специфичным для категории учета, которая используется в ходе конкретной оценки. Таким образом, он будет применяться ко всем оценкам в рамках обследования GATS для данного региона. Например, значение $Meff_w$, вычисленное для оценок по общей численности населения, должно быть подсчитано на основании всей выборки, тогда как значение $Meff_w$, вычисленное для ориентировочных оценок по сельскому населению, должно быть подсчитано только на основе ответов респондентов, проживающих в сельской местности.

Источник(и) данных

Для проведения этих вычислений должен применяться использованный для анализа файл с окончательно взвешенными данными.

Программное обеспечение для вычислений

Специального программного обеспечения не требуется. Вычисление $Meff_w$ может быть легко запрограммировано с использованием выходных данных применяемого статистического программного обеспечения.

Вычисление

Принимая во внимание итоговую величину удельного веса (w_i), вычисленную для каждого из n респондентов, к которым применяется $Meff_w$, необходимо вычислить простое среднее значение удельных весов для получения \bar{w} и затем использовать формулу $s_w^2 = \sum_{i=1}^n (w_i - \bar{w})^2 / (n - 1)$ для расчета дисперсии удельных весов элементов выборки. Расчет $Meff_w$ из окончательного набора весов выборки для респондентов GATS может быть проведен следующим образом:

$$Meff_w = 1 + \frac{s_w^2}{\bar{w}^2}, \quad (1)$$

Следует заметить, что $Meff_w \geq 1$, так как \bar{w}^2 и s_w^2 всегда имеют положительные значения.

Интерпретация

Очевидно, что предпочтительно значение $Meff_w$, максимально близкое к единице, вместе с тем $Meff_w > 2,00$ может рассматриваться как значительное и требовать принятия каких-либо мер. Стратегии корректировки веса или цензурирования являются наиболее распространенным

средством в случае чрезмерно изменяющихся показателей удельного веса (Potter, 1988). Окончательное решение о том, необходимо ли вносить корректировки в значения удельных весов зависит от баланса между преимуществом в точности результата такой корректировки и влиянием изменений весов на результат оценки обследования. В случае, если корректировка значения удельного веса приводит к уменьшению $Meff_w$, однако несущественно изменяет весовые оценки основных показателей результатов исследования, корректировка может быть оправдана с учетом необходимого времени и усилий для ее проведения.

Пример

Для выборки обследования GATS среднее значение окончательных весов выборки среди всех X,XXX респондентов обследования составляет $\bar{w} = YYY,YYYY$, а простая дисперсия составляет $s_w^2 = ZZZZ,ZZ$. Мультипликативный эффект отклонения весов в общенациональной оценке на уровне страны составляет $Meff_w = 1 + \frac{YYY,YYYY}{XXX,ZZ} = U,UU$.

Выборку с $Meff_w = U,UU < 2,00$ можно рассматривать как выборку, в которой вариация удельных весов достаточно невелика и не требует никаких мер для снижения влияния отклонения весов на оценки, например корректировки значений удельных весов (см. Potter, 1988).

С.4 Общий эффект плана выборки в точности оценок обследования и внутрикластерная однородность в пределах ПЕО соответствующих основных оценочных показателей обследования

Общие сведения

Kish (1965) первоначально определил «эффект плана выборки», который обозначается как "deff", для измерения мультипликативных изменений дисперсии оценки обследования в связи с использованием кластерной выборки для данного состава выборки по сравнению с дисперсией сопоставимой оценки простой случайной выборки такого же размера. Каждый основной показатель обследования GATS является пропорцией (p) или отношением, которое выражается как процентная доля или средняя величина (среднее). Изначальная модель Kish для эффекта плана выборки (за счет кластерной выборки) при оценке (\hat{p}) из p является

$$Deff_{CS} = 1 + \rho_{\theta}(\bar{m} - 1),$$

где ρ_{θ} – это показатель (характеризует количество респондентов внутри ПЕО) относительной внутрикластерной однородности для количественного определения p в рамках обследования (количественное определение степени, в которой члены какого-либо кластера выборки более схожи между собой по сравнению с популяцией в целом), $\bar{m} = n/a$ – среднее количество респондентов на ПЕО в выборке, n – общий объем выборки респондентов и a – количество ПЕО в выборке.

В случае, если выборка обследования охватывает как кластеры выборки, так и неравные удельные веса выборки, то, согласно Gabler с соавт.(1999), общий эффект плана выборки в оценке, обозначаемый как $Deff(\hat{\theta})$, зависит как от кластерной выборки (см. выше), так и

мультипликативного эффекта удельных весов выборочных переменных (см. обсуждение $Meff_w$); то есть

$$\begin{aligned}
 Deff(\hat{\theta}) &\equiv \frac{V(\hat{p}; design, n)}{V(\hat{p}_{srs}; srs, n)} \\
 &= [Meff_w][Deff_{CS}(\hat{\theta})] = \left[1 + \frac{S_w^2}{\bar{w}^2}\right][1 + \rho_\theta(\bar{m} - 1)]
 \end{aligned}
 \tag{1}$$

Источник(и) данных

Для проведения этих вычислений должен применяться использованный для анализа файл с окончательно взвешенными данными.

Программное обеспечение для вычислений

Те же пакеты аналитического программного обеспечения для обследования, что и указанные в задаче по пределу погрешности (см. приложение С.5).

Вычисление

Аналитическое ПО для обследования вычисляет расчетную величину дисперсии или стандартной ошибки (то есть $v(\hat{p}; design, n)$ или $\sqrt{v(\hat{p}; design, n)}$ в уравнении 1 соответственно). С тем чтобы оценить ρ_θ , уравнение 1 может быть решено для ρ_θ , в результате чего получают:

$$\rho_\theta = \frac{\frac{Deff(\hat{\theta})}{Meff_w} - 1}{\bar{m} - 1}
 \tag{2}$$

Для ориентировочной оценки ρ_θ по каждому показателю обследования из выборки может быть оценена величина $Meff_w$, \bar{m} может быть определена на основании количества ПЕО в выборке (a) и общего размера выборки респондентов (n), при этом во многих типах аналитического ПО для обследований (например, SUDAAN), наряду с оценкой ($\hat{\theta}$) может быть запрошена оценка $Deff(\hat{\theta})$. Если оценка $Deff(\hat{\theta})$ недоступна средствами ПО, используемыми при определении $Deff(\hat{\theta})$ согласно уравнению 1 или ρ_θ согласно уравнению 2, для получения ориентировочных долей может быть проведена общая оценка эффекта выборки:

$$deff(\hat{\theta}) = \frac{v(\hat{p}; design, n)}{v(\hat{p}_{srs}; srs, n)} = \frac{v(\hat{p}; design, n)}{\hat{p}(1 - \hat{p})/n},$$

так как $v(\hat{p}_{srs}; srs, n) = \hat{p}(1 - \hat{p})/n$. Затем ориентировочное значение ρ_θ может быть получено непосредственно из уравнения 2.

Интерпретация

Желательно, чтобы общий эффект плана выборки в оценке или $Deff(\hat{\theta})$ составлял менее 2,50. Как правило, ориентировочные оценки ρ_θ (для однородности внутри ПЕО) являются небольшими

положительными числами в диапазоне от 0 до 1 (как правило, от 0 до 0,05). В ряде случаев оценочные величины ρ_θ будут небольшими отрицательными числами.

Для обобщения уровней $deff(\hat{\theta})$ среди всех оценок обследования GATS рекомендуется определить $Deff(\hat{\theta})$ для ряда главных показателей основного результата (например, текущая распространенность курения, распространенность попыток курить, темпы отказа от табака и т.п.) по различным категориям учета (например, в разбивке по полу, возрасту, проживанию в городской/сельской местности, региону и различным перекрестным классификациям на основе комбинации данных переменных). Затем для каждого показателя важного конечного результата необходимо обобщить ориентировочные оценки $Deff(\hat{\theta})$ по всем категориям посредством учета следующих обобщенных статистических показателей ориентировочных значений $Deff(\hat{\theta})$: медианное значение, минимальное значение и максимальное значение.

Резюмировать значения $\hat{\rho}_\theta$ (медианное, минимальное и максимальное) среди всех оценок обследования GATS необходимо тем же образом, что и для $deff(\hat{\theta})$.

Некоторые системы обследований или организации предпочитают отражать эффект плана выборки в отчетах с помощью обозначения "deft" вместо "deff". "Deft" подсчитывают для каждой оцениваемой величины как соотношение между стандартной ошибкой данного состава выборки и стандартной ошибкой, которая имела бы место при применении простой случайной выборки.

С.5 Предел ошибки основных оценок обследования

Общие сведения

Необходимо представить определения некоторых основных показателей статистической точности выборок обследования. Во-первых, необходимо заметить, что все ориентировочные показатели по выборке зачастую несколько отличаются от таких показателей по популяции. Данная разница обычно называется **ошибкой выборки**. Вероятностные выборки, подобные тем, которые использовались в обследовании GATS, дают возможность подсчитать сводные показатели ошибки выборки, которые указывают на точность оценок. Существует несколько различных сводных показателей точности оценок обследования. Для определения этих показателей можно применять обозначение θ , которое отражает оцениваемую популяционную характеристику (например, показатель распространенности лиц, которые в настоящее время курят сигареты). Авторы применяют $\hat{\theta}$ для обозначения ориентировочной оценки θ на основе выборки.

Первый распространенный обобщающий показатель точности – это **дисперсия оценки**, обозначаемая как $V(\hat{\theta})$. Как будет видно впоследствии, все остальные показатели в той или иной степени соотносятся с $V(\hat{\theta})$. Дисперсия оценок обследования и все другие связанные с ней показатели являются количественными показателями того, в какой степени оценки всех возможных выборок отличаются от тех оценок, которые будут получены при составе выборки, использованном в конкретном обследовании. Вторым показателем точности является

стандартная ошибка оценки, определяемая как $SE(\hat{\theta}) = \sqrt{V(\hat{\theta})}$. Таким образом, стандартная ошибка оценки представляет собой квадратный корень из дисперсии этой оценки и имеет ту же размерность, что и оценка $\hat{\theta}$. Третий показатель статистической точности – это **относительная стандартная ошибка** оценки, определяемая как $RSE(\hat{\theta}) = SE(\hat{\theta})/\theta = \sqrt{V(\hat{\theta})}/\theta$. Так как $RSE(\hat{\theta})$

определяет точность по отношению к размеру оцениваемого параметра, данная величина безразмерна и, таким образом, обеспечивает большую сопоставимость при определении точности оценок, чем $SE(\hat{\theta})$.

Рекомендуемым показателем точности для страновой отчетности основных оценочных величин обследования GATS (например, показатель распространенности употребления табака и т. п.) является **предел ошибки**, определяемый как $MOE(\hat{\theta}) = [Z][SE(\hat{\theta})]$, где Z является мерой уровня доверия для показателя, и $SE(\hat{\theta})$ является стандартной ошибкой $\hat{\theta}$. Для большинства национальных оценок GATS и получения 95% уровня достоверности может быть использован $Z = 1,96$. Ориентировочная величина ($moe(\hat{\theta})$) для $\hat{\theta}$ может быть интерпретирована следующим образом:

«Имеется 95% степень уверенности в том, что учитываемое значение ($\hat{\theta}$) находится в пределах величины $moe(\hat{\theta})$ относительно θ ».

Аналитикам GATS настоятельно рекомендуется учитывать значение $moe(\hat{\theta})$ по всем основным оцениваемым параметрам, включая общенациональные оценки основных показателей потребления табака. Кроме того, значение $moe(\hat{\theta})$ соответствующих региональных оценок также необходимо учитывать в случае, если состав выборки позволяет увеличить размер региональных выборок таким образом, что могут быть получены региональные оценки, которые будут соответствовать стандартам точности обследования GATS.

Так как показатели точности могут быть легко получены с помощью статистического программного обеспечения, странам GATS настоятельно рекомендуется постоянно запрашивать оценки точности для всех генерируемых расчетных параметров, даже если их необходимо указать только для основных национальных (и, возможно, региональных) параметров. Необходимо с осторожностью подходить к учету ориентировочных параметров, для которых значение $MOE(\hat{\theta})$ приближается или превосходит размер самого параметра, либо не учитывать их вовсе. В качестве еще одного предостережения важно отметить, что, так же как и с $\hat{\theta}$, оценки $V(\hat{\theta})$, $SE(\hat{\theta})$, $MOE(\hat{\theta})$ или $RSE(\hat{\theta})$ могут содержать ошибку выборки. Это имеет важное значение, так как все эти оценки могут быть неточны в случае, если они основаны на небольших выборках, в частности тех, в которых менее 10–20 выборочных ПЕО.

Источник(и) данных

Для проведения этих вычислений должен применяться использованный для анализа файл с окончательно взвешенными данными.

Программное обеспечение для вычислений

При оценке популяционных характеристик ($\hat{\theta}$) и соответствующих значений $moe(\hat{\theta})$ необходимо принимать во внимание статистически значимые аспекты плана выборки. Расчетные характеристики должны быть взвешенными и соответствующие показатели точности должны позволять применение стратификации, кластерной выборки, выборки без замен и весов выборки. Известно, что невозможность это обеспечить, как правило, приводит к отклонениям в оценках (особенно их точности) и тем самым к получению оценок с несоответствующим интервалом и тестам значимости. Таким образом, аналитикам обследования GATS в странах настоятельно рекомендуется использовать аналитическое ПО, которое позволяет полностью учесть состав

выборки, используемый для получения данных обследования. Это означает использование весов выборки для получения всех оценок ($\hat{\theta}$ и $moe(\hat{\theta})$) для дескриптивного анализа. Кроме того, это означает необходимость применения ПО в котором реализован широко распространенный подход к оценке дисперсий и стандартных отклонений величин, изучаемых в обследовании. Тремя основными подходами к оценке дисперсии являются: метод рядов Тейлора (TSL), метод сбалансированного повторяющегося эксперимента (BRR) и метод «складного ножа» (JRR) (Wolter, 1985). В большинстве пакетов статистического ПО для определения дисперсии применяется метод рядов Тейлора (TSL), поэтому для сравнимости результатов, полученных в странах GATS, рекомендуется использование ПО, функционирующего на основе метода Тейлора.

Ряд статистических организаций во всем мире разработали программное обеспечение для анализа данных при сложных выборках, таких как выборка обследования GATS¹. Эти программные средства не только позволяют оценить потребление табака (то есть $\hat{\theta}$), но и охарактеризовать точность (обычно $V(\hat{\theta})$ или $SE(\hat{\theta})$) с надлежащим учетом основных особенностей выборки обследования GATS, то есть кластерной выборки, использования стратификации и различных вероятностей выборки (то есть весов выборки):

Дополнительная информация

Подробные сведения о представленных выше определениях можно найти в публикации “Encyclopedia of Survey Research Methods” (Энциклопедия методов обследования, 2007 г.). Подробная информация в онлайн-формате обо всем аналитическом ПО представлена на веб-сайте, действующем при поддержке Отдела исследовательских методов Американской статистической ассоциации (<http://www.hcp.med.harvard.edu/statistics/survey-soft/>). Помимо данных ссылок на аналитическое ПО, на веб-сайте ассоциации также можно найти ряд других полезных ссылок и ресурсов по теме дизайна и проведения обследований (см. <http://www.amstat.org/sections/SRMS/index.html>).

Вычисление

В результате работы перечисленного выше программного обеспечения будет получено значение $\hat{\theta}$, а также ориентировочная величина дисперсии, обозначаемая как $v(\hat{\theta})$, или стандартная ошибка, обозначаемая как $se(\hat{\theta})$. На основе полученных значений может быть вычислена ожидаемая ошибка для $\hat{\theta}$:

$$moe(\hat{\theta}) = [t][se(\hat{\theta})] = [t]\sqrt{v(\hat{\theta})} \quad (1)$$

Интерпретация

Значение $moe(\hat{\theta})$, полученное для $\hat{\theta}$, интерпретируют следующим образом:

«имеется 95% степень уверенности в том, что ориентировочное значение($\hat{\theta}$) находится в пределах величины $moe(\hat{\theta})$ относительно θ ».

¹ Координационный центр данных обследования GATS в CDC оказывает техническое содействие в использовании следующих пакетов статистических программ: SAS, SPSS и STATA.

С.6 Оценки ошибок выборки

При оценке исследуемой выборки могут иметь место два типа ошибок: (1) ошибки, не связанные с выборкой, и (2) ошибки выборки. *Ошибки, не связанные с выборкой*, являются результатами ошибок, которые не имеют отношения к составлению выборки, и были допущены при сборе и обработке данных, как например, ошибки охвата, ошибки при ответах, ошибки из-за неполучения ответа, ошибки вопросника, ошибки записей интервьюера, ошибки обработки данных и т.д. Несмотря на то, что в ходе проведения обследования GATS в [стране] предпринимались многочисленные усилия для сведения к минимуму подобных ошибок, не связанных с выборкой, их невозможно избежать и трудно оценить статистически.

Выборка респондентов, сделанная в [стране проведения обследования GATS], является лишь одной из возможных выборок, которые могут быть взяты в определенной популяции при аналогичном плане и размере выборки. Каждая из таких выборок может дать результаты, которые будут несколько отличаться от фактически взятой выборки. *Ошибки выборки* являются показателем меры рассеяния между всеми возможными выборками. Степень рассеяния неизвестна в точности, однако может быть оценена статистически на основе результатов обследования.

По каждому из выбранных показателей учитывают следующие характеристики ошибок выборки:

- **Значение (R):** взвешенная оценка распространенности показателя.
- **Стандартная ошибка (SE):** ошибки выборки, как правило, измеряют как стандартную ошибку для конкретной оценки или показателя (R). Таким образом, стандартная ошибка выборки представляет собой квадратный корень из дисперсии этой оценки и вычисляется в тех же единицах, что и оценка.
- **Размер выборки (n):** общее количество наблюдений, использованных для подсчета распространенности оценки (R).
- **Эффект плана выборки:** эффект плана выборки, обозначаемый как "deft" – это отношение фактической дисперсии показателя согласно методу выборки, использованному в обследовании, к дисперсии, вычисляемой на основе предположения о простой случайной выборке. Квадратный корень из величины эффекта плана выборки, обозначаемый как "deft", используется для того, чтобы продемонстрировать эффективность плана выборки, и подсчитывается для каждой оценки как отношение между стандартной ошибкой при определенном плане выборки и стандартной ошибкой, которая бы возникла при использовании простой случайной выборки. Значение DEFT, равное 1,0, свидетельствует о том, что эффективность плана выборки равна эффективности простой случайной выборки, тогда как величина DEFT выше 1,0 указывает на увеличение стандартной ошибки из-за использования более сложной схемы выборки. В целом при дизайне исследования надлежащего качества величина DEFT, как правило, находится в диапазоне от 1 до 3. Вместе с тем величина DEFT нередко бывает существенно больше и составляет от 7 до 8.
- **Относительная стандартная ошибка (RSE):** относительная стандартная ошибка, так же называемая коэффициентом вариации (CV), является отношением стандартной ошибки к значению показателя.
- **Предел ошибки (MOE):** предел ошибки вычисляется как произведение показателя необходимой степени достоверности и стандартной ошибки оценки. Уровень

достоверности обычно основывается на значении (Z) стандартного нормального распределения. Например, для уровня достоверности 95% может быть использована величина $Z = 1,96$.

- **Границы доверительного интервала ($R \pm 1.96SE$):** рассчитываются для определения интервала, в предел которого, по разумному предположению, будет попадать действительное значение для популяции. Для любого статистического параметра, вычисленного на основе данных исследования, значения такого параметра будут попадать в предел диапазона плюс или минус две величины стандартной ошибки этих статистических данных в 95 процентах всех возможных выборок идентичного размера и состава.

Вычисление стандартной ошибки

Если бы выборка респондентов была произведена как простая случайная выборка, было бы возможным использование простых формул для вычисления ошибок выборки. Тем не менее, выборка обследования GATS от [года] произведена многоэтапная стратифицированная и, следовательно, необходимо применение более сложных формул. Для вычисления ошибок выборки данных [страны] GATS, применялась [версия статистического программного обеспечения]. В отношении оценок обследования, которые являются средними значениями или пропорциями, для оценки дисперсии применялся метод рядов Тейлора.

В рамках метода Тейлора любая процентная доля или средняя величина рассматривается как оценка отношения $r = y/x$, где y – общее по выборке значение для переменной y , а x – общее количество случаев в изучаемой группе или подгруппе. Дисперсия r вычисляется с помощью представленной ниже формулы:

$$SE^2(r) = \text{var}(r) = \frac{1-f}{x^2} \sum_{h=1}^2 \left[\frac{m_h}{m_h-1} \left(\sum_{i=1}^{m_h} Z_{hi}^2 - \frac{Z_h^2}{m_h} \right) \right],$$

$$\text{в которой } Z_{hi} = y_{hi} - rx_{hi}, \text{ и } Z_h = y_h - rx_h,$$

где h (=1 или 2) означает страту, выделенную по критерию проживания в городской или сельской местности,

m_h – общее количество ПЕО, выбранных в h -й страте,

y_{hi} – сумма взвешенных значений переменной y в i -й ПЕО в h -й страте,

x_{hi} – сумма взвешенных количеств наблюдений в i -й ПЕО в h -й страте;

f – совокупная величина части выборки, которой можно пренебречь ввиду малого размера.

В данном добавлении результаты представлены по стране в целом, в разбивке по критерию городских и сельских районов, а также в разбивке по полу. В таблице С-5 представлен тип статистического показателя (средняя величина, пропорция или отношение) для каждой переменной или показателя, а также базисное население. Помимо описанной выше стандартной ошибки (SE), таблица С-6 содержит значение оценочного параметра (R), размер выборки, эффект плана выборки (DEFF или DEFT), относительную стандартную ошибку (SE/R), предел ошибки (МОЕ) и границы 95-процентного доверительного интервала ($R \pm 1.96SE$) для каждой переменной или показателя.

Таблица С-5: перечень показателей по ошибкам выборки, GATS [страна] [год]

Показатель	Оценка	Базисное население
Лица, которые в настоящее время потребляют табак	Доля	Взрослые в возрасте 15 лет или старше
Лица, которые в настоящее время потребляют курительный табак	Доля	Взрослые в возрасте 15 лет или старше
Лица, которые в настоящее время потребляют сигареты	Доля	Взрослые в возрасте 15 лет или старше
Лица, которые в настоящее время потребляют бездымный табак	Доля	Взрослые в возрасте 15 лет или старше
Лица, которые ежедневно потребляют табак	Доля	Взрослые в возрасте 15 лет или старше
Лица, которые ежедневно потребляют курительный табак	Доля	Взрослые в возрасте 15 лет или старше
Лица, которые в настоящее время ежедневно потребляют сигареты	Доля	Взрослые в возрасте 15 лет или старше
Лица, которые ежедневно потребляют бездымный табак	Доля	Взрослые в возрасте 15 лет или старше
Лица, которые ранее ежедневно потребляли табак (среди всех взрослых)	Доля	Взрослые в возрасте 15 лет или старше
Лица, которые ранее потребляли табак (среди всех лиц, когда-либо потреблявших табак ежедневно)	Доля	Лица в возрасте 15 лет или старше, которые когда-либо потребляли табак ежедневно
Лица, которые ранее потребляли табак курительный ежедневно (среди всех взрослых)	Доля	Взрослые в возрасте 15 лет или старше
Лица, которые ранее потребляли курительный табак (среди лиц, когда-либо куривших ежедневно)	Доля	Лица в возрасте 15 лет или старше, которые когда-либо ежедневно потребляли курительный табак
Лица, которые ранее ежедневно потребляли бездымный табак (среди всех взрослых)	Доля	Взрослые в возрасте 15 лет или старше
Лица, которые ранее ежедневно потребляли бездымный табак (среди лиц, когда-либо ежедневно потреблявших бездымный табак)	Доля	Лица в возрасте 15 лет или старше, которые когда-либо ежедневно потребляли бездымный табак
Время до первого потребления табака (в пределах 5 минут после пробуждения)	Доля	Лица в возрасте 15 лет или старше, которые потребляют табак ежедневно
Время до первого потребления табака (в пределах 6–30 минут после пробуждения)	Доля	Лица в возрасте 15 лет или старше, которые потребляют табак ежедневно
Попытка отказа от курения в течение последних 12 месяцев	Доля	Среди лиц, которые потребляют курительный табак в настоящее время или потребляли ранее и отказались от табака менее 12 месяцев назад
Попытка отказаться от потребления бездымного табака в течение последних 12 месяцев	Доля	Лица, которые потребляют бездымный табак в настоящее время, а также лица, которые потребляли бездымный табак ранее, однако отказались от такого потребления менее 12 месяцев назад
Вопрос о потреблении курительного табака со стороны поставщика медицинских услуг	Доля	Лица, которые потребляют курительный табак в настоящее время или потребляли ранее и отказались от табака менее 12 месяцев назад, а также посещали организации, предоставляющие медицинские услуги, в течение последних 12 месяцев

Показатель	Оценка	Базисное население
Вопрос об отказе от курительного табака со стороны поставщика медицинских услуг	Доля	Лица, которые потребляют курительный табак в настоящее время или потребляли ранее и отказались от табака менее 12 месяцев назад, а также посещали организации, предоставляющие медицинские услуги, в течение последних 12 месяцев
Использование фармакотерапии для прекращения курения	Доля	Среди лиц, которые потребляют курительный табак в настоящее время или потребляли ранее и отказались от табака менее 12 месяцев назад
Использование консультационных/информационных услуг или телефонных линий для помощи в прекращении курения	Доля	Среди лиц, которые потребляют курительный табак в настоящее время или потребляли ранее и отказались от табака менее 12 месяцев назад
Планы отказа от курения, рассмотрение варианта отказа от курения или намерение отказаться от курения	Доля	Лица в возрасте 15 лет или старше, которые курят в настоящее время
Пассивное курение дома	Доля	Взрослые в возрасте 15 лет или старше
Пассивное курение на рабочем месте	Доля	Взрослые, работающие в закрытых помещениях
Пассивное курение в зданиях государственных учреждений/служб	Доля	Взрослые в возрасте 15 лет или старше
Пассивное курение в учреждениях здравоохранения	Доля	Взрослые в возрасте 15 лет или старше
Пассивное курение в ресторанах	Доля	Взрослые в возрасте 15 лет или старше
Пассивное курение в общественном транспорте	Доля	Взрослые в возрасте 15 лет или старше
Последняя покупка сигарет в магазине	Доля	Лица в возрасте 15 лет или старше, которые в настоящее время курят сигареты промышленного производства
Последняя покупка сигарет у уличного торговца	Доля	Лица в возрасте 15 лет или старше, которые в настоящее время курят сигареты промышленного производства
Последняя покупка сигарет в киоске	Доля	Лица в возрасте 15 лет или старше, которые в настоящее время курят сигареты промышленного производства
Заметность антитабачной информации по радио или телевидению	Доля	Взрослые в возрасте 15 лет или старше
Заметность предостережений о вреде здоровью на пачках сигарет	Доля	Лица в возрасте 15 лет или старше, которые курят в настоящее время
Рассмотрение варианта отказа от курения из-за предостережений о вреде здоровью на пачках сигарет	Доля	Лица в возрасте 15 лет или старше, которые курят в настоящее время
Заметность любой рекламы или стимулирования продаж сигарет	Доля	Взрослые в возрасте 15 лет или старше
Мнение о том, что курение табака является причиной серьезных заболеваний	Доля	Взрослые в возрасте 15 лет или старше
Мнение о том, что курение табака является причиной инсультов	Доля	Взрослые в возрасте 15 лет или старше
Мнение о том, что курение табака является причиной инфарктов	Доля	Взрослые в возрасте 15 лет или старше

Показатель	Оценка	Базисное население
Мнение о том, что курение табака является причиной рака легких	Доля	Взрослые в возрасте 15 лет или старше
Мнение о том, что потребление бездымного табака является причиной серьезных заболеваний	Доля	Взрослые в возрасте 15 лет или старше
Мнение о том, что пассивное курение является причиной серьезных заболеваний у некурящих	Доля	Взрослые в возрасте 15 лет или старше
Количество выкуриваемых за день сигарет (ежедневными курильщиками)	Среднее значение	Лица в возрасте 15 лет или старше, которые в настоящее время ежедневно потребляют сигареты
Время с момента отказа от курения (в годах)	Среднее значение	Лица в возрасте 15 лет или старше, которые курили ранее
Ежемесячные расходы на сигареты промышленного производства	Среднее значение	Лица в возрасте 15 лет или старше, которые в настоящее время курят сигареты промышленного производства
Возраст начала ежедневного курения	Среднее значение	Лица в возрасте 15 лет или старше, которые когда-либо ежедневно потребляли курительный табак

Показатель	Оценочное значение (R)	Стандартная ошибка (SE)	Размер выборки (n)	Эффект плана выборки (DEFF или DEFT)	Относительная ошибка (SE/R)	Предел ошибки (MOE)	Границы доверительного интервала	
							Нижняя граница (R-1.96SE)	Верхняя граница (R+1.96SE)
Количество выкуриваемых за день сигарет (ежедневными курильщиками)								
Время с момента отказа от курения (в годах)								
Ежемесячные расходы на сигареты промышленного производства								
Возраст начала ежедневного курения								

Примечания.

1. Таблицы должны быть разработаны по подгруппам (например, по критерию пола, проживания в городской/сельской местности и региону). Определения и вычисления перечисленных показателей представлены в документе «Определения показателей обследования GATS».
2. Показатели, представленные в этой таблице, получены из основного вопросника. Они могут быть соответствующим образом адаптированы для каждой страны. Кроме того, в определенных случаях страны могут посчитать необходимым получение оценок по подгруппам показателей, а не по всем показателям.
3. Страны могут захотеть получить либо 'deff', либо 'deft', основываясь на стандартной практике, применимой в стране /организации.
4. Приведенные выше таблицы могут быть включены в приложение странового отчета в качестве оценок ошибок выборки для описания качества данных.

С.7 Уровень охвата домохозяйств рамкой выборки

Общие сведения

Основной частью процесса составления выборки в обследовании домохозяйств является формирование перечня домохозяйств на стадии выборки, в котором выбираются адреса. Могут быть использованы различные подходы к формированию перечня домохозяйств, и обычным конечным результатом является неполный перечень домохозяйств и, соответственно, неполный охват в конечной выборке домохозяйств. Так как эффект ошибки, вызванной неполнотой перечня, зависит от: (i) совокупной разницы между основными показателями обследования по домохозяйствам, охваченным и не охваченным данной рамкой, так и от (ii) степени неполноты охвата (или 100 процентов минус уровень охвата домохозяйств) в выборке, целесообразно оценить уровень охвата домохозяйств рамкой выборки.

Уровень охвата рамкой выборки (FCR) определяется как

$$FCR = \frac{N_F}{N} = \frac{\text{Количество домохозяйств, охваченных рамкой выборки домохозяйств}}{\text{Количество домохозяйств в популяции}}. \quad (1)$$

Источник(и) данных

Для этих расчетов будут необходимы два источника данных. Один из них – файл с окончательно взвешенными аналитическими данными, используемый для анализа, с помощью которого определяют N_F в уравнении 1. Другой – оптимальный из имеющихся источник численности домохозяйств в стране, который используется как показатель N уравнении 1. В качестве последнего источника может выступать самая актуальная перепись населения, позднейшие обновления последней переписи населения с использованием демографических методов или недавнее национальное обследование, в котором размер и качество выборки, по меньшей мере, соответствуют таковым в обследовании GATS.

Программное обеспечение для вычислений

Какого-либо специализированного оборудования не требуется, так как необходимые вычисления могут быть легко запрограммированы в любой программной среде.

Вычисление

Количество домохозяйств, охваченных рамкой выборки домохозяйств (N_F), может быть оценено любым из двух способов. В обоих случаях необходимы специфичные для конкретной стадии выборочные вероятности, применявшиеся для вычисления базисного удельного веса по каждому респонденту. Одна является взвешенной оценкой N_F с использованием основных весов для «сегментов» выборки в схеме (подход с применением взвешенного суммарного значения); другая представляет собой сумму базисных весов для всех домохозяйств в выборке обследования GATS, которая также является оценкой N_F (подход с применением суммы удельных весов домохозяйств).

Подход с применением взвешенного суммарного значения: в большинстве схем дизайна обследования GATS сегментами будут являться вторичные единицы отбора (ВЕО), однако в целом они представляют собой единицы выборки в конкретном районе, для которого составлен нумерованный перечень домохозяйств и получены рамки выборки, с использованием которых выбирались конкретные домохозяйства. Исходя из предположения о том, что сегментом является ВЕО, W_j определяют как базисный удельный вес для j -го сегмента выборки и затем проводят

вычисления путем деления единицы на произведение вероятностей попадания в выборку для ПЕО, в которой расположен сегмент, и вероятности попадания в выборку сегмента, находящегося внутри ПЕО. Если M_j обозначает количество домохозяйств в нумерованном перечне в j -м сегменте выборки, оценка N_F в рамках данного подхода вычисляется как

$$\hat{N}_F^{(i)} = \sum_j^{All \text{ sample segments}} W_j M_j \quad (2)$$

Подход с применением суммы удельных весов домохозяйств: в рамках этого подхода используется базисный вес каждого выбранного домохозяйства. Набор выбранных домохозяйств включает в себя те домохозяйства, которые назначены для выездных работ с домохозяйствами во время сбора данных GATS. Пусть W_{jk} обозначает базисный вес для k -го выбранного домохозяйства в j -м сегменте выборки. W_{jk} вычисляются как отношение W_j к вероятности попадания в выборку внутри сегмента k -го домохозяйства. Затем в рамках данного подхода определяют оценочную величину N_F как

$$\hat{N}_F^{(ii)} = \sum_{j,k}^{All \text{ selected households}} W_{jk} \quad (3)$$

Оптимальными источниками пригодной величины, которая отражает реальное количество домохозяйств в популяции (N) являются: (1) наиболее актуальная перепись населения (при условии, что она является качественной), или (2) ориентировочная численность домохозяйств для N_F (полученная с использованием уравнений 2 или 3) на основе недавнего обследования с большей статистической выборкой (например, контрольная выборка, из которой была произведена случайная подвыборка для обследования GATS).

Интерпретация

Расчетная величина (FCR) на уровне 95% или выше будет означать, что качество нумерованного перечня домохозяйств в выборке обследования GATS являлось приемлемым.

Global Adult Tobacco Survey (GATS)