



Global Adult Tobacco Survey (GATS)



**Руководство по определению удельного веса
элементов выборки**

**Глобальный опрос взрослого населения
о потреблении табака (GATS)
Руководство по определению
удельного веса элементов
выборки**

Сентябрь 2020 г.

Глобальное обследование употребления табака среди взрослых (GATS) Комплексный стандартный протокол

Вопросник GATS

Основной вопросник с дополнительными вопросами
Повопросные разъяснения

План выборки обследования GATS

Рекомендации по проведению инструктажа в сфере работы с данными
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 в сфере работы с данными

Обеспечение качества GATS: руководства и документация

Пакет материалов для проведения анализа и подготовки отчетности по обследованию GATS

Шаблоны информационных бюллетеней
Страновой отчет: план и руководство по составлению таблиц
Определения показателей

Опубликование 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). Руководство по определению удельного веса элементов выборки. 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 США (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 настоящего руководства	1-2
2. Обзор удельного веса элементов выборки в обследовании GATS	2-1
3. Рекомендуемый подход	3-1
3.1 Описание наглядного примера	3-1
3.2 Базовое значение удельного веса	3-2
3.3 Корректировка с учетом доли не ответивших	3-5
3.4 Калибровка методом корректировки после стратификации	3-12
4. Обеспечение качества удельного веса элементов выборки GATS	4-1
4.1 Тщательное документирование этапов выборки и соответствующих вероятностей отбора	4-1
4.2 Контролирование отрицательного статистического эффекта различных значений удельного веса	4-2
4.3 Тщательное соблюдение рекомендуемых в настоящем руководстве вычислительных процедур и процедур проверки качества	4-3
4.4 Подготовка подробного документа с описанием фактической процедуры вычисления удельных весов	4-4
4.5 Консультирование с сотрудниками центрального органа GATS при вычислении удельных весов	4-4
5. Библиография	5-1

1. Введение

Употребление табака является основной предотвратимой причиной преждевременной смертности и заболеваемости во всем мире, вместе с тем порядка 1,4 миллиарда человек 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.

Глобальный опрос взрослого населения о потреблении табака (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 уровнях 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 интервью) 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; электронных сигарет; бездымного табака; о воздействии табачного дыма; экономические аспекты; средства массовой информации; и знания, отношение и восприятие различных аспектов, связанных с потреблением табака..

1.2 Использование настоящего руководства

Целью настоящего руководства является предоставление странам требований по вычислению удельного веса элементов выборки для обследования GATS. В первой главе сообщаются общие сведения о проводимом обследовании. Краткое изложение остальных глав руководства представлено ниже.

- В главе 2 содержится определение удельного веса элемента выборки, указывается, на каких этапах обследования следует проводить вычисление удельных весов, приводится

краткое обоснование вычислительного процесса и описывается, как удельные веса используются специалистами по анализу данных.

- В главе 3 представлено достаточно подробное описание всех этапов рекомендуемого подхода к вычислению удельных весов. Для каждого из этих этапов приводится пример, в котором используется тип респондента, который может присутствовать в выборке GATS.
- В главе 4 содержится ряд рекомендуемых мер по обеспечению качества вычисления удельных весов.

В настоящем руководстве предлагается поэтапный шаблон процедуры вычисления удельного веса элементов выборки обследования. Однако следует отметить, что в связи со спецификой каждой отдельной страны в отношении плана выборки, задач анализа и наличия вспомогательных данных для вычисления корректировок могут потребоваться определенные изменения данного шаблона. Для обеспечения качества вычисления и корректировки удельного веса элементов выборки существенно важное значение имеет сотрудничество национальных специалистов по статистике, национальных координаторов CDC и членов Комитета по рассмотрению выборки (КРВ).

2. Обзор удельного веса элементов выборки в обследовании GATS

Удельный вес элемента выборки представляет собой статистическое измерение, связанное с записью данных по тому или иному респонденту обследования в выборке населения, при построении которой в полной мере использовались методы случайного отбора. В общих чертах удельный вес отдельного респондента выборки обычно представляет собой обратную величину скорректированной вероятности получения данных по этому респонденту. В большинстве случаев эта вероятность является изначальной вероятностью отбора респондента, обусловленной планом выборки. Как правило, для устранения непреднамеренной несбалансированности выборки, возникающей в ходе проведения обследования, проводится корректировка обратной величины вероятности, или базового значения удельного веса (B_j)⁴. Могут применяться несколько методов корректировки удельных весов. Все они носят мультипликативный характер. Если для целей анализа не проводится нормализация удельного веса (например, «калибровка» к общей численности охваченного населения), значение удельного веса можно интерпретировать как показатель числа членов населения, представленных респондентом. При сборе подлежащих анализу данных по различным видам данных или единицам анализа, связанным с респондентом, может потребоваться подготовка отдельных наборов удельных весов. Например, если в рамках обследования домохозяйств собираются данные по выбранным домохозяйствам и по одному участнику, случайным образом выбранному в каждом из этих домохозяйств, то для данных по домохозяйствам и для данных по участникам будут подготовлены отдельные наборы удельных весов. Однако, поскольку в обследовании GATS обрабатываются только данные индивидуального уровня, вычисление удельных весов потребуется только на уровне респондентов.

Несмотря на общепризнанную статистическую обоснованность использования удельных весов для получения оценок по выборке населения (Horvitz and Thompson, 1952), универсального протокола по их вычислению не существует. Это отчасти объясняется различной спецификой выборок с точки зрения плана выборки и качества документации по процедурам составления выборки и набора в обследование, а также наличия вспомогательных данных по выборке и группе населения, которые могут использоваться для выявления и устранения несбалансированности выборки, обусловленной различиями в охвате и уровнях отсутствия ответа. Таким образом, фактические этапы вычисления удельного веса элементов выборки в разных обследованиях могут различаться. Вместе с тем при расчете удельного веса для записи данных по каждому (например, j -му) респонденту обычно используется определенная комбинация этапов, при этом итоговое скорректированное значение удельного веса (W_j) является произведением значений, полученных на каждом этапе. Подробное описание данных этапов приводится ниже.

1. Определить базовое значение удельного веса (B_j) для учета всех этапов случайного отбора, в результате которых была сформирована данная выборка членов населения.

⁴ Понятие «несбалансированности» в данном случае просто указывает на то, что демографическая представленность включенной в выборку группы населения в некоторой степени искажается факторами, связанными с этапами составления выборки и набора в обследование. Другими словами, в иных отношениях репрезентативная выборка, полученная методом случайного отбора, представляет данную группу населения с несколько меньшей точностью.

2. Провести корректировку с учетом доли не ответивших ($A_j^{(nr)}$) для устранения несбалансированности выборки, обусловленной различными результатами набора в обследование.
3. Провести дополнительную корректировку с учетом неполноты охвата ($A_j^{(cov)}$) для устранения несбалансированности выборки, связанной с тем, что структура выборки не в полной мере охватывает выбранную для обследования группу населения.
4. Провести дополнительную корректировку для калибровки ($A_j^{(cal)}$) итогового набора скорректированных значений удельного веса к распределению населения по характеристикам, напрямую связанным с основными показателями итоговых результатов обследования (т. е. моделями потребления табака в обследовании GATS).

Поскольку удельные веса, полученные на каждом отдельном этапе, зависят от результатов вычисления на предыдущем этапе, важное значение имеет фактическая последовательность этапов подготовки набора статистических весов для выборки населения. Итоговый удельный вес любого респондента представляет собой произведение результатов вычислений для данного респондента на всех этапах в следующей последовательности:

$$W_j = B_j A_j^{(nr)} A_j^{(cov)} A_j^{(cal)} \quad (1)$$

Чтобы понять логическое построение процедуры вычисления удельных весов для выборки обследования, в которой членами целевой группы населения являются жители выбранных домохозяйств, следует прежде всего принять во внимание, что данные по тому или иному члену населения в обследовании GATS могут быть получены только в случае наступления всех трех описанных ниже событий (т. е. событий F, S и R):

- **событие F:** домохозяйство данного члена населения и сам член населения включены в структуры выборок домохозяйств и респондентов домохозяйств, используемые для построения выборки обследования GATS;
- **событие S:** при условии наступления события F домохозяйство данного члена населения и сам член населения случайным образом выбраны для участия в обследовании GATS; и
- **событие R:** при условии наступления события S согласие на участие в обследовании GATS и заполнение соответствующего вопросника получено как на уровне домохозяйства данного члена населения, так и от самого члена населения.

Вероятность того, что для получения оценок по выборке будут использоваться данные этого члена населения, является произведением вероятности наступления трех указанных событий. Далее будет показано, что для выполнения этапов (1), (2) и (3) необходимо определить или оценить вероятность наступления событий F, S и R. Для предотвращения потери внешней значимости выборки из-за факторов, не учтенных при проведении корректировок на отсутствие ответа и неполноту охвата на этапах (2) и (3) используется калибровка (или «корректировка после стратификации»). К числу таких факторов относятся случайные изменения демографического состава выборки по переменным, не используемым для выделения страт, а также различные факторы отсутствия ответа и неполноты охвата, связанные с другими переменными, помимо тех, что используются для определения корректирующих ячеек для корректировки на этапах (2) и (3).

Поскольку проведение этапа (3) часто практически нецелесообразно, корректировка с учетом неполноты охвата структурой выборки обычно проводится в процессе калибровки.

Удельные веса имеют важное значение для многих типов аналитической деятельности. Например, они используются для получения точечных оценок по характеристикам населения (например, показателю нынешней распространенности потребления табака) и для оценки дисперсии этих характеристик, хотя для последней важны и другие особенности плана выборки (такие как стратификация, кластерная выборка, отбор без возвращения). Таким образом, удельные веса имеют важное значение для многих типов использования данных, включая описательный анализ и сравнительное тестирование (например, в случае значительных различий в показателях распространенности потребления табака между регионами одной страны). Иногда они также используются при регрессионном моделировании (например, для определения предикторов моделей потребления табака). Поэтому удельные веса зачастую являются необходимым, но не всегда достаточным компонентом плана выборки для анализа данных в рамках обследования. Например, если при вычислении дисперсии оценок обследования используются только удельные веса, но не учитываются стратификация и кластерная выборка, то полученные значения дисперсии и выводы, сделанные в ходе проверки гипотез, вполне могут оказаться неверными.

Для каждого набора единиц наблюдения в обследовании рассчитывается один набор удельных весов. В обследованиях часто используются сразу несколько наборов единиц наблюдения. В частности, если в плане сбора данных для обследования предусмотрен сбор данных по домохозяйствам (например, доход, социально-экономическое положение, удаленность от учреждений здравоохранения) и по отдельным участникам этих домохозяйств (например, модели потребления табака, демографические характеристики и пр.), то для каждого набора единиц наблюдения (т. е. домохозяйств и респондентов) может потребоваться отдельный набор удельных весов. Кроме того, если для получения данных на индивидуальном уровне в домохозяйствах проводится отбор респондентов, удельные веса домохозяйств и респондентов из числа респондентов этих домохозяйств могут различаться. Если бы в рамках GATS проводился сбор данных как на уровне домохозяйств, так и на уровне респондентов, в настоящем обследовании имели бы место такого рода различия, поскольку из каждого включенного в обследование домохозяйства выбирается один подходящий для участия в обследовании житель.

В качестве заранее сформированных выборок геополитических кластеров, отвечающих исследовательским потребностям различных организаций, могут использоваться основные выборки при наличии таковых в стране (см. раздел 3.3 Руководства GATS по составлению плана выборки обследования). Затем для получения кластеров домохозяйств, необходимых для составления перечней домохозяйств и проведения выборки в соответствии с Руководством GATS по составлению карт и перечней, из основных выборок географических единиц формируется подвыборка. План выборки GATS должен соответствовать плану основной выборки страны, и на момент проведения выборки обследования GATS должна быть обеспечена возможность отбора единиц основной выборки. Таким образом, вероятность отбора кластера GATS является произведением первоначальной вероятности отбора единицы основной выборки и вероятности отбора кластера домохозяйств GATS. Ввиду того, что в разных странах используется различный план подвыборки из основных выборок, для обеспечения точного вычисления вероятностей отбора и соответствующих удельных весов существенно важное значение имеет тесное

сотрудничество национального статистического управления, национального координатора по GATS, координаторов и специалистов по статистике CDC и членов Комитета по рассмотрению выборки. В приведенных далее вычислениях предполагается, что при отборе не использовались основные выборки, а геополитические кластеры были выбраны непосредственно из общенациональной структуры всех таких единиц; нередко такую структуру предоставляет национальное статистическое управление по вопросам переписи населения. Страны, использующие основные выборки для проведения первого этапа отбора, при вычислении вероятности отбора кластеров GATS должны учитывать вероятность отбора единиц основной выборки.

Вычисление удельных весов обычно проводится после сбора, обработки и очистки данных для дальнейшего использования специалистами по анализу. Поскольку удельные веса применяются к конечной выборке респондентов, и для их вычисления используются итоговые результаты сбора данных, вычисление удельных весов не может быть проведено до завершения полевых работ. Кроме того, поскольку удельные веса будут использоваться специалистами по анализу, работающими с данными обследования, их вычисление должно быть завершено до начала анализа. Как правило, в силу специфических знаний о целевой группе населения и методах составления выборки и проведения набора в обследование, для вычисления удельных весов наилучшим образом подготовлены технические специалисты, участвующие в составлении выборки и оказывающие поддержку в проведении полевых работ. По мере необходимости помощь и содействие в вычислении удельных весов могут оказывать сотрудники центрального органа GATS. Наконец, чтобы гарантировать соблюдение соответствующих процедур по обеспечению качества удельных весов, весь процесс их вычисления проходит проверку внешних экспертов (см. раздел 5 документа «Обеспечение качества GATS: рекомендации и документация»).

3. Рекомендуемый подход

В настоящей главе достаточно подробно излагается трехступенчатый подход, соответствующий основным этапам, рекомендуемые техническим специалистам страны при вычислении удельных весов для выборки обследования GATS. В данном подходе предполагаются следующие этапы: вычисление базовых значений удельного веса для всех респондентов выборки; корректировка базовых значений удельного веса с учетом доли не ответивших; и калибровка скорректированных значений удельного веса к известной численности населения. Более общее описание каждого из этих этапов можно найти в обзорах Лесслера и Калсбика (Lessler and Kalsbeek, 1992) и Калтона и Флорес-Сервантеса (Kalton and Flores-Cervantes, 2003). Изложение данных этапов в настоящем руководстве подкрепляется реалистичным примером респондента, который может присутствовать в обследовании GATS. Описание каждого из трех этапов начинается с перечисления требований, необходимых для его выполнения; далее, в качестве наглядного примера, представлены формулы, рекомендуемые для расчета соответствующего компонента удельного веса. Сотрудникам стран, проводящих обследование GATS, предлагается рассмотреть альтернативные варианты некоторых рекомендаций, однако выбор таких альтернатив должен быть в полной мере обоснован и осуществляться при всестороннем сотрудничестве с национальными координаторами CDC и Комитетом по рассмотрению выборки до начала их применения.

3.1 Описание наглядного примера

Приведенный ниже пример используется в качестве иллюстрации процесса вычисления удельных весов для обследования GATS; основное внимание уделяется определению итогового скорректированного значения удельного веса для вымышленного респондента (R) в выборке GATS, при построении которой не требуется проведение рандомизации по критерию пола. Мы исходим из того, что план выборки, в рамках которой был выбран этот респондент, представляет собой стратифицированную трехступенчатую выборку домохозяйств, в которой первичными единицами отбора (ПЕО) являются признанные геополитические территориальные единицы, вторичными единицами отбора (ВЕО) являются «сегменты» приблизительно одинакового размера, состоящие примерно из 200 домохозяйств, и во включенных в выборку сегментах выбраны отдельные домохозяйства. В данном примере предполагается, что при составлении перечней домохозяйств в составе сегмента респондента R не было выявлено пропущенных домохозяйств и что в ходе отбора респондента R не применялось распределение по критерию пола, однако при практической реализации эти вопросы, возможно, будет необходимо рассмотреть (оптимальные

РЕКОМЕНДАЦИЯ.

Странам рекомендуется применять к каждой выборке GATS трехступенчатый подход:

- (1) вычисление базового значения удельного веса для всех респондентов выборки;**
- (2) корректировка базовых значений удельного веса с учетом доли не ответивших; и**
- (3) калибровка скорректированных значений удельного веса к известной численности населения.**

НАГЛЯДНЫЙ ПРИМЕР.

Справочная информация по вымышленному респонденту GATS (R).

методы их решения приводятся далее)⁵. Наконец, предполагается, что респондент R был выбран из реестра респондентов домохозяйства с использованием портативного устройства и что это устройство также использовалось для проведения интервью с респондентом R. При описании каждого этапа будет приводиться другая предполагаемая информация, касающаяся отбора респондента R и его набора в обследование.

3.2 Базовое значение удельного веса

3.2.1 Исходные данные, требуемые для выполнения вычислительного этапа

- Тщательно задокументированные вероятности отбора на каждом этапе процесса отбора, если они уже были рассчитаны (что является предпочтительным вариантом).

ИЛИ

- Специальные знания о вычислении вероятностей в соответствии с методами случайного отбора, используемыми на каждом этапе. Процесс отбора включает следующие этапы: построение выборки, распределение домохозяйств по группам участия лиц мужского/женского пола (при необходимости) и, в применимых случаях, дополнительный отбор пропущенных домохозяйств, обнаруженных в ходе процедуры по выявлению пропущенных домохозяйств.
- Формы отбора и/или код для программного обеспечения, с использованием которого проводится отбор, для каждого этапа, *если* одновременно с вычислением удельных весов требуется определить вероятность отбора.

3.2.2 Наглядный пример вычисления базового значения удельного веса

В любой вероятностной выборке базовое значение удельного веса того или иного респондента представляет собой единицу, деленную на общую вероятность отбора данного респондента с учетом проведенных этапов, в результате которых был выбран данный респондент. Таким образом, для вычисления базового значения удельного веса респондента GATS требуется определить статистическую вероятность отбора данного респондента в рамках используемого плана выборки.

Соответствующие компоненты общей вероятности отбора для рекомендуемого многоступенчатого подхода к построению выборки излагаются в Руководстве GATS по составлению плана выборки обследования (глава 11). В приведенном далее описании нижние индексы α и k (для α -й ПЕО

⁵ Дополнительный отбор пропущенных домохозяйств требуется в тех случаях, когда в ходе процедуры по выявлению пропущенных домохозяйств (например, методом «полуоткрытого интервала») было обнаружено достаточно большое число пропущенных домохозяйств, при котором требуется проведение дополнительного случайного отбора отсутствующих в перечне домохозяйств. Распределение выборки GATS по критерию пола требуется в тех случаях, когда в силу причин культурного характера в стране планируется обеспечить совпадение пола интервьюеров с полом отобранных респондентов и привлекать к работе группы интервьюеров того же пола, что и респонденты; либо когда в силу причин статистического характера необходимо сформировать непропорционально большую выборку респондентов того или иного пола. Каждое включенное в выборку домохозяйство случайным образом распределяется в группу участия лиц мужского пола (в реестр респондентов для выбора респондента включаются только жители мужского пола, подходящие для участия в обследовании) или в группу участия лиц женского пола (в реестр включаются только жители женского пола, подходящие для участия в обследовании).

(первичной единицы отбора) и k -й ВЕО (вторичной единицы отбора) соответственно) совместно соответствуют «области» (т. е. сегменту) g , выбранной в два этапа для отбора респондента R:

- $p_{ak}^{(1)}$ = безусловная вероятность отбора α -й ПЕО (геополитической территориальной единицы, в которой проживает респондент R) и k -й ВЕО (сегмента, в котором проживает респондент R);
- $p_{ak}^{(2)}$ = условная вероятность (при условии выбора ПЕО и ВЕО) отбора домохозяйства, в котором проживает респондент R;
- $p_{aki}^{(3)}$ = для вымышленного респондента GATS (R) эта вероятность равна единице; если применяется рандомизация по критерию пола, необходимо вычислить условную вероятность (при условии выбора ПЕО, ВЕО и домохозяйства) случайного распределения домохозяйства респондента R к группе участия лиц женского пола;
- $p_{aki}^{(4)}$ = условная вероятность (при условии выбора ПЕО, ВЕО, домохозяйства и распределения по критерию пола) случайного отбора домохозяйства респондента R, если данное домохозяйство отсутствовало в изначальной структуре домохозяйств ak -го сегмента и было обнаружено в ходе процедуры по выявлению пропущенных домохозяйств; в противном случае эта вероятность равна единице; и
- $p_{akij}^{(5)}$ = условная вероятность (при условии выбора ПЕО, ВЕО, домохозяйства, распределения по критерию пола и выбора пропущенного домохозяйства) случайного отбора респондента R из реестра респондентов домохозяйства респондента R, соответствующих критериям участия в обследовании.

Следует отметить, что для отбора респондента R в выборку GATS требуется наступление всех событий отбора, соответствующих этим вероятностям. В этом случае совместная безусловная вероятность отбора респондента R ($akij$ -го лица) в выборку GATS вычисляется как

$$p_{akij} = p_{ak}^{(1)} \cdot p_{ak}^{(2)} \cdot p_{aki}^{(3)} \cdot p_{aki}^{(4)} \cdot p_{akij}^{(5)}, \quad (2)$$

поскольку совместная вероятность последовательных событий является безусловной вероятностью первого события в последовательности, помноженного на условные вероятности каждого события, следующего за предыдущим. Таким образом, соответствующее базовое значение удельного веса респондента R вычисляется как

$$B_{akij} = \frac{1}{p_{akij}} = \frac{1}{p_{ak}^{(1)} \cdot p_{ak}^{(2)} \cdot p_{aki}^{(3)} \cdot p_{aki}^{(4)} \cdot p_{akij}^{(5)}}. \quad (3)$$

Как видно из уравнения (3), для вычисления базового значения удельного веса респондента R необходимо определить каждый компонент совместной безусловной вероятности отбора этого респондента. Предположим, что для выбора ПЕО из страты выборки на первом этапе используется определенная форма отбора с вероятностью, пропорциональной размеру (ВНР), без возвращения, тогда, если $N_i = 2\,462$ соответствует показателю размера ПЕО респондента R (числу домохозяйств по состоянию на последнюю перепись населения), $I = 2$ соответствует числу выбранных ПЕО в страте выборки, из которой была выбрана ПЕО респондента R, и сумма всех показателей размера всех ПЕО в этой страте составляет $\sum_{\alpha} N_{\alpha} = 338\,754$, то безусловная вероятность отбора ПЕО респондента R составляет

$$p_{\alpha}^{(1)} = \frac{I \cdot N_{\alpha}}{\sum_{\alpha} N_{\alpha}} = \frac{(2) \cdot (2\,462)}{338\,754} = 1,4536 \times 10^{-2}. \quad (4)$$

Если ВЕО (вторичная единица отбора) респондента R является одним из $K_{\alpha} = 2$ сегментов, выбранных методом простого случайного отбора (без возвращения) из $S_{\alpha} = 12$ сегментов в составе ПЕО (первичной единицы отбора) респондента R, тогда условная вероятность (при условии выбора ПЕО) отбора ВЕО респондента R составляет

$$p_{k(\alpha)}^{(1)} = \frac{K_{\alpha}}{S_{\alpha}} = \frac{2}{12} = \frac{1}{6}, \quad (5)$$

и совместная безусловная вероятность отбора ПЕО и ВЕО респондента R составляет

$$p_{\alpha k}^{(1)} = p_{\alpha}^{(1)} p_{k(\alpha)}^{(1)} = \left[\frac{I \cdot N_{\alpha}}{\sum_{\alpha} N_{\alpha}} \right] \cdot \left[\frac{K_{\alpha}}{S_{\alpha}} \right] = (1,4536 \times 10^{-2}) \cdot (1/6) = 2,4226 \times 10^{-3}. \quad (6)$$

В Руководстве GATS по составлению плана выборки обследования описываются два метода отбора домохозяйств в выбранных сегментах или ПЕО. Если для отбора домохозяйств используется систематическая выборка, тогда условная вероятность отбора каждого домохозяйства составляет $\frac{1}{K}$, где K соответствует интервалу отбора. Если для отбора домохозяйств используется простая случайная выборка, тогда эта вероятность отбора представляет собой отношение размера выборки домохозяйств в каждом сегменте к общему числу домохозяйств в перечне структуры каждого сегмента. Предположим, что из $L_{\alpha k} = 212$ включенных в перечень домохозяйств в составе αk -го сегмента, в котором расположено домохозяйство респондента R, методом случайного отбора без возвращения выбраны $H_{\alpha k} = 28$ домохозяйств, тогда условная вероятность отбора домохозяйства респондента R составляет

$$p_{\alpha k}^{(2)} = \frac{H_{\alpha k}}{L_{\alpha k}} = \frac{28}{212} = 0,13208. \quad (7)$$

Поскольку в стране, где находится вымышленный респондент, не требуется рандомизация по критерию пола, то вероятность отбора респондента составляет

$$p_{\alpha k i}^{(3)} = 1. \quad (8)$$

Отметим, что по общему правилу, если требуется распределение домохозяйств по критерию пола, и в αk -й «области» (т. е. сегменте) выбрано $H_{\alpha k} = M_{\alpha k} + F_{\alpha k}$ домохозяйств, где $M_{\alpha k}$ и $F_{\alpha k}$ обозначают домохозяйства, отнесенные к группам участия лиц мужского и женского пола соответственно, тогда для респондентов мужского пола

$$p_{\alpha k i}^{(3)} = \frac{M_{\alpha k}}{H_{\alpha k}}, \quad (9)$$

и для респондентов женского пола

$$p_{\alpha k i}^{(3)} = \frac{F_{\alpha k}}{H_{\alpha k}}. \quad (10)$$

По общему правилу, если респондент выбран из домохозяйства, обнаруженного в ходе процедуры по выявлению пропущенных домохозяйств (например, методом полуоткрытого интервала), и при проведении дополнительной выборки домохозяйств методом простого случайного отбора

применяется показатель доли выборки ($f_{\alpha ki}$), тогда вероятность дополнительного отбора для любого (т. е. αki -го) выбранного домохозяйства составляет

$$p_{\alpha ki}^{(4)} = f_{\alpha ki}. \quad (11)$$

Поскольку мы предполагаем, что отбор домохозяйства респондента R не проводился в ходе процедуры по выявлению пропущенных домохозяйств, то

$$p_{\alpha ki}^{(4)} = 1. \quad (12)$$

Что касается конкретного случая вычисления вероятности отбора респондента R из домохозяйства, следует отметить, что данный респондент был выбран случайным образом из $R_{\alpha ki} = 4$ членов домохозяйства респондента, включенных в реестр респондентов домохозяйства. Таким образом, для домохозяйства

$$p_{\alpha kij}^{(5)} = \frac{1}{R_{\alpha ki}} = \frac{1}{4} = 0,25. \quad (13)$$

После объединения всех вероятностей процесса отбора, в результате которого был выбран респондент R, общая безусловная вероятность отбора для респондента R составляет

$$p_{\alpha kij} = p_{\alpha k}^{(1)} \cdot p_{\alpha k}^{(2)} \cdot p_{\alpha ki}^{(3)} \cdot p_{\alpha ki}^{(4)} \cdot p_{\alpha kij}^{(5)} = (2,4226 \times 10^{-3}) \cdot (0,13208) \cdot (1) \cdot (1) \cdot (0,25) = 7,9992 \times 10^{-5}, \quad (14)$$

при этом базовое значение удельного веса респондента R составляет

$$B_{\alpha kij} = \frac{1}{p_{\alpha kij}} = \frac{1}{7,9992 \times 10^{-5}} = 12501,3081, \quad (15)$$

и базовое значение удельного веса домохозяйства составляет

$$B_{\alpha ki} = p_{\alpha kij}^{(5)} / p_{\alpha kij} = (0,25) \cdot (12501,3081) = 3125,2370. \quad (16)$$

Следует отметить, что в следующих разделах настоящего руководства $B_{\alpha kij}$ обозначается как B_j .

3.3 Корректировка с учетом доли не ответивших

3.3.1 Исходные данные, требуемые для выполнения вычислительного этапа

- Базовые значения удельного веса для всех респондентов GATS.
- Итоговые диспозиционные категории для **всех выбранных домохозяйств**, указывающие на завершённый этап набора в обследование (например: «получен ответ», «отказ», «никого нет дома» и пр.), которые будут использоваться при вычислении доли ответивших на вопросник для домохозяйств в выборке домохозяйств, соответствующих критериям включения в обследование, в каждом «сегменте» выборки. Это означает, что для каждого сегмента выборки потребуется подсчет числа выбранных, соответствующих критериям обследования и включенных в обследование домохозяйств. В случае использования плана выборки, в рамках которого отбор домохозяйств проводится непосредственно на этапе отбора первичных единиц, для корректировки с учетом доли не ответивших на вопросник



для домохозяйств будут использоваться итоговые диспозиционные категории в каждой ПЕО.

- Итоговые диспозиционные категории для **всех выбранных респондентов** всех включенных в обследование домохозяйств, указывающие на заверченный этап набора в обследование, и данные из реестров респондентов домохозяйств по полу, возрасту и употреблению табака (курит/ не курит на данный момент). Для корректировки с учетом доли не ответивших может также использоваться дополнительная информация, полученная в ходе отбора ПЕО, например распределение по городской/сельской местности и данные по регионам. Эти вспомогательные сведения об изучаемой выборке будут использоваться для вычисления доли ответивших по выбранным переменным класса взвешивания. Важно, чтобы все переменные, используемые для вычисления доли ответивших, не содержали пропусков. Если неизвестно, употребляет ли респондент табак, то в целях корректировки с учетом взвешенной доли не ответивших на индивидуальный вопросник следует условно отнести этого респондента к категории «не употребляющих табак».

3.3.2 Наглядный пример корректировки с учетом доли не ответивших

Смещение из-за доли не ответивших, наблюдаемое в оценках по простым характеристикам населения, таким как численность в абсолютном, среднем и процентном выражениях, обусловлено исключительно данными на уровне респондентов, вместе с тем базовое значение удельного веса частично определяется ковариацией на уровне членов населения ($\sigma_{\pi y}$) между склонностью (т. е. вероятностью в стохастическом смысле) j -го члена населения отвечать на вопросы обследования (π_j) и показателями обследования (y_j) по оцениваемым областям (Lessler and Kalsbeek, 1992)⁶. Следует отметить, что мы временно считаем всех респондентов членами населения в целом и, таким образом, снимаем индекс (i) для домохозяйства, членом которого является данный респондент.

Для (целевых) групп населения размером N респондентов (который предположительно известен) можно записать смещение нескорректированной оценки ($\hat{y}_r = \frac{\hat{t}_r}{N} = \frac{\sum_{j=1}^N s_j r_j y_j / p_j}{v}$) средней численности населения, возникшее вследствие отсутствия ответа, как

($\bar{Y} = t/N$), поскольку

$$\text{Bias}(\hat{y}_r) = \frac{\sigma_{\pi y}}{\bar{\pi}}, \quad (17)$$

где для j -го члена населения: s_j — показатель 0/1 для события S , p_j — общая вероятность отбора (т. е. $Pr\{s_j = 1\}$), r_j — показатель 0/1 для события R и $\bar{\pi}$ — средняя величина всех значений склонности к ответу (π_j) в данной группе населения. Оценка склонности того или иного респондента к ответу используется для корректировки базового значения удельного веса данного респондента с учетом доли не ответивших. Корректировочным коэффициентом на долю не ответивших становится обратная величина этой предполагаемой склонности (т. е. единица, деленная на эту цифру), которая умножается на базовое значение удельного веса, полученное на

⁶ Поскольку итоговый результат процесса по установлению того, будет ли получен ответ от того или иного члена выборки обследования, определяется не рандомизацией по явным признакам, а предполагаемым стохастическим поведением, то применительно к доле не ответивших на обследование используется понятие «склонности», а не «вероятности».

предыдущем этапе, что позволяет вычислить удельный вес данного респондента, скорректированный на долю не ответивших.

Точнее говоря, при корректировке удельного веса элементов выборки на долю не ответивших требуется эмпирическим путем оценить каждое значение π_j исходя из полученных уровней отсутствия ответа в данной выборке. Таким образом, мультипликативная корректировка с учетом доли не ответивших на уровне членов населения представляет собой обратную величину предполагаемой склонности респондента к ответу:

$$A_j^{(nr)} = \hat{\pi}_j^{-1}, \quad (18)$$

и удельный вес респондента, скорректированный на долю не ответивших, составляет

$$W_j^{(2)} = B_j A_j^{(nr)}. \quad (19)$$

Таким образом, одной из главных задач при проведении такой корректировки является выбор метода оценки склонности респондента к ответу. Для корректировки с учетом доли не ответивших используются два подхода: доля ответивших в классе взвешивания и прогнозируемая склонность к ответу с использованием специально подобранной логистической модели. Поскольку подход на основе классов взвешивания более прост в применении, и было установлено, что прогнозирование склонности в большинстве случаев не обладает превосходством в том, что касается ограничения (**не** устранения) величины смещения, обусловленного отсутствием ответа, то рекомендуемым подходом для обследования GATS является корректировка по классу взвешивания.

В обследованиях GATS оценка склонности представляет собой несколько более сложный процесс, поскольку при наборе респондентов в выборку обследования отказ может иметь место как на уровне домохозяйств, так и на уровне отдельных лиц. Таким образом, совокупная склонность к ответу для j -го респондента (π_j) имеет два мультипликативных компонента, которые должны оцениваться по отдельности: один компонент отражает склонность к ответу на уровне домохозяйства респондента при составлении реестра членов домохозяйства ($\pi_j^{(HH)}$), а другой — склонность респондента к ответу в ходе интервью GATS, проводимого после выбора данного респондента из реестра респондентов домохозяйства ($\pi_j^{(person)}$).

В соответствии с рекомендуемым подходом к корректировке с учетом доли не ответивших каждый компонент склонности к ответу для j -го респондента оценивается как доля ответивших на обследование в подгруппе (т. е. «классе») выбранных членов выборки, обладающих схожими с респондентом характеристиками и предрасположенностью к ответу. Вычисление доли «ответивших» для компонента на уровне домохозяйств и компонента на уровне респондентов должно проводиться в соответствии с рекомендациями относительно взвешенной или невзвешенной доли ответивших RR1 согласно определению AAPOR (American Association of Public Opinion Research, 2009) и указаниям, изложенным в Руководстве GATS по составлению плана выборки обследования. Использование взвешенной доли ответивших с применением базовых значений удельного веса является более предпочтительным вариантом, чем использование невзвешенной доли ответивших. Окончательное решение о выборе того или иного варианта принимают страны. При использовании диспозиционных кодов домохозяйств доля ответивших на

вопросник для домохозяйств вычисляется отдельно для каждого сегмента выборки (см. описание ниже) по следующей формуле:

$$\text{доля ответивших на вопросник для домохозяйств} = \frac{[1]}{[1] + [3] + [4] + [5] + [6] + [9]} \quad (20)$$

где:

- 1 = завершен опрос по вопроснику для домохозяйств; выбран один участник;
- 2 = завершен опрос по вопроснику для домохозяйств; участники не выбраны;
- 3 = опрос по вопроснику для домохозяйств завершен частично, не удалось составить реестр (опрос завершен не полностью);
- 4 = опрос по вопроснику для домохозяйств не завершен; не удалось выявить соответствующего респондента скрининга;
- 5 = никого нет дома;
- 6 = отказ от заполнения вопросника для домохозяйств;
- 9 = другие причины отсутствия ответа на вопросник для домохозяйств.

Следует отметить, что итоговый диспозиционный код домохозяйств «2» исключен из числителя и знаменателя доли ответивших на вопросник для домохозяйств, поскольку такие домохозяйства считаются не соответствующими критериям включения в обследования вне зависимости от применения рандомизации по критерию пола.

Не соответствующие критериям обследования домохозяйства исключаются из знаменателя доли ответивших на вопросник для домохозяйств, при этом предполагается, что в обследование GATS будут включены все выбранные домохозяйства, для которых неизвестно соответствие критериям обследования (итоговые диспозиционные коды 3, 4, 5, 6, 9). Если домохозяйства, для которых неизвестно соответствие критериям обследования, достоверно не соответствуют этим критериям, доля ответивших на вопросник для домохозяйств может оказаться заниженной. Странам рекомендуется оценить эту процентную долю (e) путем вычисления доли домохозяйств, достоверно соответствующих критериям обследования, посредством деления числа домохозяйств, достоверно соответствующих критериям обследования (диспозиционный код 1) на сумму числа домохозяйств, достоверно соответствующих критериям обследования, и числа домохозяйств, достоверно не соответствующих критериям обследования (диспозиционный код 2):

$$e^{household-level} = \frac{[1]}{[1] + [2]} \quad (21)$$

Если эта доля во всей выборке составляет менее 0,90, странам следует скорректировать неизвестный компонент доли ответивших на вопросник для домохозяйств путем умножения числа домохозяйств, для которых неизвестно соответствие критериям обследования (итоговые диспозиционные коды 3, 4, 5, 6, 9), на данную процентную долю (e). Для вычисления доли ответивших RR3 согласно определению AAPOR (2009 г.) используется следующая формула:

$$\text{доля ответивших на вопросник для домохозяйств} = \frac{[1]}{[1] + e^{household-level} [3 + 4 + 5 + 6 + 9]} \quad (22)$$

При корректировке по классу взвешивания соответствующий компонент на уровне домохозяйств ($A_{hi}^{(nr,HH)}$) для каждого сегмента выборки будет рассчитываться как единица, деленная на взвешенную долю ответивших на вопросник для домохозяйств.

Доля ответивших на индивидуальный вопросник вычисляется по стратегически сформированным подгруппам (см. описание ниже) как

$$\text{доля ответивших на индивидуальный вопросник} = \frac{[11]}{[11]+[12]+[14]+[15]+[16]+[17]}, \quad (23)$$

где:

- 11 = завершен опрос по индивидуальному вопроснику;
- 12 = опрос завершен не полностью;
- 13 = выбранный респондент был впоследствии признан не соответствующим критериям участия в обследовании GATS;
- 14 = в настоящее время выбранный респондент отсутствует;
- 15 = отказ от заполнения индивидуального вопросника;
- 16 = выбранный респондент некомпетентен;
- 17 = другая причина отсутствия ответа на индивидуальный вопросник.

Не соответствующие критериям обследования респонденты (итоговый диспозиционный код 13) исключаются из знаменателя доли ответивших на индивидуальный вопросник, при этом предполагается, что в обследовании GATS могут участвовать все выбранные респонденты, для которых неизвестно соответствие критериям обследования (итоговый диспозиционный код 14). Если интервьюеры часто выбирают респондентов, которые при проведении интервью оказываются не соответствующими критериям обследования, доля ответивших на индивидуальный вопросник может оказаться заниженной. Странам рекомендуется оценить процентную долю респондентов, которые были выбраны из реестра и достоверно соответствуют критериям участия в обследовании GATS (e), с использованием взвешенной частоты диспозиционных кодов как

$$e^{person-level} = \frac{[11]+[12]+[15]+[16]+[17]}{[11]+[12]+[13]+[15]+[16]+[17]}. \quad (24)$$

Если эта доля в общей выборке составляет менее 0,90, странам следует скорректировать неизвестный компонент доли ответивших на индивидуальный вопросник путем умножения числа респондентов, для которых неизвестно соответствие критериям обследования (итоговый диспозиционный код 14) на данную процентную долю (e). Для вычисления доли ответивших RR3 согласно определению AAPOR (2009 г.) используется следующая формула:

$$\text{доля ответивших на индивидуальный вопросник} = \frac{[11]}{[11]+[12]+[e^{person-level}[14]+[15]+[16]+[17]}}. \quad (25)$$

При корректировке по классу взвешивания компонент на уровне респондентов ($A_{hj}^{(nr,person)}$) для j -го респондента в i -ом домохозяйстве будет вычисляться как единица, деленная на взвешенную долю ответивших на индивидуальный вопросник.

Корректировка по подгруппе или кластеру небольшого размера может характеризоваться значительной дисперсией и приводить к получению чрезмерно больших значений корректировки. Поэтому для всех вычисляемых корректировок по классам взвешивания на уровне домохозяйств и на уровне респондентов будет установлена верхняя граница со значением 3,00. Значение выше 3,00 в любом компоненте удельного веса корректировки должно быть установлен на уровне 3,00.

Наконец, для получения совокупного значения корректировки с учетом доли не ответивших для j -го респондента в i -м домохозяйстве ($A_j^{(nr)}$) необходимо умножить соответствующий компонент корректировки на уровне домохозяйств ($A_{hi}^{(nr,HH)}$) на компонент корректировки на уровне респондентов ($A_{hj}^{(nr,person)}$)

Поскольку при проведении корректировки по классам взвешивания уменьшение величины смещения прямо пропорционально повышению степени корреляции между долей ответивших и изучаемым параметром этих классов (Kalton, 1983), стратегически важное значение имеет выбор характеристик для выделения подгрупп (т. е. классов взвешивания) по каждому компоненту. В большинстве случаев классы взвешивания выделяются путем перекрестной классификации нескольких категориальных переменных (например, непрерывных переменных в разбивке по нескольким категориям, таких как переменные возраста респондентов в разбивке по возрастным группам).

Эти подгруппы могут представлять собой страты либо кластеры выборки или могут быть выделены по любому другому признаку, известному для всех выбранных членов выборки. При корректировке с учетом доли не ответивших в обследовании GATS рекомендуется, чтобы классы взвешивания для компонента на уровне домохозяйств определялись набором выбранных домохозяйств в составе «сегментов» выборки. Если для составления перечня домохозяйств отбор географических кластеров проводился в один этап, классами взвешивания для компонента на уровне домохозяйств будут ПЕО выборки. Применительно к компоненту корректировки на уровне респондентов рекомендуется отдельно определить классы взвешивания и обеспечить участие 8 000 респондентов (в применимых случаях 4 000 респондентов) в каждом регионе, по которому требуется получение оценок GATS. Кроме того, для корректировки следует использовать все следующие переменные классов взвешивания на уровне респондентов или их комбинацию: городская/сельская местность и возраст (15–24, 25–34, 35–44, 45–54, 55), пол (мужской/женский) и потребление табака на данное время (курит/ не курит) согласно данным из реестра респондентов домохозяйств. В случае если получение оценок по регионам не планируется, классы взвешивания для компонента на уровне респондентов могут

РЕКОМЕНДАЦИЯ.

Любой компонент корректировки с учетом отсутствия ответа на уровне домохозяйств или респондентов, превышающий 3,00, должен быть установлен как равный 3,00.

РЕКОМЕНДАЦИЯ.

При корректировке на долю не ответивших ячейки класса взвешивания для компонента на уровне респондентов следует сформировать по критериям пола, возраста и потребления табака согласно данным из реестра респондентов; а также по регионам, если необходимы качественные региональные оценки.

быть сформированы комбинацией следующих переменных: городская/сельская местность, пол, возраст и потребление табака.

Далее представлено вычисление двух мультипликативных составляющих корректировок по классу взвешивания для респондента R, которое проводится после определения классов взвешивания для всех компонентов. В приведенных ниже примерах числителями и знаменателями являются суммы базовых значений удельного веса. Применительно к компоненту на уровне домохозяйств, если все из $H_{\alpha^*} = 56$ выбранных домохозяйств в составе α -й ПЕО, в которой расположено домохозяйство респондента R, соответствуют критериям включения в обследование, и сумма базовых значений удельного веса этих домохозяйств составляет 170 013,3, при этом согласие на участие в обследовании было получено в $r_{\alpha^*}^{(HH)} = 50$ домохозяйствах из их числа (взвешенное общее значение = 156 261,9), тогда компонент на уровне домохозяйств для корректировки с учетом доли не ответивших будет вычисляться исходя из полученной опытным путем взвешенной доли ответивших в выборке домохозяйств, выбранной из ПЕО или сегмента, где расположено домохозяйство респондента R. Для этого примера предположим, что $\epsilon^{household-level}$ превышает 90%, что позволяет использовать приведенную выше формулу (20). Взвешенная доля ответивших на вопросник для домохозяйств в классе взвешивания для респондента R (проживающего в i -м домохозяйстве) составляет

$$RR_h^{(weighted,HH)} = \hat{\pi}_{hi}^{(wca,HH)} = \sum_{i=1}^{r_{\alpha^*}^{(HH)}} B_{hi} / \sum_{i=1}^{H_{\alpha^*}} B_{hi} = \frac{156\,261,9}{170\,013,3} = 0,9191. \quad (26)$$

Тогда компонент на уровне домохозяйств для корректировки по классу взвешивания для респондента R (проживающего в i -м домохозяйстве) будет вычисляться как

$$A_{hi}^{(nr,HH)} = \frac{1}{[\hat{\pi}_{hi}^{(wca,HH)}]} = \frac{1}{\sum_{i=1}^{r_{\alpha^*}^{(HH)}} B_{hi} / \sum_{i=1}^{H_{\alpha^*}} B_{hi}} = \frac{1}{0,9191} = 1,088, \quad (27)$$

где B_{hi} обозначает переменные базовых значений удельного веса домохозяйств в h -м классе взвешивания.

Если было установлено, что менее 90% домохозяйств страны, для которых известно соответствие критериям включения в обследование GATS, достоверно соответствуют этим критериям, то неизвестный компонент соответствия критериям обследования на уровне домохозяйств умножается на (ϵ), в результате чего получается несколько более высокая доля ответивших и, соответственно, более низкое значение корректировки на уровне домохозяйств.

Поскольку компонент корректировки на уровне домохозяйств (1,088) составляет менее 3,00, нет необходимости ограничивать значение корректировки.

Если процентная доля (ϵ) выбранных респондентов, соответствующих критериям для заполнения вопросника GATS, составляет 0,90 и выше, то для вычисления взвешенной доли ответивших на индивидуальный вопросник используется приведенная выше формула (23). Если (ϵ) составляет менее 0,90, то для вычисления доли ответивших на индивидуальный вопросник следует использовать формулу (25). Как и в случае компонента корректировки на уровне домохозяйств, компонент корректировки на уровне респондентов для респондента R равен единице, деленной на

взвешенную долю ответивших в классе взвешивания для респондента R. Теперь предположим, что респондент R — женщина, которая употребляет табак, проживает в городской местности региона X и является одним из $r_h = 680$ респондентов (взвешенное число = 119 009 025) среди $n_h = 771$ жителя домохозяйств (взвешенное значение = 134 935 233) в классе взвешивания на уровне респондентов, состоящего из респондентов женского пола, проживающих во всех выбранных домохозяйствах в городских районах региона X и на данный момент употребляющих табак. Также предположим, что взвешенная доля ответивших в классе взвешивания для респондента R составляет

$$RR_h^{(weighted, person)} = \hat{\pi}_{hj}^{(wca, person)} = \sum_{j=1}^{680} B_{hj} / \sum_{j=1}^{771} B_{hj} = \frac{119\,009\,025}{134\,935\,233} = 0,8820. \quad (28)$$

Тогда компонент корректировки на уровне респондентов для респондента R вычисляется как

$$A_{hj}^{(nr, person)} = \frac{1}{[\hat{\pi}_{hj}^{(wca, person)}]} = \frac{1}{\sum_{j=1}^{680} B_{hj} / \sum_{j=1}^{771} B_{hj}} = \frac{1}{0,8820} = 1,1338. \quad (29)$$

Компонент корректировки на уровне респондентов также не превышает 3,00, поэтому ограничения значения корректировки не требуется. Итоговая корректировка на долю не ответивших для респондента R составляет

$$A_j^{(nr)} = \frac{1}{\pi_j^{(wca)}} = \frac{1}{[\hat{\pi}_i^{(wca, HH)}][\hat{\pi}_j^{(wca, person)}]} = [A_i^{(nr, HH)}][A_j^{(nr, person)}] = [1,0880] \cdot [1,1338] = 1,2336. \quad (30)$$

С учетом величины базового значения удельного веса и итоговой корректировки на долю не ответивших для респондента R удельный вес респондента R, скорректированный на долю не ответивших, вычисляется как

$$W_j^{(nr)} = [B_j] \cdot [A_j^{(nr)}] = [12501,3081] \cdot [1,2336] = 15421,61. \quad (31)$$

3.4 Калибровка методом корректировки после стратификации

3.4.1 Исходные данные, требуемые для выполнения вычислительного этапа

- Частотное распределение населения по данным переписи, проведенной в течение пяти лет до обследования GATS, или другого авторитетного источника актуальных демографических данных, совместно по категориальным переменным, связанным с моделями потребления табака, и по сохраняющейся несбалансированности выборки (например, по охвату). К числу возможных переменных калибровки относятся пол, образование, возраст и городская/сельская местность, а также регион, если имеются данные о численности населения в регионах и размеры выборок респондентов на уровне регионов достаточны для получения качественных региональных оценок (рекомендуется 8 000 респондентов)
- Наличие в вопроснике GATS сопоставимых по формулировке вопросов, категорий ответов на вопросы в разбивке по полу респондентов, а также вопросов об образовании.
- Удельные веса, скорректированные на долю не ответивших ($W_j^{(nr)}$), для всех респондентов выборки.

Наглядный пример вычисления корректировки после стратификации

Описанные выше два типа корректировки на долю не ответивших являются эффективными методами устранения несбалансированности выборки, возникшей из-за тех или иных переменных, используемых для определения классов взвешивания и моделирования склонности к ответу. Однако выборка может обладать другими важными характеристиками, по которым корректировка не проводилась. Например, полученные в выборке различные уровни ответа могут быть обусловлены наличием дополнительных характеристик. Также могут существовать характеристики населения, по которым имеется различный охват структурой выборки, либо могут быть различия в размере выборки по иным характеристикам, помимо используемых в процессе отбора для выделения страт. Одним из общих решений проблемы сохранения несбалансированности выборки является проведение дополнительной калибровки, но на этот раз по группе населения, из которой была сформирована выборка.

Первыми, кто ввел в обиход понятие «калибровки» в увязке со взвешиванием выборки, были Девилл и Сарндал (Deville and Sarndal, 1992), однако методы, позволяющие эффективно ограничить свойства удельных весов, существуют уже более 60 лет. Как правило, цель калибровочной корректировки удельного веса заключается в приведении взвешенных сумм полученных в выборке значений в соответствие с аналогичными значениями в целевой группе населения. Первыми важными методами калибровки удельного веса стали корректировка после стратификации и балансировка, которые могут являться частными случаями применения общей системы калибровки, описанной Девиллом и Сарндалом (1992). Оба метода широко используются по сей день.

Задача калибровки зависит от того, применялись ли другие методы корректировки, и, если они применялись, то какие именно и в каком порядке. Например, если при использовании всего трех методов корректировки сначала проводятся корректировки с учетом доли не ответивших и затем калибровка, то калибровка позволяет устранить любую несбалансированность выборки, которая сохранилась после корректировок на долю не ответивших. С другой стороны, если практически применимым методом является только калибровка, она становится единственным средством устранения всех источников несбалансированности выборки.

Калибровка итогового набора удельных весов к распределению населения может проводиться на основе демографических данных, полученных из внешнего источника превосходящего статистического качества (например, последней переписи населения или результатов другого современного национального обследования с оценками численности населения равного или более высокого качества). В качестве источника для калибровки могут также использоваться надежные и общепринятые прогнозы численности населения. В случае если последняя перепись проводилась как минимум за пять лет до даты сбора данных для обследования GATS, следует рассмотреть вопрос об использовании других источников. Страны, не располагающие источниками данных для калибровки или располагающие устаревшими данными, могут быть не в состоянии выполнить этот этап корректировки. В этом случае странам следует проконсультироваться с национальными координаторами CDC и специалистами по статистике Координационного центра данных и Комитета по рассмотрению выборки. Данный этап в основном включает в себя корректировку взвешенной выборки (на основе $W_j^{(nr)}$ из выборки GATS) по отношению к распределению населения по набору категориальных переменных калибровки одним из двух способов: (1) путем *корректировки после*

стратификации (или взвешивания ячеек) по отношению к общему или перекрестному распределению населения относительно этих переменных либо (2) путем *балансировки* (или итерационного пропорционального подбора) для приведения предельных значений совокупного распределения этих переменных в выборке в соответствие с их распределением среди населения. Хотя вероятнее всего, что при использовании балансировки дисперсия итоговых скорректированных значений удельного веса будет несколько ниже, в рамках GATS предполагается формирование выборок относительно большего размера, которые лучше поддаются корректировке после стратификации, что позволит точнее откалибровать выборку по отношению к населению. Таким образом, рекомендуемым методом калибровки для GATS является корректировка после стратификации. Подробные инструкции по ее проведению излагаются ниже. Если национальные специалисты по статистике заинтересованы в использовании процедур балансировки, перед их применением следует проконсультироваться с национальным координатором CDC и Комитетом по рассмотрению выборки (КРВ).

Как и в случае других видов корректировки, наибольший эффект калибровки достигается в тех случаях, когда переменные, используемые для определения контрольных распределений, характеризуются высокой степенью корреляции с основными переменными обследования. Несмотря на то, что в обследованиях здоровья населения выбор наилучшего набора предикторов часто зависит от изучаемых переменных, сильными предикторами моделей потребления табака обычно являются факторы пола и образования; таким образом, они представляются оптимальным вариантом для выборки GATS. В тех случаях, когда размер выборки достаточен для получения оценок по регионам, в качестве переменных для калибровки также можно рассмотреть возраст, проживание в сельской/городской местности и регион. В тех случаях, когда считается, что сравнение городской и сельской местности имеет более важное значение, чем преимущества калибровки по критерию образования как предиктору потребления табака, следует заменить в корректирующих ячейках критерий образования на критерий проживания в городской/сельской местности. Наконец, используемый для анализа итоговый удельный вес (W_j) для записи данных по j -му респонденту рассчитывается по уравнению (1) как $W_j = V_j A_j^{(nr)} A_j^{(cov)} A_j^{(cal)}$, где $A_j^{(cal)}$ вычисляется с помощью той или иной стратегии калибровки. Следует отметить, что в отношении охвата структурой выборки не рекомендуется проводить никаких специальных корректировок, поэтому $A_j^{(cov)} = 1$.

РЕКОМЕНДАЦИЯ.

Ячейки для корректировки после стратификации следует сформировать по критериям пола, возраста и образования согласно сообщенной респондентами информации; а также по регионам, если требуются качественные региональные оценки. Если считается, что сравнение городской и сельской местности имеет более важное значение, чем преимущества калибровки по критерию образования как предиктору потребления табака, следует заменить в корректирующих ячейках критерий образования на критерий проживания в городской/сельской местности.

Корректировка после стратификация (КПС) должна осуществляться следующим образом. Во-первых, полный набор респондентов выборки, опрос которых в рамках обследования был

завершен, распределяется по корректирующим ячейкам калибровки. Эти ячейки выделяются путем многосторонней классификации нескольких категориальных переменных калибровки, для которых известна корреляция с основными показателями потребления табака в целевой группе населения (например, пол и образование либо возраст, пол и образование). Некоторые переменные калибровки являются более сильными предикторами, чем другие. При прогнозировании важных дихотомических (0/1) переменных потребления табака одним из поддающихся измерению показателей прогностической силы является уровень статистической значимости коэффициента логистической регрессии. Затем в каждой корректирующей ячейке рассчитывается корректирующий коэффициент калибровки, который вычисляется как отношение следующих двух значений. Числитель представляет собой подсчет численности целевой группы населения в ячейке согласно надежному источнику (например, недавней переписи населения). Знаменатель представляет собой сумму удельных весов, скорректированных на долю не ответивших, для всех респондентов в ячейке, которая представляет собой оценку численности целевой группы населения (числителя). Скорее всего, полученное соотношение составит 1,00. Значение КПС менее 1,00 указывает на то, что в данной ячейке чрезмерно представлена включенная в выборку группа населения, в то время как значение КПС более 1,00 указывает на то, что группа населения представлена в ячейке в недостаточной степени. Умножение КПС на удельный вес, скорректированный на долю не ответивших, позволяет устранить любую чрезмерную или недостаточную репрезентативность членов группы населения в ячейке таким образом, чтобы итоговое взвешенное распределение выборки по наборам корректирующих ячеек точно соответствовало распределению населения по ячейкам. Таким образом, будет проведена калибровка итоговой взвешенной выборки респондентов к целевой группе населения.

Как указывалось ранее, для калибровки выборки в обследовании GATS рекомендуется использовать как минимум следующие предикторные переменные: пол (мужской или женский) и четыре категории уровня законченного формального образования. Категории образования должны быть определены таким образом, чтобы предельное процентное распределение населения по этим категориям было приближено к равномерному (т. е. в каждой группе должно содержаться приблизительно 25% населения по данным переписи или другого статистически превосходящего внешнего источника).

В некоторых случаях число респондентов в корректирующих ячейках снижается ниже определенного минимального стандарта (например, менее 50 респондентов). Во избежание появления в отдельных ячейках экстремальных значений корректировки после стратификации из-за небольшого числа респондентов выборки, рекомендуется рассмотреть объединение или слияние каждой небольшой ячейки выборки с одной или несколькими ячейками. Этот процесс слияния называется «схлопыванием ячеек». В этой связи необходимо определить, с какой(ими) ячейкой(ами) необходимо объединить ячейку небольшого размера. Поскольку принцип корректировки после стратификации связан со стратификацией выборки, Киш (Kish, 1965) рекомендует (как и в случае стратификации) объединить схожие ячейки (то есть ячейки, для которых имеются сопоставимые средства подготовки основных показателей потреблением табака для обследования GATS). Это приводит к объединению категорий переменных калибровки, обладающих наименьшей прогностической силой (т. е. имеющих наименьшие коэффициенты логистической регрессии при прогнозировании важных дихотомических (0/1) переменных потребления табака). Возьмем, к примеру, ячейку небольшого размера, включающую респондентов мужского пола наиболее старшей возрастной группы с наиболее низким уровнем

формального образования; поскольку возраст достоверно является наиболее слабой предикторной переменной калибровки, следует провести совмещение этой ячейки с ячейкой, включающей респондентов мужского пола следующей по убыванию старшей возрастной группы с наиболее низким уровнем формального образования так, чтобы в этих ячейках было получено минимальное число респондентов. Если требуется провести совмещение ячейки небольшого размера с двумя ячейками или более, можно либо перейти в следующую ячейку выше или ниже порядком, в которой имеется переменная с наименьшей прогностической силой, либо рассмотреть возможность совмещение по второй по прогностической силе переменной калибровки, как описано выше. Конечная цель заключается в формировании ячеек, в которых число респондентов превышает стандарт в отношении минимального размера выборки.

Возвращаясь к примеру вычисления удельного веса для вымышленного респондента R в обследовании GATS, респондент R будет отнесен к корректирующей ячейке калибровки, включающей респондентов женского пола, проживающих в регионе X и имеющих аналогичный с респондентом уровень образования. Если по данным последней переписи число членов населения с такими же характеристиками составляет $N_h = 2\,724\,182$ и взвешенная сумма выборки с такими же характеристиками составляет $\sum_{j=1}^{r_h} W_{hj}^{(nc)} = 2\,919\,669,3442$, тогда корректировка после стратификации ($A_j^{(cal)}$) для респондента R и всех других респондентов GATS в корректирующей ячейке респондента R вычисляется как

$$A_j^{(cal)} = \frac{N_h}{\sum_{j=1}^{r_h} W_{hj}^{(nc)}} = \frac{2\,724\,182}{2\,919\,669,3442} = 0,9330, \quad (32)$$

где N_h — численность населения согласно внешнему источнику данных о численности населения, и, таким образом, $N = \sum_{h=1}^H N_h$ — общая численность населения согласно внешнему источнику данных, используемому для калибровки, а r_h — размер выборки в корректирующей ячейке. Следует отметить, что $A_j^{(cal)}$ одинаков для всех респондентов в корректирующей ячейке h . При проверке набора корректировок ($A_j^{(cal)}$) по всем корректирующим ячейкам по завершении данного этапа значения большинства из них должны быть либо несколько больше, либо несколько меньше единицы, причем ячейки, в которых после корректировки с учетом доли не ответивших сохраняется недостаточная представленность сегментов населения выборки, имеют значения больше единицы, а те, в которых чрезмерно представлены сегменты населения, имеют значения меньше единицы.

При умножении полученного удельного веса респондента R на значение корректировки для корректирующей ячейки респондента R ($A_j^{(cal)}$) итоговый скорректированный удельный вес респондента R (который следует сохранить в записи данных по респонденту R в файле для анализа) будет составлять

$$W_j = W_j^{(nr)} \cdot A_j^{(cal)} = [15421,61] \cdot [0,9339] = 14402,24, \quad (33)$$

и взвешенное относительное распределение численности выборки с использованием этого итогового удельного веса будет совпадать с соответствующим распределением численности населения относительно переменных калибровки; т. е.

$$\sum_{j=1}^{r_h} W_{hj} / \sum_{h=1}^H \sum_{j=1}^{r_h} W_{hj} = N_h / N. \quad (34)$$

Это уравнение можно использовать для проверки вычислений, проведенных в рамках данной корректировки.

Альтернативный подход

При небольших размерах выборки респондентов и более выраженном непропорциональном отборе по регионам может быть целесообразно использовать балансировку вместо корректировки после стратификации. Балансировка для приведения сумм итоговых скорректированных значений удельного веса в соответствие с общими предельными категориальными величинами по набору переменных калибровки (но не по межклассовым ячейкам) была впервые предложена Демингом и Стефаном (Deming and Stephan, 1940). Сначала проводится тот же тип многосторонней классификации категориальных переменных калибровки, что и при корректировке после стратификации; на каждом повторе балансировки проводится приведение общих категориальных сумм по отдельной переменной в соответствие с общими категориальными суммами среди населения (см. пример вычисления при калибровке по двум переменным: Kalton and Flores-Cervantes, 2003). Страны, заинтересованные в использовании балансировки для калибровки итоговых удельных весов в обследовании GATS, должны обсудить этот вариант с национальным координатором CDC и Комитетом по рассмотрению выборки.

4. Обеспечение качества удельного веса элементов выборки GATS

В заключительной части настоящего руководства представлен ряд предложений, направленных на обеспечение последовательного и точного вычисления всеми странами удельного веса элементов выборки GATS.

4.1 Тщательное документирование этапов выборки и соответствующих вероятностей отбора

Невозможно переоценить важность тщательного документирования процесса отбора, используемого на каждом этапе выборки в рамках GATS. Поэтому крайне важно, чтобы технические сотрудники каждой страны, проводящей обследование GATS, взяли на себя строгие обязательства по ведению данной документации. Существует несколько способов обеспечения полноты этой документации и ее полного соответствия задаче вычисления удельных весов. Во-первых, после выбора алгоритма отбора для каждого этапа сотрудникам следует обратиться к надежному источнику по составлению выборки, чтобы найти правильную вычислительную формулу для используемой стратегии отбора. К числу некоторых широко используемых в обследованиях стратегий отбора относятся простой случайный отбор, систематический отбор, случайный отбор пошаговым методом, систематический отбор с ВПО, отбор с ВПО без возвращения и отбор с ВПО с возвращением. В качестве справочных материалов можно ознакомиться с работами следующих авторов: Hansen, Hurwitz, and Madow (1953), Kish (1965), Raj (1968), Cochran (1977), Sukhatme, et al. (1984) и Lohr (1999). Во-вторых, для обеспечения корректности вычислений в процессе отбора необходимо предусмотреть наличие средств для вычисления вероятностей отбора. В практическом смысле это означает, что предпочтение отдается компьютеризированному отбору по сравнению с методом ручного отбора и что в программном коде для построения выборки также должны быть предусмотрены шаги по вычислению и сохранению вероятностей отбора одновременно с проведением отбора отдельных членов выборки. Это также означает, что, если некоторые этапы выборки должны выполняться вручную, то формы, используемые для составления выборки, должны обеспечивать возможность получения всей информации, необходимой для расчета вероятности отбора на том или ином этапе одновременно с проведением данного этапа отбора или же позднее, если такой вариант более предпочтителен. В-третьих, на технического сотрудника страны должна быть возложена общая ответственность за обеспечение правильного и своевременного вычисления вероятностей отбора и их надлежащего хранения до этапа определения удельных весов. В-четвертых, процедуры и формулы, используемые для вычисления всех компонентов удельного веса элементов выборки, должны быть разработаны и подробно задокументированы заблаговременно до начала вычисления удельных весов, с тем чтобы другой компетентный технический сотрудник, не задействованный в обследовании GATS, мог правильно произвести указанные расчеты. По завершении процесса вычисления каждая страна должна внести соответствующие изменения в разработанную на этапе планирования техническую документацию с тем, чтобы она могла использоваться в качестве официального документа по вычислению удельного веса элементов выборки GATS. И наконец, должен быть обеспечен общий доступ к документации по процедуре определения удельных весов для партнеров GATS и Комитета GATS по рассмотрению выборки.

4.2 Контролирование отрицательного статистического эффекта различных значений удельного веса

Как показано в работе Киша (Kish (1965)), дисперсия значений удельного веса элементов выборки может привести к увеличению дисперсии оценок обследования на коэффициент мультипликативного эффекта, обозначаемый как $Meff$. Например, если $\bar{w} = 12394,0091$ соответствует средней величине итоговых скорректированных значений удельного веса W_j элементов выборки, и $s_{\bar{w}}^2 = (5463,3760)^2$ соответствует дисперсии этих же значений удельного веса для членов выборки GATS, частью которой является вымышленный респондент R, тогда

$$Meff_w = 1 + \frac{s_{\bar{w}}^2}{\bar{w}^2} = 1 + \left(\frac{5463,3760}{12394,0091} \right)^2 = 1,1943. \quad (35)$$

Некоторые факторы, способствующие появлению этой дисперсии, объясняются характеристиками плана выборки, предназначенными для уменьшения дисперсии оценок (например, стратификация выборки), в то время как другие факторы связаны со стратегиями, применяемыми для ограничения величины смещения этих оценок (например, корректировка с учетом доли не ответивших и калибровка выборки). Для эффективного контролирования этой дисперсии важно найти баланс между снижением отрицательного воздействия $Meff_w$ на точность оценок и сохранением преимуществ уменьшения величины смещения, полученного при корректировке удельных весов.

Рекомендуемый подход к контролированию эффекта различных значений удельного веса в выборке GATS заключается в отслеживании $Meff_w$ в итоговых скорректированных значениях удельного веса (W_j) и в обращении за консультацией к сотрудникам центрального органа GATS при выявлении $Meff_w > 2,00$. По итогам такой консультации могут быть выбраны различные имеющиеся стратегии (например, усечение удельных весов), позволяющие ограничить размер корректировки удельных весов, либо различные стратегии «урезания» выпадающих значений удельного веса. Например, в работе Калтона и Флорес-Сервантеса (Kalton and Flores-Cervantes, 2003) описывается процедура

дополнительного совмещение корректирующих ячеек, применяемая для ограничения величины корректировок, производимых по итоговому набору ячеек. Этот упреждающий подход может использоваться не только для ограничения размера корректировки с учетом доли не ответивших, но и в качестве компромиссного решения между корректировкой после стратификации и балансировкой, применяемых для ограничения размера калибровочной корректировки (см. Deville and Sarndal, 1992). С другой стороны, Поттер (Potter, 1988, 1990; 1993) описывает несколько подходов к подравнению выборки, которые заключаются в создании нескольких наборов уменьшенных удельных весов по различным уровням уменьшения и выборе одного набора, характеризующегося минимальной предполагаемой среднеквадратичной ошибкой в основных оценках обследования. Недостаток данных подходов заключается в том, что уменьшение удельных весов, позволяющее получить наименьшую среднеквадратическую ошибку для одной оценки обследования, может не дать аналогичный результат для другой оценки.

РЕКОМЕНДАЦИЯ.
Следует контролировать эффект различных значений удельного веса путем отслеживания его мультипликативного эффекта на дисперсию оценок обследования и при необходимости обращаться за консультацией к специалистам GATS.

4.3 Тщательное соблюдение рекомендуемых в настоящем руководстве вычислительных процедур и процедур проверки качества

Цель настоящего руководства заключается в предоставлении рекомендаций по выработке стратегии вычисления удельных весов для каждой выборки GATS, успешному осуществлению данной стратегии, и, в конечном итоге, обеспечении того, чтобы удельные веса, используемые при анализе данных GATS, способствовали получению наиболее точных демографических оценок в отношении потребления табака. Для содействия достижению этой цели предусмотрен ряд конкретных процедур, которые позволяют обеспечить соответствие удельных весов для данных GATS самым высоким стандартам в области обследований и которые, в случае их единообразного применения всеми странами, проводящими обследование GATS, будут способствовать качественному сравнению итоговых результатов обследования в различных странах. Способы обеспечения качества удельного веса элементов выборки также предлагались или подразумевались при описании этапов их вычисления. Эти способы излагаются ниже.

РЕКОМЕНДАЦИЯ.
Необходимо следовать рекомендованным этапам вычисления удельного веса элементов выборки и выполнять предлагаемые проверки в ходе этих вычислений.

1. Привлечь специалистов, не участвующих в вычислении удельных весов для обследования GATS, но знакомых с содержанием изложенного в настоящем руководстве рекомендуемого протокола, к проверке вычислений удельных весов в процессе этих вычислений. Эту проверку лучше всего проводить не по завершении всего вычислительного процесса, а после каждого этапа (например, после вычисления базовых значений удельных весов, корректировки на долю не ответивших и калибровки).
2. Определить, является ли средний размер базовых значений удельного веса (B_j), деленный на средний размер значений удельного веса, скорректированных на долю не ответивших ($W_j^{(nr)}$), приблизительно равным итоговой общей доле ответивших RR1, как описано в разделе 3.2 настоящего руководства.
3. Использовать уравнение (34) для проверки (по каждой корректирующей ячейке калибровки) соответствия взвешенного распределения численности выборки с использованием итоговых удельных весов (W_j) после калибровки распределению населения, по отношению к которому проводилась такая калибровка.
4. Убедиться, что сумма итоговых скорректированных значений удельного веса (W_j) по всей выборке совпадает по размеру с общей численностью населения (лиц в возрасте 15 лет и старше), указанной в используемом для калибровки источнике. Если таким источником является последняя перепись населения, тогда сумма удельных весов должна совпадать с указанной в ней численностью населения.
5. Убедиться, что в большинстве случаев значение корректировки после стратификации ($A_j^{(cal)}$), производимой по набору корректирующих ячеек, несколько больше или меньше единицы.

4.4 Подготовка подробного документа с описанием фактической процедуры вычисления удельных весов

Как и в случае большинства обследований, проведение которых предусматривает разработку предварительного плана, зачастую фактические процедуры расчетов несколько отличаются от запланированных (т. е. по сути процедур, представленных в настоящем руководстве, если применяется рекомендуемый процесс вычисления удельных весов). При определении удельного веса элементов выборки отклонения от запланированных требований могут возникать на каждом из трех

вычислительных этапов. Например, в случае использования для обследования другого надежного метода отбора респондентов из домохозяйств, условная вероятность отбора респондента из включенного в обследование GATS домохозяйства будет вычисляться в соответствии с методом отбора с возвращением. Во избежание снижения степени детализации этого важного процесса, предшествующего аналитической работе в рамках GATS, сразу же после вычисления удельных весов следует подробно

задокументировать фактические процедуры их расчета в соответствии с основными положениями настоящего руководства, а затем обеспечить интеграцию этого документа с остальной официальной документацией по всем видам деятельности, проводимым в рамках обследования GATS. Этот документ будет использоваться специалистами по медико-биологической статистике Комитета по рассмотрению выборки для проведения формальной оценки процедур вычисления и калибровки удельного веса элементов выборки и впоследствии будет включен в итоговый страновой отчет в качестве официальной документации о методах, используемых в GATS. Подробные сведения о требуемых для этой проверки формате и вычислениях содержится в документе «Обеспечение качества GATS: руководства и документация». Полученные странами итоговые значения удельного веса не должны быть опубликованы до рассмотрения и официального утверждения этой документации как соответствующей установленным в настоящем руководстве стандартам в отношении качества удельных весов.

РЕКОМЕНДАЦИЯ.

Каждая страна, проводящая обследование GATS, должна подготовить письменный документ с подробным описанием процесса вычисления удельных весов.

4.5 Консультирование с сотрудниками координаторами GATS при вычислении удельных весов

Наконец, координаторы GATS готовы оказать каждой стране, проводящей обследование GATS, содействие в вычислении удельного веса элементов выборки. Это содействие будет включать разъяснение процессуальных методов, изложенных в настоящем руководстве, а также помощь в устранении ошибок, возникающих при выполнении этого важного комплекса задач обследования в силу непредвиденных обстоятельств.

5. Библиография

- American Association for Public Opinion Research. 2009. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 6th edition. AAPOR.
- Berry, C. C., Flatt, S. W., and Pierce, J. P. (1996). Correcting Unit Nonresponse via Response Modeling and Raking in the California Tobacco Survey. *Journal of Official Statistics*, 12(4), 349-363.
- Cochran, W.G. (1977). *Sampling Techniques*, 3rd Ed., New York: Wiley and Sons.
- Deming, W. E. and Stephan, F. F. (1940). On the Least Squares Adjustment of a Sample Frequency Table When the Expected Marginal Totals are Known. *Annals of Mathematical Statistics*, 11(4), 427-444.
- Deville, J. C. and Sarndal, C. E. (1992). Calibration Estimators in Survey Sampling. *Journal of American Statistics Association*, 87, 376-382.
- Сотрудничающая группа по проведению Глобального обследования потребления табака взрослым населением. Глобальный опрос взрослого населения о потреблении табака (GATS). Рекомендации по проведению инструктажа в сфере работы с данными. Атланта, Джорджия: Центры США по контролю и профилактике заболеваний, 2020 г.
- Сотрудничающая группа по проведению Глобального обследования потребления табака взрослым населением. Глобальный опрос взрослого населения о потреблении табака (GATS). Руководство для выездного интервьюера. Атланта, Джорджия: Центры США по контролю и профилактике заболеваний, 2020 г.
- Сотрудничающая группа по проведению Глобального обследования потребления табака взрослым населением. Глобальный опрос взрослого населения о потреблении табака (GATS). Руководство для полевого супервайзера. Атланта, Джорджия: Центры США по контролю и профилактике заболеваний, 2020 г.
- 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 г.
- Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). *Sample Survey Methods and Theory*, Vols. I and II, 1953.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.
- Iannacchione, V. G., Milne, J. G., and Folsom, R. E. (1991). Response Probability Weight Adjustments Using Logistic Regression. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 637-642.
- Kalsbeek, W. D., Morris, C., and Vaughn, B. (2001). Effects of Nonresponse on the Mean Squared Error of Estimates from a Longitudinal Study. *Proceedings of the Section on Survey Research Methods, American Statistical Association*
- Kalton, G. (1983). *Compensating for Missing Survey Data*. Ann Arbor, MI: University of Michigan.
- Kalton, G. and Flores-Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics*, 19(2), 81-97.
- Kish, L. (1965). *Survey Sampling*, Second Printing, New York: Wiley and Sons.
- Lepkowski, J. M., Kalton, G., and Kasprzyk, D. (1989). Weighting Adjustments for Partial Nonresponse in the 1984 SIPP Panel. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 296-301.
- Lessler, J. T. and Kalsbeek, W. D. (1992). *Nonsampling Error in Surveys*. New York: Wiley and Sons.
- Lohr, S. (1999). *Sampling: Design and Analysis*, Pacific Grove: Duxbury Press. 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.
- Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 225-230.
- Potter, F. J. (1993). The Effect of Weight Trimming on Nonlinear Survey Estimates. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 758-763.
- Raj, D. (1968). *Sampling Theory*, New York: McGraw-Hill.
- Särndal, C.-E., Swensson, B., and Wretman, J. H. a. (1992). *Model assisted survey sampling*. New York: Springer-Verlag.
- Sukhatme, P. V. et al, (1984). *Sampling Theory of Surveys with Applications*, Ames: Iowa State University Press.

Global Adult Tobacco Survey (GATS)