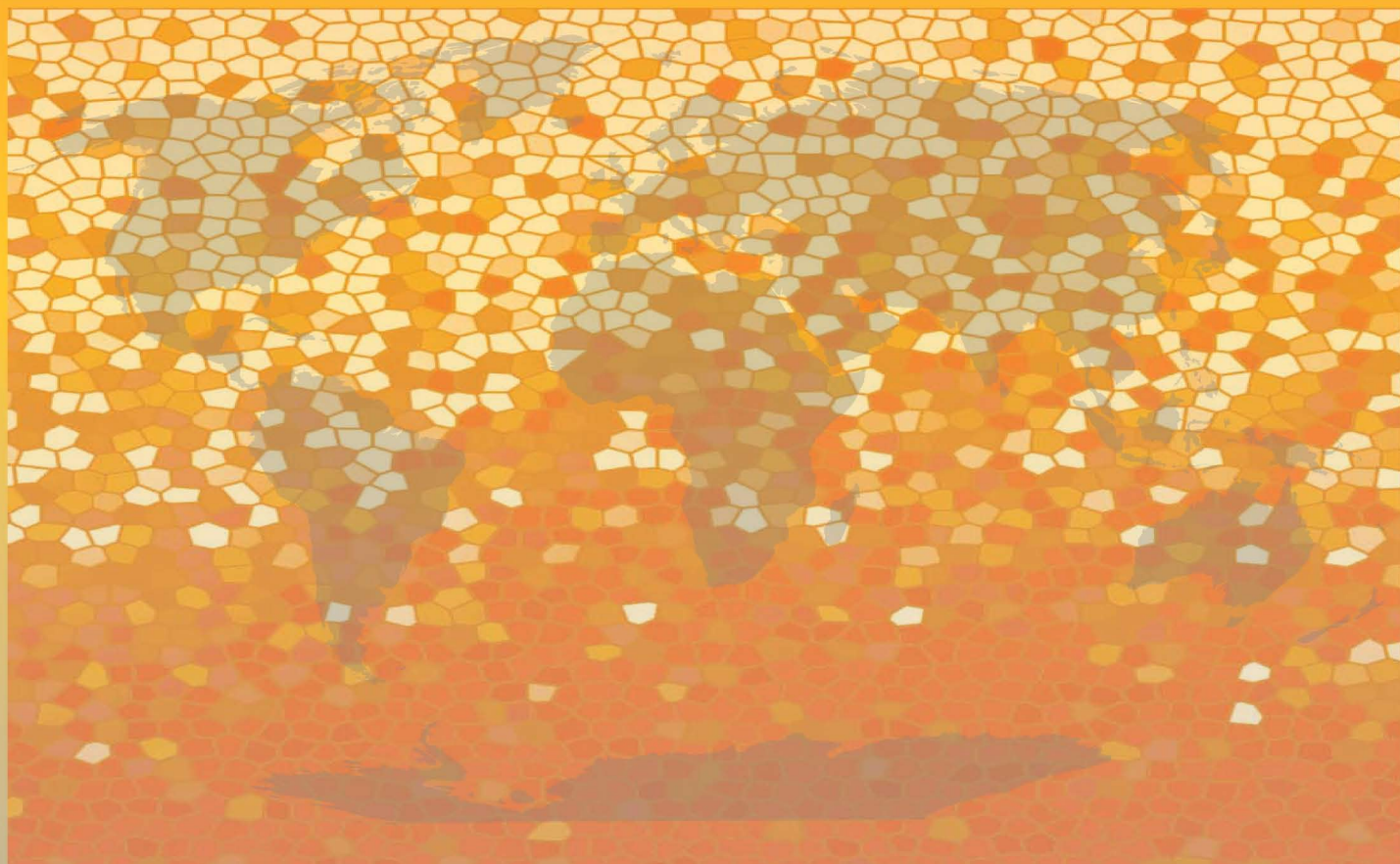


Руководство

ПО РЕДАКТИРОВАНИЮ ДАННЫХ

ПЕРЕПИСИ НАСЕЛЕНИЯ И ЖИЛОГО ФОНДА

ПЕРВОЕ ПЕРЕСМОТРЕННОЕ ИЗДАНИЕ



Организация
Объединенных Наций

Департамент по экономическим и социальным вопросам
Статистический отдел

Методологические исследования

Серия F № 82/Rev. 1

**Руководство
по редактированию
данных переписи населения
и жилого фонда**

Первое пересмотренное издание



Организация Объединенных Наций
Нью-Йорк, 2011 год

Департамент по экономическим и социальным вопросам

Департамент по экономическим и социальным вопросам Секретариата Организации Объединенных Наций играет исключительно важную роль по согласованию глобальной экономической, социальной и экологической политики и деятельности на национальном уровне. Работа Департамента осуществляется в трех основных взаимосвязанных областях: i) сбор, генерирование и анализ разнообразных экономических, социальных и экологических данных и информации, которые используются государствами — членами Организации Объединенных Наций для изучения общих проблем и оценки стратегий; ii) содействие проводимым государствами-членами в целом ряде межправительственных органов переговорам с целью выработки общих подходов к решению возникающих новых проблем; и iii) предоставление заинтересованным правительствам рекомендаций относительно путей и средств воплощения основ политики, выработанных на конференциях и встречах на высшем уровне Организации Объединенных Наций, в страновые программы и, по линии технической помощи, оказание содействия укреплению национального потенциала.

Примечание

Используемые в настоящем издании обозначения и представление материала не подразумевают выражения со стороны Секретариата Организации Объединенных Наций какого бы то ни было мнения в отношении правового статуса той или иной страны, территории, города или района или их полномочных органов, или в отношении делимитации их границ.

Термин «страна», используемый в настоящей публикации, также относится к территориям или районам в зависимости от контекста.

Обозначения «развитые» и «развивающиеся» страны или районы, а также «более развитые», «менее развитые» и «наименее развитые» регионы предназначены для удобства статистической работы и не означают какой-либо оценки этапа в процессе развития, достигнутого какой-либо определенной страной или районом.

Условные обозначения документов Организации Объединенных Наций состоят из прописных букв и цифр. Когда такое обозначение встречается в тексте, оно служит указанием на соответствующий документ Организации Объединенных Наций.

ST/ESA/STAT/SER.F/82/Rev.1

ИЗДАНИЕ ОРГАНИЗАЦИИ ОБЪЕДИНЕННЫХ НАЦИЙ
В продаже под №. R.09.XVII.11

ISBN: 978-92-1-461029-8

Авторское право © Организация Объединенных Наций, 2011 год
Все права защищены

Отпечатано в Организации Объединенных Наций, Нью-Йорк

Предисловие

В течение ряда лет Организация Объединенных Наций выпускает серии руководств и технических докладов, предназначенных для оказания помощи странам в планировании и проведении эффективных с точки зрения затрат переписей населения и жилого фонда по усовершенствованной методике. Эти руководства и доклады время от времени пересматриваются и обновляются, с целью отражения новых тенденций и учета возникающих при проведении переписей проблем. Настоящее издание входит в серию руководств, разработанных с целью оказания помощи странам в подготовке переписей 2000 года и последующих раундов переписей. Данная серия включает также:

- a) *Руководство по управлению переписями населения и жилого фонда*, Серия F, № 83 (издание Организации Объединенных Наций, в продаже под № R.00.XVII.15);
- b) *Руководство по географическим информационным системам и цифровому картированию*, Серия F, № 79 (издание Организации Объединенных Наций, в продаже под № R.00.XVII.12);
- c) *Guide for the Collection of Economic Characteristics in Population Censuses* [«Справочник по сбору данных в области экономических характеристик населения при переписях» (готовится к изданию)].

В публикации *Принципы и рекомендации в отношении переписей населения и жилого фонда, второе пересмотренное издание* (Организация Объединенных Наций, 2008 год), рассматриваются достоинства системы контроля и повышения качества на первоначальном этапе переписи, который является крайне важным с точки зрения успеха всей работы по проведению переписи, и значимость плана редактирования, который должен разрабатываться как часть всей программы проведения переписи и объединяться с другими планами и процедурами переписи. Для пользователей *Руководства по редактированию данных переписи населения и жилого фонда* будут особенно полезны *Принципы и рекомендации*, где содержится значительная доля исходной информации относительно процедур редактирования, изложенных в главах III, IV и V.

Цель настоящего *Руководства* заключается в том, чтобы дать странам широкое представление о методологии редактирования данных переписей и обследований и предоставить информацию соответствующим должностным лицам о применении различных подходов к редактированию данных переписи. В частности, целью публикации является обеспечение странам руководства по сохранению накопленного ими опыта редактирования и по активизации связей между специалистами, занимающимися существом содержания переписи, и специалистами по обработке данных, а также по составлению документации в ходе переписи или обследования, с тем чтобы избежать дублирования в работе во время следующей переписи или обследования.

В *Руководстве* рассматриваются преимущества и недостатки ручного и автоматизированного редактирования. В больших переписях корректирование

вручную почти невозможно с экономической точки зрения. Условия для такого корректирования, как правило, указываются в специально разработанных компьютерных программах, с помощью которых проводится автоматический поиск ошибок и вменение значений на основе других сведений о данном лице или домохозяйстве, или о других лицах или домохозяйствах. В основном *Руководство* посвящено автоматическому исправлению данных.

Редактирование с помощью компьютера играет важную роль в обнаружении и исправлении ошибок. На этапе компьютерного редактирования подробная проверка на согласованность может проводиться с помощью консультаций со специалистами, занимающимися существом содержания переписи. Обнаруженные ошибки можно исправить либо путем ссылки на первоначальные вопросники, либо автоматически. Хотя автоматическое редактирование ускоряет обработку данных, оно требует тщательного контроля качества поступающих данных.

Настоящее *Руководство* состоит из введения и пяти глав. Во введении описывается процесс проведения переписи и различные виды ошибок, которые при этом выявляются. Глава I посвящена основам редактирования данных переписи. В главах II–V говорится о процедурах и способах редактирования данных переписи на различных этапах обработки данных. Рекомендации технического порядка, в частности те, которые касаются программирования, изложены в приложениях.

Хотя настоящее *Руководство* посвящено главным образом редактированию данных переписей населения и жилого фонда, многие понятия и методы применимы также к проведению обследований.

Особая признательность выражается г-ну Майклу Дж. Левину из Гарвардского центра исследований народонаселения и развития за вклад в подготовку проекта данного *Руководства*. Благодарность выражается также Майклу Банкиеру, Уэсли Бенджамину, Марселю Бюро, Сильвейну Делислу, Дэррилу Джейнсу и Шону Кроу из Статистического управления Канады, которые рецензировали проект и представили ценные замечания в связи с подготовкой окончательной версии настоящей публикации.

Содержание

	Стр.
Предисловие	iii
Введение	1
A. Цель <i>Руководства</i>	1
B. Процесс переписи	2
C. Ошибки в процессе переписи	3
1. Ошибки в охвате	3
2. Ошибки в содержании	3
a) Ошибки в разработке переписного листа	4
b) Ошибки счетчиков	4
c) Ошибки респондентов	4
d) Ошибки кодирования	5
e) Ошибки при вводе данных	5
f) Ошибки при компьютерном редактировании	5
g) Ошибки при составлении таблиц	6
D. Структура <i>Руководства</i>	7
<i>Глава</i>	
I. Редактирование при проведении переписей и обследований	9
A. Редактирование: историческая справка	9
B. Группа редактирования	10
C. Практика редактирования: отредактированные и неотредактированные данные	12
D. Основы редактирования	15
1. Каким образом чрезмерное редактирование оказывает пагубное воздействие	18
a) Своевременность	18
b) Финансы	18
c) Искажение истинных значений	18
d) Ложное чувство уверенности	19
2. Обработка неизвестных величин	19
3. Кажущиеся изменения	20
4. Определение уровня допустимых отклонений	20
5. Извлечение уроков из процесса редактирования	21
6. Гарантия качества	21
7. Затраты на редактирование	21
8. Вменение	22
9. Архивирование	22

<i>Глава</i>	<i>Стр.</i>
II. Прикладные программы редактирования	25
A. Вопросы кодирования	27
B. Ручная и автоматизированная коррекция данных	32
C. Руководящие принципы коррекции данных	35
D. Проверка на достоверность и согласованность	38
1. Нисходящий принцип редактирования	39
2. Редактирование по принципу групповых переменных	40
E. Методы коррекции и вменения данных	43
1. Метод статического вменения («cold deck»)	44
2. Метод динамического вменения («hot deck»)	44
3. Вопросы динамического вменения («hot deck»)	48
a) Географические соображения	48
b) Использование смежных элементов данных	49
c) Как порядок расположения переменных влияет на матрицы	49
d) Сложность матриц вменения	50
e) Разработка матрицы вменения	50
f) Стандартизованные матрицы вменения	51
g) Когда динамическое вменение не используется	54
h) Какого размера должны быть матрицы вменения?	56
i) Проблемы, возникающие в случае слишком большого размера матрицы вменения	56
ii) Осмысление принципа функционирования матрицы вменения	57
iii) Проблемы, возникающие в случае слишком маленького размера матрицы вменения	57
iv) Элементы данных, представляющие трудности для матрицы вменения	57
4. Проверка матриц вменения	58
a) Создание исходной статической матрицы	58
b) Сообщения об ошибках	58
c) Заказная распечатка ошибок	59
d) Сколько раз следует прогонять программу редактирования?	61
5. Флажки вменения	62
F. Другие системы редактирования	64
III. Редактирование структуры данных	67
A. Редактирование географических параметров	68
1. Местоположение жилых помещений (населенный пункт)	68
2. Постоянное проживание в городских и сельских районах	68

	<i>Стр.</i>
В. Проверка охвата	69
1. Регистрация де-факто и де-юре	69
2. Иерархия домохозяйств и жилищных единиц	70
3. Фрагменты переписных листов	70
С. Структура записей о жилом фонде	71
Д. Соответствие между записями о жилом фонде и населении	71
1. Свободные и занятые жилища	71
a) Оставить жилищную единицу в категории «свободная»	72
b) Повторное неоднократное посещение жилищной единицы для заполнения переписных листов	72
c) Замена данных об отсутствующих лицах данными из другой жилищной единицы	72
2. Двойной учет домохозяйств и жилищных единиц	72
3. Недостающие домохозяйства и жилищные единицы	73
4. Соответствие между числом жильцов и суммарным числом жильцов	73
a) Когда число жильцов больше суммарного числа жильцов	73
b) Проверка числа лиц по признаку пола	74
c) Порядковая нумерация	74
5. Соответствие между жильцами и типом строения/ домохозяйства	75
Е. Двойные записи	75
Ф. Специальные совокупности	76
1. Коллективно проживающие лица	76
a) Когда для коллективов используется другой тип записи	76
b) Когда переменная отличает коллективы от других записей	76
c) Когда отсутствует код «тип коллектива»	77
d) Когда есть код коллектива, но все лица находятся в родственных отношениях	77
e) Проведение различий между разными типами коллективов	77
2. Сложные для регистрации группы населения	77
a) Сезонные мигранты	77
b) Бездомные лица	78
c) Кочевники и лица, проживающие в труднодоступных местах	78
d) Гражданские постоянно проживающие лица, временно отсутствующие в стране	78
e) Гражданские иностранцы, не пересекающие границу ежедневно и находящиеся в стране временно, включая лиц без документов или лиц, временно находящихся на момент переписи на судах в порту	78

	<i>Стр.</i>
<i>f)</i> Беженцы	79
<i>g)</i> Военнослужащие армии и флота, дипломатические сотрудники и их семьи, находящиеся за пределами страны, и иностранные военнослужащие армии и флота, дипломатические сотрудники и их семьи, находящиеся в стране.	79
<i>h)</i> Гражданские иностранцы, пересекающие границу ежедневно для работы в стране	80
<i>i)</i> Гражданские постоянно проживающие лица, пересекающие границу ежедневно для работы в другой стране.	80
<i>j)</i> Моряки торгового и рыболовного флота, постоянно проживающие в стране, но находящиеся в открытом море на момент переписи (включая лиц, не имеющих иного места жительства помимо своих помещений на борту судна).	80
G. Определение главы домохозяйства и супруги(а)	80
1. Редактирование переменной, касающейся главы домохозяйства	80
<i>a)</i> Порядок родственных связей	81
<i>b)</i> Когда глава не является первым лицом	81
<i>i)</i> Установление указателя для записи о главе домохозяйства	82
<i>ii)</i> Как сделать главой первое лицо	83
<i>iii)</i> Изменение кодов родства, чтобы первое лицо сделать главой домохозяйства.	83
<i>c)</i> Более одного главы	83
<i>d)</i> Если глава отсутствует.	84
2. Редактирование данных о супруге	84
<i>a)</i> Когда присутствует точно один супруг в моногамном обществе	84
<i>b)</i> Когда обнаруживается более одного супруга в моногамном обществе	84
<i>c)</i> Супруги в полигамном обществе	85
<i>d)</i> Прочие характеристики глав домохозяйств и их супругов	85
H. Возраст и дата рождения	85
1. Когда имеется дата рождения, но не указан возраст	85
2. Когда возраст и дата рождения не согласуются между собой	85
I. Учет недостоверных данных	86
<i>Глава</i>	
IV. Редактирование элементов данных о населении.	87
A. Демографические характеристики	89
1. Степень родства	89
<i>a)</i> Редактирование данных о степени родства	89

	<i>Стр.</i>
b) Когда главу домохозяйства необходимо указывать первым	90
c) Когда степень родства кодируется в обратном порядке ..	90
d) Когда имеются супруги в полигамном браке	90
e) Наличие нескольких родителей	91
f) Когда при переписи собираются данные о родстве по признаку пола	91
g) Когда данные о родстве и семейном положении не согласуются между собой	91
2. Пол	92
a) Когда код пола достоверен, но глава домохозяйства и супруг(а) одного пола	92
b) Когда в отношении мужчины имеется информация о рождаемости, а в отношении взрослой женщины нет ..	92
c) Когда код пола является недостоверным и имеется супруг(а)	93
d) Когда код пола для супруга(и) недостоверен	93
e) Когда код пола недостоверен и имеется информация о женщине	93
f) Когда код пола недостоверен и это лицо является мужем супруги	93
g) Когда код пола недостоверен, а информации для определения пола недостаточно	93
h) Примечание об вмененных соотношениях полов	94
3. Дата рождения и возраст	94
a) Возраст и дата рождения	94
b) Взаимосвязь между датой рождения и возрастом	95
c) Когда рассчитанный возраст выходит за верхний предел	96
d) Редактирование данных о возрасте	96
e) Редактирование данных о возрасте при наличии данных о возрасте главы домохозяйства и супруги(а) ..	96
f) Редактирование данных о возрасте главы домохозяйства при отсутствии данных о возрасте супруги(а) главы, но при наличии данных о возрасте ребенка	97
g) Редактирование данных о возрасте главы домохозяйства при наличии данных о возрасте родителя главы домохозяйства	97
h) Редактирование данных о возрасте главы домохозяйства при наличии данных о возрасте внука/внучки	98
i) Редактирование данных о возрасте главы домохозяйства при отсутствии данных о возрасте других лиц	98
j) Редактирование данных о возрасте супруги(а), когда уже определен возраст главы домохозяйства	98

	<i>Стр.</i>
k) Редактирование данных о возрасте других супружеских пар в домохозяйстве, когда известен возраст одного из супругов.	99
l) Редактирование данных о возрасте ребенка, когда уже определен возраст главы домохозяйства	99
m) Редактирование данных о возрасте родителя, когда уже определен возраст главы домохозяйства	100
n) Редактирование данных о возрасте внука/внучки, когда уже определен возраст главы домохозяйства	100
o) Редактирование данных по возрасту всех прочих лиц. .	100
4. Семейное положение	101
a) Редактирование данных о семейном положении	101
b) Отнесение к категории «семейное положение» в случаях, когда не применяется метод динамического вменения	101
c) Отнесение к категории «семейное положение», когда применяется метод динамического вменения	102
d) Супруг(а) должен(должна) состоять в браке	102
e) Супруг(а) супружеской пары	102
f) Если имеется супруг(а), то глава домохозяйства должен(должна) состоять в браке	102
g) Глава домохозяйства без супруги(а) и без детей.	102
h) Если все остальные вычисления не дают результатов, следует производить вменение	102
i) Взаимосвязь между возрастом и семейным положением молодых людей	103
5. Возраст вступления в первый брак	103
a) Графа возраста вступления в первый брак для лиц, никогда не состоявших в браке, не должна заполняться	104
b) В графе для лиц, когда-либо состоявших в браке, должна быть соответствующая запись	104
6. Рождаемость: дети, родившиеся живыми, и дети, живущие на момент переписи.	104
a) Собираемые элементы данных о рождаемости	105
b) Общие правила редактирования данных о рождаемости	105
c) Связь между данными о детях, родившихся живыми, и детях живущих на момент переписи	106
d) Редактирование, когда регистрируются только дети, родившиеся живыми	106
e) Редактирование, когда регистрируются дети, родившиеся живыми, и дети, живущие на момент переписи	107
f) Редактирование, когда имеются данные о детях, родившихся живыми, детях, живущих на момент переписи, и умерших детях	109
i) Когда имеются данные по всем трем элементам . . .	109
ii) Когда имеются данные по двум элементам	110

	<i>Стр.</i>
iii) Когда имеются данные по одному элементу	110
iv) Когда не имеется данных ни по одному из элементов	110
g) Редактирование, когда имеются данные о детях, родив- шихся живыми, детях, живущих дома, детях, живущих вне дома, и умерших детях	111
i) Когда имеются данные по всем четырем элементам.	111
ii) Когда имеются данные по трем из четырех элементов	112
iii) Когда имеются данные по двум из четырех элементов	112
iv) Когда имеются данные только по одному элементу	114
v) Когда не имеется данных ни по одному из элементов	114
h) Особый случай при наличии пяти или более элементов	114
i) Значение единого источника данных для всех элементов по рождаемости	115
j) Связь между данными о собственных детях, детях, проживающих в доме, и детях, живущих на момент переписи	115
7. Рождаемость: дата рождения последнего ребенка, родившегося живым, и рождение детей за 12 месяцев до переписи	115
8. Рождаемость: возраст при рождении первого ребенка	117
9. Смертность	118
a) Возраст и пол умершего лица	119
b) Причина смерти	119
c) Материнская смертность	119
d) Младенческая смертность	120
10. Сиротство в случае смерти матери/отца и номер строки матери	120
V. Характеристики миграции	121
1. Место рождения	122
a) Взаимосвязь между данными о стране рождения и количеством лет, прожитых в данном районе	122
b) Присвоение категории «неизвестно» для недоверен- ных записей о месте рождения	122
c) Применение метода статического вменения для данных о месте рождения	122
d) Применение метода динамического вменения для данных о месте рождения	123
e) Присвоение места рождения, если у данного лица есть мать	123
f) Присвоение места рождения для ребенка главы домохозяйства	123

	<i>Стр.</i>
g) Присвоение места рождения для ребенка, не являющегося ребенком главы домохозяйства	124
h) Присвоение места рождения для взрослых замужних женщин	124
i) Присвоение места рождения для взрослых незамужних женщин	124
j) Присвоение места рождения для лиц мужского пола ..	124
2. Гражданство	125
a) Редактирование данных о гражданстве	125
b) Взаимосвязь между этнической/расовой принадлежностью и гражданством	125
c) Взаимосвязь между натурализацией и гражданством ..	126
d) Взаимосвязь между продолжительностью проживания и гражданством	126
3. Продолжительность проживания	126
a) Редактирование данных о продолжительности проживания	126
b) Проживание и продолжительность де-факто/де-юре...	127
c) Взаимосвязь между возрастом и продолжительностью проживания	127
d) Взаимосвязь между местом рождения и продолжительностью проживания	127
e) Для тех лиц, которые всегда жили в данном месте	128
f) Определение продолжительности проживания данного лица по продолжительности проживания матери	128
g) Определение продолжительности проживания того или иного лица по продолжительности проживания ребенка	129
h) Продолжительность проживания лица, когда нет никакой другой информации	129
4. Прежнее место жительства	129
a) Редактирование данных о прежнем месте жительства ..	130
b) Прежнее место жительства при изменении границ	130
c) Если человек никуда не переезжал с рождения	130
d) Использование данных о других лицах в единице	130
e) Когда нет данных о других приемлемых лицах для редактирования прежнего места жительства	130
5. Место жительства в определенный момент в прошлом	131
6. Год прибытия	131
a) Взаимосвязь между возрастом и годом прибытия	132
b) Взаимосвязь между местом рождения и годом прибытия.	132
c) Для лиц, которые всегда жили в данном месте	132
d) Определение прибытия данного лица исходя из года прибытия матери	133
e) Определение года прибытия ребенка исходя из года прибытия главы домохозяйства	133

	<i>Стр.</i>
f) Определение года прибытия данного лица при отсутствии какой-либо другой информации	134
7. Взаимосвязь между продолжительностью проживания и годом прибытия	134
8. Постоянное место жительства	135
C. Социальные характеристики	136
1. Способность читать и писать (грамотность)	136
2. Посещение учебных заведений	137
a) Редактирование данных о посещении учебных заведений	137
b) Прием на очное или заочное обучение	138
c) Согласованность между сведениями о посещении учебных заведений и экономической деятельностью	138
d) Присвоение той или иной категории недостоверным или несогласующимся записям о посещении учебных заведений	138
3. Образовательная подготовка (последняя завершенная ступень или уровень)	138
a) Редактирование данных об образовательной подготовке	138
b) Минимальный возраст для образовательной подготовки	139
c) Взаимосвязь между возрастом и образовательной подготовкой	139
4. Область знаний и образовательный ценз	139
5. Вероисповедание	140
a) Редактирование данных о вероисповедании	140
b) Отсутствие данных о вероисповедании главы домохо- зяйства, но наличие данных о вероисповедании какого- либо другого лица в жилищной единице	140
c) Отсутствие данных о вероисповедании главы домохозяйства или любого другого лица в жилищной единице	140
d) Отсутствие данных о вероисповедании лица, не являющегося главой домохозяйства	141
6. Язык	141
a) Редактирование данных о языке	141
b) Редактирование данных о языке: глава домохозяйства	141
c) Редактирование данных о языке: лица, отличные от главы домохозяйства	141
d) Редактирование данных о языке: использование данных об этническом происхождении или месте рождения	142
e) Редактирование данных о языке: родной язык	142
f) Редактирование данных о языке: умение говорить на обозначенном языке	142

	<i>Стр.</i>
7. Этническая принадлежность и коренные народы	142
a) Редактирование данных об этнической принадлежности	143
b) Редактирование данных об этнической принадлежности: глава домохозяйства	143
c) Редактирование данных об этнической принадлежности: лица, не являющиеся главой домохозяйства	143
d) Редактирование данных об этнической принадлежности: использование сведений о языке и месте рождения	144
8. Нетрудоспособность	144
a) Вопросы переписи, касающиеся нетрудоспособности	144
b) Редактирование данных о нетрудоспособности	145
c) Многие виды нетрудоспособности	145
d) Редактирование данных о причинах нетрудоспособности	145
D. Экономические характеристики	145
1. Статус экономической активности	146
a) Категории, относящиеся к статусу экономической активности	147
i) Незанятое население	147
ii) Поиск работы	148
iii) Экономически неактивное на данный момент население	148
iv) Почему не ведется поиск работы	149
b) Редактирование данных о статусе экономической активности	149
i) Работающие лица	149
ii) Экономическая активность неработающих лиц	149
iii) Экономическая активность учащихся и пенсионеров	150
iv) Когда запись об экономической активности недостоверна и указаны переменные, касающиеся работающих лиц	150
v) Когда запись об экономической активности недостоверна и указаны переменные, касающиеся неработающих лиц	150
vi) Когда запись об экономической активности недостоверна и не указаны никакие экономические переменные	150
2. Отработанное время	150
3. Род занятий	151
4. Отрасль	151
5. Статус занятости	152
6. Доход	153

	<i>Стр.</i>
7. Институциональный сектор	154
8. Занятость в неформальном секторе	154
9. Место работы	154
<i>Глава</i>	
V. Редактирование данных о жилом фонде	157
A. Основные и дополнительные темы	159
1. Жилые помещения: тип жилых помещений (основная тема)	159
2. Местонахождение жилых помещений (основная тема)	161
3. Сведения о занятости (основная тема)	161
4. Форма собственности (основная тема)	162
5. Число комнат (основная тема)	163
6. Число спальных комнат (дополнительная тема)	163
7. Полезная площадь помещения (дополнительная тема)	164
8. Система водоснабжения (основная тема)	164
9. Основной источник питьевой воды (основная тема)	165
10. Тип туалета (основная тема)	166
11. Канализационная система (основная тема)	166
12. Ванные помещения и оборудование (основная тема)	167
13. Наличие кухни (основная тема)	168
14. Топливо, используемое для приготовления пищи (основная тема)	169
15. Освещение и/или электроснабжение (основная тема)	170
16. Основной метод удаления твердых отходов (основная тема)	170
17. Тип отопления и используемая энергия (дополнительная тема)	171
18. Наличие горячей воды (дополнительная тема)	171
19. Наличие магистрального газа (дополнительная тема)	172
20. Использование жилищной единицы (дополнительная тема)	172
21. Занятость более чем одним домохозяйством (основная тема)	173
22. Число жильцов (основная тема)	173
23. Тип здания (основная тема)	173
24. Год или период строительства (дополнительная тема)	174
25. Число жилищных единиц в здании (дополнительная тема) ..	174
26. Строительный материал для наружных стен (основная тема)	174
27. Строительный материал для пола и крыши (дополнитель- ная тема)	175
28. Наличие лифта (дополнительная тема)	175
29. Сельскохозяйственное строение (дополнительная тема)	176
30. Эксплуатационное состояние (дополнительная тема)	176
31. Характеристики главы или другого основного члена домохозяйства (основная тема)	177

	<i>Стр.</i>
32. Условия владения жильем (основная тема)	177
33. Арендная плата за жилье и жилищные расходы владельца (дополнительная тема)	178
34. Меблированные и немеблированные (дополнительная тема)	178
35. Наличие информационно-коммуникационных устройств (основная тема)	178
36. Число автомобилей (дополнительная тема)	179
37. Наличие бытовой техники длительного пользования (дополнительная тема)	180
38. Наличие свободной площади вне помещения (дополнительная тема)	180
В. Занятые и незанятые жилищные единицы	181

Приложения

I. Производные переменные	183
А. Производные переменные для записей о жилом фонде	183
1. Доход домохозяйства	183
2. Семейный доход	184
3. Нуклеарная семья	184
4. Тип домохозяйства	185
5. Состав домохозяйства	185
6. Состав семьи	186
7. Статус домохозяйства и семьи	187
8. Влияние ВИЧ/СПИДа на структуру домохозяйства	188
9. Лица, связанные родственными отношениями	188
10. Работники в семье	189
11. Полный комплект сантехнического оборудования	189
12. Полный комплект кухонного оборудования	189
13. Валовая квартирная плата	190
14. Индекс благосостояния	190
В. Производные переменные для записей о населении	191
1. Статус экономической активности	191
2. Собственные дети	191
3. Родители в доме	192
4. Текущий год в учебном заведении	192
5. Число месяцев с даты последних родов	192
II. Взаимосвязь между форматом переписного листа и вводом данных с клавиатуры	195
III. Сравнение сканирования и ввода данных с клавиатуры	201
А. Ввод данных	201
1. Сканирование	201

	<i>Стр.</i>
2. Ввод данных с клавиатуры	202
a) Ввод данных в режиме «heads down» без схем пропуска элементов	202
b) Ввод данных в режиме «heads down» со схемами пропуска элементов	203
3. Интерактивный ввод данных с клавиатуры	204
V. Проверка данных	205
1. Зависимая проверка	206
2. Независимая проверка	206
C. Соображения по вопросам редактирования отсканированных данных	206
D. Выводы	208
IV. Образцы рабочих блок-схем	209
V. Методы вменения	215
VI. Программные пакеты компьютерного редактирования	221
Глоссарий	225
Библиография	233

Таблицы

1. Выборка населения с разбивкой по возрастным группам с интервалом 15 лет и полу с использованием неотредактированных и отредактированных данных	13
2. Население и изменения в его составе по возрастным группам с интервалом 15 лет с неизвестными величинами: 2000 и 2010 годы	14
3. Население и изменения в его составе по возрастным группам с интервалом 15 лет без неизвестных данных: 2000 и 2010 годы	15

Рисунки

1. Примеры общих кодов для отдельных элементов данных	31
2. Типовое гипотетическое домохозяйство с указанием родства, пола и данных о рождаемости его членов	36
3. Пример домохозяйства, где глава и супруг одного пола	36
4. Пример домохозяйства с указанием возраста некоторых членов домохозяйства	37
5. Пример домохозяйства с возможной несогласованностью указанных возрастных данных	38
6. Пример правил редактирования отдельных характеристик населения по принципу групповых переменных	41

	<i>Стр.</i>
7. Пример с главой и супругом одного пола в неотредактированном наборе данных и его разрешение	41
8. Выборочные спецификации редактирования для исправления переменной пола с символическим кодом	42
9. Пример анализа данных об очень молодой вдове с тремя детьми, отредактированных по принципу групповых переменных	42
10. Выборочные домохозяйства в качестве примера входных данных для динамического вменения	46
11. Исходная статическая матрица для возраста, исчисляемого исходя из пола и степени родства	47
12. Пример матрицы динамического вменения после одного изменения ..	47
13. Пример матрицы динамического вменения после многократных изменений	48
14. Пример главы домохозяйства и отца главы без указания языка	52
15. Исходные значения для матрицы динамического вменения языка ..	52
16. Пример членов домохозяйства без указания языка	53
17. Пример главы домохозяйства с ребенком при отсутствии данных о возрасте ребенка	55
18. Пример главы домохозяйства с ребенком при отсутствии данных о возрасте и уровне образования ребенка	55
19. Выборочный набор величин для матрицы «cold deck» и выборочный код вменения	58
20. Пример сводного отчета для ряда вменений на одну ошибку	59
21. Выборочное сообщение об ошибках в переписном листе	59
22. Пример дополнительной распечатки ошибок по переписному листу, включая множественные переменные	60
23. Выборочные демографические записи с признаками для вменяемых значений	63
24. Пример флага для молодой женщины с пробелом в графе «рождаемость» и добавлением флага	64
25. Пример домохозяйства, где глава домохозяйства зарегистрирован как первое лицо	81
26. Пример домохозяйства, где глава домохозяйства зарегистрирован как пятое лицо	82
27. Иллюстрация домохозяйства с информацией о рождаемости	108
28. Исходные значения для определения детей, живущих на момент переписи, когда возраст и число детей, родившихся живыми, достоверны	109
29. Образец матрицы вменения, которую необходимо составить для пар известных элементов данных	113

	<i>Стр.</i>
<i>Вставки</i>	
1. Для чего нужно редактировать данные переписи	11
2. Основные руководящие принципы коррекции данных	35
3. Руководящие принципы редактирования структуры данных.	67
4. Редактирование и вменение возраста	95
<i>Рисунки приложений</i>	
A.II.1. Образец бланка вопросника с персональными страницами	195
A.II.2. Пример последовательности информации в вопроснике с персональными страницами	196
A.II.3. Образец вопросника по домохозяйству с данными по всем лицам на одной странице	196
A.II.4. Пример последовательности для вопросника со страницами по домохозяйству, причем на одной странице содержатся данные о нескольких лицах	197
A.II.5. Пример страницы по домохозяйству с указанием нескольких лиц, без проблем в отношении ввода данных с клавиатуры	198
A.II.6. Пример страницы по домохозяйству с указанием нескольких лиц с возможными проблемами в отношении ввода данных с клавиатуры	198
A.IV.1. Образец блок-схемы для определения главы домохозяйства (ГД).	210
A.IV.2. Образец блок-схемы для определения супруги (супруга) в домохозяйстве.	211
A.IV.3. Образец блок-схемы для редактирования переменной для пола главы домохозяйства и супруги (супруга)	212

Сокращения

ГИС	географические информационные системы
МВБС	Методология вменения по ближайшему соседу
МКРНИ	Международная классификация расстройств здоровья, нетрудоспособности и инвалидности
МОТ	Международная организация труда
ОРМ	оптическое распознавание меток
ОРС	оптическое распознавание символов
ПРО	пострегистрационное обследование
СНС	Система национальных счетов
СПОН	статистический пакет для общественных наук
ССА	система статистического анализа
СУ	счетный участок
ЮНЕСКО	Организация Объединенных Наций по вопросам образования, науки и культуры
<i>CD</i>	умершие дети
<i>CEB</i>	дети, родившиеся живыми
<i>CLA</i>	дети, живущие вне дома
<i>CLH</i>	дети, живущие дома
<i>CS</i>	число детей, живущих на момент переписи

Введение

А. Цель Руководства

1. Хорошо спланированная перепись или обследование¹ с минимальными ошибками в конечном продукте является бесценным ресурсом для государства. Для получения точных результатов переписи или обследования данные должны быть в максимально возможной степени свободны от ошибок и несоответствий, особенно после этапа обработки данных.

Процедура обнаружения ошибок в записях данных и между ними во время и после сбора данных и корректирования отдельных элементов известна как редактирование данных переписи населения и жилого фонда.

2. Абсолютно совершенных данных переписи или обследования не бывает. Страны уже давно признали наличие проблем в связи с данными переписей и обследований и приняли различные подходы к тому, как устранять пробелы в данных и непоследовательность ответов. Вместе с тем из-за длительных перерывов между переписями, процедуры редактирования данных зачастую должным образом не документируются. В результате для новой переписи или обследования страны вынуждены заново придумывать процедуры, которые уже использовались во время предыдущей работы по сбору данных.

3. Настоящее *Руководство по редактированию данных переписи населения и жилого фонда* призвано заполнить этот пробел в знаниях в области методологии редактирования данных переписей и обследований и предоставить информацию соответствующим должностным лицам о применении различных подходов к редактированию данных переписи. *Руководство* преследует также цель содействовать тому, чтобы страны сохраняли накопленный ими опыт редактирования данных, укрепляли связь между специалистами, занимающимися содержательной частью переписи, и специалистами по обработке данных и документально отражали то, что делается в ходе текущей переписи населения или обследования во избежание дублирования работы в будущем.

4. Настоящее *Руководство* является справочным пособием как для специалистов, занимающихся содержательной частью переписи², так и для специалистов по обработке данных или методистов, поскольку они ведут совместную работу по разработке спецификаций редактирования и программ для переписей и обследований. Оно построено по принципу «книги по кулинарии», которая позволяет странам использовать такие приемы редактирования, которые наиболее приемлемы с точки зрения сложившейся у них ситуации в статистической сфере. Настоящее издание предназначено также для того, чтобы содействовать более тесному общению между специалистами в разработке и осуществлении программы редактирования.

¹ Перепись представляет собой полный подсчет. В рамках обследования обычно регистрируется небольшая часть общей совокупности. Описанные в этом издании методы редактирования применимы к обоим видам деятельности.

² По определению, содержащемуся в данном *Руководстве*, к специалистам, занимающимся содержательной частью переписи относятся демографы, социологи, экономисты и другие специалисты, работающие в области народонаселения, жилья и в других смежных отраслях.

5. Во введении описывается процесс переписи, приводятся различные виды ошибок, которые возникают во время переписи, и основы редактирования данных. В последующих главах рассказывается о процедурах и методах редактирования данных переписи на различных этапах их обработки. Хотя настоящее *Руководство* посвящено главным образом редактированию результатов переписи населения и жилого фонда, многие из этих понятий и методов можно применять и при проведении обследований.

В. Процесс переписи

6. Перепись населения и/или жилого фонда — это весь процесс сбора, обработки, оценки, анализа и опубликования демографических, экономических, социальных данных и данных о жилом фонде в отношении всех лиц и помещений, в которых они проживают (United Nations, 2007). Традиционно переписи проводятся в конкретное время по всей стране или в ее четко обозначенной части. Недавно некоторые страны начали проводить непрерывные обследования в масштабе всей страны, используя «подробный» формат переписных листов для обеспечения полного охвата в динамике по времени. В рамках любого из этих сценариев перепись отображает положение населения и состояние жилого фонда на определенную дату.

7. Главная цель переписи заключается в том, чтобы предоставить информацию о численности, распределении и характеристиках населения страны. Данные переписи используются при выработке политики, планировании и управлении, а также при ведении и оценке программ в области образования, трудовых ресурсов, планировании семьи, жилья, здравоохранения, транспорта и развития сельских районов. В административных целях данные переписи находят свое основное применение в демаркации избирательных округов и распределении представительства в органах управления. Перепись является также неоценимым источником для проведения научных исследований, представляя данные для научного анализа состава и распределения населения и для статистических моделей прогнозирования будущего роста. Перепись обеспечивает деловые круги и промышленность основными данными, необходимыми им для оценки спроса на жилье, учебные заведения, предметы обстановки, продукты питания, одежду, места отдыха и развлечений, предметы медицинского назначения и другие товары и услуги.

8. Переписи и обследования имеют ряд общих основных признаков, которые включают: *a)* подготовительную работу; *b)* регистрацию или сбор данных; *c)* обработку данных, в том числе ввод данных (с клавиатуры или путем сканирования), редактирование и составление таблиц; *d)* создание баз данных и распространение результатов; *e)* оценку результатов; и *f)* анализ результатов.

9. Подготовительная работа включает множество элементов, таких как определение правовой основы для проведения переписи; составление бюджета; разработка календарного плана; административная организация; составление карт; составление перечня жилых единиц; разработка программы табулирования; подготовка переписного листа; а также разработка планов и подготовка работников для проведения регистрации, предварительных проверок, обработки данных и их распространения.

10. Процесс регистрации зависит от избранного метода, сроков и длительности периода регистрации, уровня инспектирования и от того, используется ли и как используется та или иная выборка. После сбора данных их необходимо закодировать, ввести, обработать, отредактировать и свести в таблицы. Во время обработки данных создаются базы микро- и макроданных. Национальные бюро переписи/статистические управления используют эти базы данных для составления таблиц, анализа временных рядов, составления диаграмм и картограмм, а также в географических информационных системах (ГИС) для составления тематических карт и других методов распространения данных. Результаты оцениваются как с точки зрения содержания, так и с точки зрения охвата с использованием целого ряда методов, включая демографический анализ и контрольные обследования после регистрации. И наконец, результаты анализируются с использованием самых различных методов, включая описательные сводки результатов, стратегический анализ результатов переписи и детализированные аналитические исследования одного или более аспектов демографической и социальной ситуации в стране.

С. Ошибки в процессе переписи

11. Данные переписи изобилуют многими ошибками, которые в целом можно классифицировать как ошибки в охвате и ошибки в содержании.

1. Ошибки в охвате

12. Ошибки в охвате возникают в результате пропусков или двойных записей о лицах или жилищных единицах при регистрации в ходе переписи. Источниками ошибок в охвате являются, в частности, неполные или неточные карты или списки счетных участков, неспособность счетчиков охватить все единицы на закрепленных за ними участках, двойной учет, пропуск лиц, которые не желают регистрироваться, ошибочный учет некоторых категорий лиц, таких как гости или иностранцы из числа нерезидентов, и утрата или уничтожение записей переписи после регистрации. Ошибки в охвате следует исправлять, насколько это возможно, на месте. В ходе редактирования в офисе фактические двойные записи устраниваются. Однако при этом необходимо определить, не зарегистрированы ли эти лица или домохозяйства повторно. Например, о близнецах может быть одинаковая информация, за исключением порядкового номера. Следовательно, применяемые в ходе такой обработки правила редактирования определяют, когда учитывать, когда исключать кажущуюся дублированной информацию, а когда вносить изменения путем вменения значений.

13. При структурном редактировании, о котором говорится в главе III, проводится проверка домохозяйств на правильность указания числа лиц и порядковых номеров и на наличие дважды зарегистрированных лиц.

2. Ошибки в содержании

14. Ошибки в содержании возникают в результате неправильной отчетности или регистрации характеристик лиц, домохозяйств и жилищных единиц. Ошибки в содержании могут быть вызваны плохо сформулированными вопросами, неудачной последовательностью вопросов или недоста-

точным общением между респондентом и счетчиком, а также ошибками в кодировании и вводе данных, ошибками в ручном и компьютерном редактировании и ошибочном табулировании результатов. На каждом этапе данного процесса необходимо должным образом вести и сохранять журнал редактирования (известный также как контрольный журнал), с тем чтобы исключить потерю данных. В следующих разделах дается разъяснение каждой из вышеупомянутых ошибок.

a) Ошибки в разработке переписного листа

15. Одним из источников ошибок в содержании являются плохо сформулированные вопросы или инструкции. Тип переписного листа, его формат и точность формулировок или расположение пунктов вопросника заслуживают самого тщательного рассмотрения, поскольку недостатки плохо составленного переписного листа нельзя устранить ни во время регистрации, ни после нее. Для того чтобы свести к минимуму ошибки, которые могут возникать из-за плохо составленных переписных листов, следует использовать предварительное тестирование. Если, например, схемы пропусков указаны нечетко или расположены не на месте, счетчик может по ошибке пропустить разделы вопросника и не собрать всю необходимую информацию.

b) Ошибки счетчиков

16. Счетчики и респонденты взаимодействуют между собой, если только перепись не проводится с использованием переписных листов, заполняемых самими респондентами. Счетчик может ошибиться, задавая вопросы и при этом сокращая или изменяя формулировку вопросов или не полностью разъясняя респонденту смысл вопросов. Счетчик может также делать ошибки при записи ответов. Качественный уровень счетчиков и их подготовки оказывает решающее влияние на качество собираемых данных. Счетчики должны быть надлежащим образом подготовлены по всем аспектам проведения переписи. Они должны понимать, почему столь важна их роль в процессе переписи и как регистрация сочетается с другими этапами переписи. Кроме того, поскольку счетчики могут обладать самой разной квалификацией и иметь разные уровни образования, подготовка должна вестись таким образом, чтобы добиться от счетчиков умения задавать вопросы так, чтобы получить правильный ответ.

c) Ошибки респондентов

17. Ошибки могут вноситься в данные, если респонденты неправильно понимают некоторые вопросы. Ошибки могут возникать также в результате преднамеренного сообщения неправильных сведений, или когда ответы на вопросы даются через доверенное лицо (когда ответы на вопросы дает не тот, кого касается запрашиваемая информация). Качество отдельных ответов можно улучшить, рекламируя перепись и занимаясь подготовкой счетчиков, с тем чтобы разъяснять им цель переписи и причины постановки различных вопросов. Некоторые страны используют переписные листы, заполняемые опрашиваемыми лицами, поэтому взаимодействия между счетчиком и респондентом не происходит. Если бланки заполняются самими опрашиваемыми

ми лицами, ошибки возникают, если респонденты неправильно понимают вопросы или инструкции.

18. Ошибками респондентов и счетчиков лучше всего заниматься на этапе регистрации, когда еще доступны и бланки, и респонденты, и счетчики. Инспекторы должны уметь готовить счетчиков. Инспекторы должны также уметь проверять данные, собранные счетчиками, и делать это регулярно во время регистрации, с тем чтобы счетчики не вносили систематические погрешности в данные. Инспекторы должны исправлять ошибки счетчика и респондента на месте до того, как переписные листы будут направлены в региональные или центральные бюро.

d) Ошибки кодирования

19. Ошибки могут возникать в ходе кодирования, поскольку кодировщик может неправильно закодировать информацию. Во время ввода данных ошибки могут вноситься в результате неправильного использования клавиатуры. В целом отсутствие контроля и проверки на этом этапе задерживает публикацию данных, поскольку на более позднем этапе обнаружить и исправить ошибки становится сложнее. Зачастую до или во время кодирования ведется ручное редактирование.

e) Ошибки при вводе данных

20. Диапазонные проверки и некоторые основные проверки на согласованность можно встроить в программное обеспечение ввода данных для предотвращения недостоверных вводов. Интеллектуальная система ввода данных гарантирует, что величина каждого поля или элемента данных будет в пределах допустимого диапазона величин для этого элемента. С такой системой увеличиваются шансы на то, что оператор по вводу данных будет набирать допустимые данные и снимать часть нагрузки, связанной с редактированием данных на более поздних этапах процесса подготовки информации. Вместе с тем эти проверки могут замедлить скорость ввода данных. В связи с этим во время ввода данных следует тщательно взвесить объем проверок на согласованность с учетом необходимости поддержания приемлемой скорости ввода данных. Для этого требуется заранее добиться определенного баланса, с тем чтобы операторы по вводу данных не тратили слишком много времени на эту работу. Проверка набора данных неизбежно улучшает их качество. Набранные бланки можно проверить путем повторного ввода с клавиатуры той же информации, зачастую на выборочной основе. Взаимосвязь между форматом переписного листа и вводом данных с клавиатуры рассматривается в приложении II, а сравнительные характеристики ввода данных с клавиатуры и сканирования рассматриваются в приложении III.

f) Ошибки при компьютерном редактировании

21. Одним из важнейших этапов обработки данных переписи является редактирование. В процессе редактирования меняются или корректируются недостоверные или несогласующиеся данные путем вменения отсутствующих ответов или несогласующейся информации на основе достоверных данных.

Как это ни парадоксально, любая из этих операций в процессе редактирования может повлечь за собой новые ошибки.

г) Ошибки при составлении таблиц

22. Ошибки могут возникать на этапе составления таблиц из-за ошибок в обработке данных или использования «неизвестной» (не предоставленной) информации. На этом этапе ошибки трудно исправлять, не делая при этом новых ошибок. Отсутствие перекрестной проверки таблиц и ошибки при печатании ведут к ошибкам на этапе публикации. Вместо того чтобы пытаться исправлять сами таблицы, необходимо применять систему обработки данных таким образом, чтобы при появлении несоответствий в таблицах производилось дополнительное редактирование. Если ошибки имеют место на всех этапах работы вплоть до публикации, они становятся заметными, и тогда ценность результатов будет сомнительной. Если «исправления» вносятся на этапе составления таблиц, например если обнаружены несколько различных неизвестных величин и включены в «суммарные» результаты, но не включены в распределение величин, другие аналитики не смогут воспроизвести такие таблицы. Более разумно рассматривать обработку данных переписи как систему обратной связи, с тем чтобы вносить изменения в набор данных в процессе редактирования, а не на этапе составления таблиц. Прежде чем выпускать таблицы, необходимо их тщательно проверить, чтобы убедиться в том, что все запланированные таблицы подготовлены для всех намеченных территориальных единиц. Хотя введенная на этапе редактирования диапазонная проверка и проверка на согласованность позволяют устранить большинство ошибок, после составления таблиц необходимо провести комплексную проверку, иногда называемую «макроредактированием». Подготовленные и опытные сотрудники должны просмотреть различные таблицы, с тем чтобы проверить, соответствуют ли указанные цифры в различных ячейках таблицы известной ситуации на местах. В ограниченном количестве случаев ошибки кодирования можно выявлять путем быстрых просмотров переписных листов. Может оказаться полезным и расчет отдельных отношений и темпов роста, а также сопоставление с показателями предыдущей переписи и другими показателями, опубликованными по результатам выборочных обследований. Однако сравнение с другими показателями, основанными на результатах обследований или административных данных, можно проводить лишь в том случае, если используемые понятия сопоставимы. В случае обнаружения ошибок в итоговых таблицах исправления следует вносить прежде всего в набор микроданных, в частности для того, чтобы позволить составлять сопоставимые таблицы другим специалистам по обработке данных в национальных бюро переписи/статистических управлениях. Кроме того, поскольку национальные бюро переписи/статистические управления иногда предоставляют части файлов микроданных ученым и другим пользователям в государственном и частном секторах, таблицы должны быть воспроизводимыми.

23. Как указывалось выше, процесс переписи состоит из целого ряда последовательных, взаимосвязанных операций, и ошибки могут возникать при каждой операции. Важно не забывать о том, что компьютерное редактирование является частью системы обратной связи и не только направлено на построение таблиц, но и также имеет обратную связь в направлении сбора и обработки

данных на местах. Лучше всего национальное бюро переписи/статистическое управление может избежать проблем, связанных с автоматизированным редактированием, путем максимального применения редактирования на местах. Национальное бюро переписи/статистическое управление должно также быть уверено в том, что кодирование и ввод данных произведены точно, и должно иметь постоянную обратную связь по всем операциям, включая ввод данных, их редактирование и построение таблиц.

D. Структура Руководства

24. В главе I рассматривается роль редактирования в проведении переписей и обследований. Другие главы посвящены конкретным вопросам. В главе II говорится о прикладном назначении редактирования, включая методы измерения величин. Глава III посвящена редактированию структуры данных, которое одновременно охватывает жилищные и демографические единицы, а также определенные процедуры, способствующие редактированию остальных элементов, например уточнению вопроса о том, есть ли один и только один глава домохозяйства. В главе IV рассматривается вопрос редактирования данных по народонаселению, а в главе V — редактирование данных по жилищным единицам. И наконец, ряд приложений посвящен конкретным вопросам, связанным с редактированием и вменением данных переписей населения и жилого фонда.

Глава I

Редактирование при проведении переписей и обследований

А. Редактирование: историческая справка

25. До появления компьютеров большинство операций в рамках переписи требовали привлечения большого числа конторских сотрудников средней квалификации для редактирования каждого бланка переписного листа. Однако в силу сложных взаимосвязей между даже небольшим количеством позиций данных простая проверка не могла охватить все вероятные несоответствия в данных. Разные сотрудники могут по-разному интерпретировать правила, и даже один и тот же сотрудник может быть непоследовательным.

26. С появлением компьютеров редактирование данных переписи заметно изменилось. С помощью компьютера можно выявить намного больше несоответствий, чем при ручном редактировании. Спецификации редактирования становятся все более утонченными и сложными. Стало возможным автоматизированное вменение данных с соответствующими правилами такой работы (Nordbotten, 1963; Naus, 1975). В то же время такой процесс позволяет все чаще общаться с респондентами или по крайней мере иметь вопросники, заполненные этими респондентами. Многие группы редактирования стали понимать, что чем больше объем редактирования, тем лучше, и чем сложнее редактирование, тем точнее результаты. Благодаря этим программам появились тысячи сообщений об ошибках, требующих просмотра первоначальных бланков в ручном режиме, а при некоторых обследованиях — и повторного опроса респондентов³.

27. С компьютерами становится все легче вносить изменения в набор данных. Иногда эти изменения корректируют записи или позиции данных. Многие записи проходят через компьютер множество раз с ошибками и несоответствиями, и каждый раз их просматривают разные лица (Boucher 1991; Granquist, 1997).

28. В результате всего этого процесса появился целый ряд обобщенных пакетов редактирования данных переписи, и некоторые из них используются до сих пор. Первоначально такие пакеты были разработаны для больших компьютеров; позднее некоторые из них были модифицированы для использования в персональных компьютерах. За этот период Фелледжи и Холт (Fellegi and Holt, 1976) разработали новый метод обобщенного редактирования и вменения, который не сразу вошел в практику, но стал все шире применяться сегодня, по мере того как национальные бюро переписи/статистические управления все больше совершенствуют процесс редактирования.

29. Крупный шаг вперед в редактировании данных переписи был сделан в 1980-е годы, когда национальные бюро переписи/статистические управления

³ В приложении V рассматриваются различные методы вменения; в приложении VI обсуждаются компьютерные пакеты редактирования.

стали пользоваться персональными компьютерами для ввода и редактирования своих данных и их сведения в таблицы. При этом обработчики данных сразу же получили возможность осуществлять редактирование в интерактивном режиме на этапе ввода данных или вскоре после этого. В рамках обследований и переписей в небольших странах появилась возможность разрабатывать программы для выявления ошибок во время сбора или введения данных непосредственно в компьютер. Компьютерное редактирование позволило расширить непрерывные контакты с респондентами для устранения проблем, возникающих в процессе редактирования (Pierzchala, 1995).

30. В первые годы процесс проведения все более усложненных и тщательных проверок данных переписей и обследований представлялся весьма успешным. Группы редактирования создавали все более сложные спецификации редактирования, а специалисты по обработке данных месяцами занимались разработкой графических схем или схем принятия решений и программных кодов. Аналитики редко занимались оценкой пакетов редактирования. Казалось, что путем редактирования можно решить любые проблемы, возникающие на более ранних этапах сбора, кодирования и ввода данных с клавиатуры. Однако при этом для многих аналитиков стало очевидным, что во многих случаях все эти операции с дополнительным редактированием наносили ущерб данным или по меньшей мере задерживали получение результатов или служили причиной погрешностей в этих результатах. Иногда программа делала так много проходов через данные, корректируя сначала один элемент, затем другой элемент, что результаты получались несопоставимыми с первоначальными, необработанными данными.

31. При проведении многих переписей и крупных обследований такое широкое редактирование вызвало значительные задержки в процессе проведения переписи или обследования. Сотрудники тратили много времени на поиски бланков вручную. Специалисты по обработке данных продолжали разрабатывать прикладные программы, которые относились к весьма небольшому количеству случаев. Гранквист (Granquist, 1997) отмечает, что как показали многие исследования, во многих случаях при такой дополнительной работе «качество повышается незначительно, не повышается совсем или даже ухудшается; многие виды серьезных систематических ошибок путем редактирования выявить нельзя».

32. По мере того как национальные статистические организации продолжают совершенствовать методику проведения переписей и обследований, компьютерное редактирование в широких масштабах становится возможным и даже реальным. Следовательно, вопрос, который должен стоять перед каждым национальным бюро переписи/статистическим управлением, заключается в том, какой уровень компьютерного редактирования подходит для их целей.

В. Группа редактирования

33. При подготовке к проведению переписи национальным статистическим управлениям необходимо рассмотреть множество возможностей повышения качества своей работы. Одной из таких возможностей является создание группы редактирования. Ответственность за процесс редактирования должна возлагаться на группу редактирования в составе руководителей переписи, спе-

специалистов по содержательной части переписи и обработчиков данных. Такую группу следует создавать, как только начинается подготовка к переписи, желательно — во время составления переписного листа. Группа редактирования играет важную роль с самого начала, и такая роль сохраняется за ней в течение всего процесса редактирования. Тщательный подход к подбору группы и к разработке и выполнению правил редактирования и вменения обеспечивает более оперативное и более эффективное проведение переписи.

Вставка 1

Для чего нужно редактировать данные переписи

Редактирование данных переписи преследует следующие цели:

- ☞ обеспечить пользователей высококачественными данными переписи;
- ☞ выявить типы и источники ошибок;
- ☞ предоставить откорректированные результаты переписи.

34. Совместные совещания должностных лиц переписи с пользователями по вопросам построения таблиц и создания других информационных продуктов могут позволить проникнуть в суть редактирования, которое необходимо осуществить. Зачастую пользователи просят составить конкретную таблицу или тип таблиц, которые требуют дополнительного редактирования, с тем чтобы устранить возможные несоответствия. Группа редактирования должна запланировать построение таких таблиц на первоначальном этапе редактирования, вместо того чтобы составлять их специально после обработки результатов переписи. Разработка правил редактирования и компьютерных программ во время предварительного обследования или «генеральной репетиции» позволяет проверить сами программы и сократить длительность цикла обработки на различных этапах процесса редактирования и вменения. Затем группа редактирования устанавливает степень воздействия этих различных процессов и в случае необходимости принимает меры по исправлению положения.

35. Разработкой правил редактирования и вменения должны заниматься совместно специалисты по содержательной части и по обработке данных. На самом начальном этапе подготовки к переписи группа редактирования составляет план выявления и редактирования ошибок. Группа редактирования данных переписи или обследования составляет в письменном виде свод правил проверки соответствия и исправления ошибок.

36. Наряду с разработкой правил редактирования и вменения специалисты по содержательной части переписи и специалисты по обработке данных должны работать вместе на всех этапах проведения переписи или обследования, в том числе и в период проведения анализа. Риск проведения чрезмерного редактирования столь же велик, как и риск отредактировать слишком мало данных и получить необработанную или неверную информацию в наборе данных. В связи с этим обе группы должны взять на себя ответственность за надлежащее ведение баз метаданных. Группа редактирования должна также эффективно использовать имеющиеся административные источники и регистры обследований, с тем чтобы лучше проводить последующие операции во время переписи или обследования.

37. Когда национальные бюро переписи/статистические управления использовали большие компьютеры, общение между специалистами по содержательной части переписи и специалистами по обработке данных было ограничено. Такое разделение продолжалось и некоторое время после появления микрокомпьютеров, но пакеты компьютерных программ стали более удобными для пользователя, и сейчас многие специалисты по содержательной части переписи могут фактически разрабатывать и испытывать собственные планы редактирования и составления таблиц. Хотя специалисты по содержательной части, как правило, сами не обрабатывают данные, они зачастую понимают те шаги, которые осуществляют специалисты по обработке данных в ходе работы.

С. Практика редактирования: отредактированные и неотредактированные данные

38. Страны проводят редактирование результатов переписи, с тем чтобы улучшить данные и их представление. В этом разделе *Руководства* освещается проблема, с которой сталкиваются национальные бюро переписи/статистические управления, когда публикуются неотредактированные данные переписи. Для иллюстрации используется гипотетический набор данных.

39. Национальное бюро переписи/статистическое управление некой вымышленной страны сталкивается с необходимостью выбора возможности постараться удовлетворить запросы многочисленных пользователей. Некоторым пользователям, возможно, потребуются неизвестные введенные данные, используемые для анализа или научных исследований, а другие, возможно, захотят иметь данные с минимальными искажениями (возможными ошибками) для целей планирования или проведения своей политики. Если национальное бюро переписи/статистическое управление распространит необработанную таблицу с такими данными, как, например, в левой части таблицы 1, то и аналитики, и директивные органы будут вынуждены строить свои допущения, исходя из этих данных. В таблице 1 это показано на примере лишь небольшого количества человек. Из нее видно, что в отношении 23 человек в этой стране не зарегистрирован пол, а в отношении 15 человек — возраст. Эти пропуски могли возникнуть в результате отсутствия ответа или ошибок при вводе данных с клавиатуры. В двух случаях не зарегистрирован ни пол, ни возраст.

40. Большинство пользователей сами решают, что им делать с неизвестными величинами. Логический и, возможно, наивный подход заключается в том, чтобы распределить неизвестные величины в той же пропорции, что и известные. Если национальное бюро переписи/статистическое управление примет решение произвести вменение для неизвестных величин, то группа редактирования может решить оставить 12 мужчин и 11 женщин, то есть примерно в одинаковой пропорции, однако этот показатель будет искаженным, потому что при переписи было зарегистрировано больше женщин. В этом случае результаты будут соотноситься с отредактированными данными, указанными в правой части таблицы 1.

41. Для обработки неизвестных величин существуют и другие варианты. Например, группа редактирования может решить использовать вменение данных, исходя только из распределения по полу, игнорируя другую имеющуюся информацию, например отношения между супругами, зарегистрировано ли лицо неуказанного пола как мать другого лица или дается ли положительный

Таблица 1

Выборка населения с разбивкой по возрастным группам с интервалом 15 лет и полу с использованием неотредактированных и отредактированных данных

Возрастная группа	Неотредактированные данные				Отредактированные данные		
	Всего	Мужчины	Женщины	Не зарегистрировано	Всего	Мужчины	Женщины
Всего	4 147	2 033	2 091	23	4 147	2 045	2 102
Младше 15 лет	1 639	799	825	15	1 743	855	888
15–29 лет	1 256	612	643	1	1 217	603	614
30–44 года	727	356	369	2	695	338	357
45–59 лет	360	194	166	0	341	182	159
60–74 года	116	54	59	3	114	53	61
75 лет и старше	34	12	22	0	37	14	23
Не зарегистрировано	15	6	7	2			

ответ в отношении лица неуказанного пола об общем числе родившихся детей. Альтернативная стратегия вменения может заключаться в том, чтобы учитывать одну или более из этих других переменных величин.

42. Национальное бюро переписи/статистическое управление может еще в качестве альтернативы взять за основу вмененные значения по возрастному распределению. Для выборки населения, показанной в таблице 1, незарегистрированный возраст отмечается в общей сложности в пятнадцати случаях. Эти данные можно также распределить в тех же пропорциях, что и известные величины, то есть опять в соответствии с логической стратегией вменения. При этом группа редактирования может, вероятно, получить более точные результаты, учитывая другие переменные и их комбинации, например относительный возраст мужа и жены, родителя и ребенка или бабушки и бабушки, внука или внучки или наличие детей школьного возраста, пенсионеров и лиц из числа экономически активного населения.

43. В таблице 1 отредактированные данные справа более достоверны, поскольку неизвестные величины изъяты (см. колонки в графе «отредактированные данные»). В этой части таблицы нет неизвестных величин, поскольку программа распределяет их по другим ответам. Тем не менее многие демографы и другие специалисты по содержательной части переписи по традиции хотят, чтобы неизвестные величины указывались в таблицах, как это сделано в графе «неотредактированные данные» в таблице 1. Они считают, что такая процедура позволяет им производить различные виды оценок по показателям, чтобы определить эффективность процедур переписи или оказать содействие в планировании дальнейших переписей и обследований. Обе задачи вполне выполнимы — составить отредактированную таблицу для реальных пользователей и неотредактированную таблицу для оценки, включая в них неизвестные величины и обходясь без них.

44. Статистические управления должны предпринимать максимальные усилия к сохранению изначально собранных данных. Полный набор введенных с клавиатуры данных следует поместить в архив не только в качестве справочного материала по истории вопроса, но и в качестве ссылки, в случае если будет

принято решение о повторном редактировании с самого начала любой части набора данных. При этом, однако, первоначальные величины по важнейшим позициям данных, таким как возраст, пол и рождаемость, следует сохранять в какой-либо части каждого файла, для того чтобы позволить демографам и другим специалистам проводить анализ результатов редактирования.

45. Еще одна проблема с использованием неизвестных величин в публикуемых таблицах заключается в том, что неизвестные величины могут влиять на результаты анализа тенденций. Новая технология позволяет намного упростить такой анализ по сравнению с принятой практикой. Например, в таблице 2 показано распределение по возрасту из двух проводимых одна за другой переписей. Количество неизвестных величин снизилось в отношении этой небольшой страны с 217, или около 6,5 процента зарегистрированных ответов в 2000 году, до лишь 15, или менее 1 процента ответов в 2010 году.

Таблица 2

Население и изменения в его составе по возрастным группам с интервалом 15 лет с неизвестными величинами: 2000 и 2010 годы

Возрастная группа	Число		Изменение в числовом выражении	Изменение в процентном выражении	Проценты	
	2010 год	2000 год			2010 год	2000 год
Всего	4 147	3 319	828	24,9	100,0	100,0
Младше 15 лет	1 639	1 348	291	21,6	39,5	40,6
15–29 лет	1 256	902	354	39,2	30,3	27,2
30–44 года	727	538	189	35,1	17,5	16,2
45–59 лет	360	200	160	80,0	8,7	6,0
60–74 года	116	89	27	30,3	2,8	2,7
75 лет и старше	34	25	9	36,0	0,8	0,8
Не зарегистрировано	15	217	–202	–93,1	0,4	6,5

46. В этом случае национальное бюро переписи/статистическое управление должно учитывать, как несогласующееся число неизвестных величин влияет на отдельную перепись и на изменения в период между переписями. Например, при показателе 6,5 процента неизвестных величин в переписи 2000 года трудно сопоставить изменения в процентном распределении для возрастных групп с интервалом 15 лет по двум переписям. За десятилетие процент лиц в возрасте от 15 до 29 лет по показателям увеличился с 27 всего лишь до 30 процентов, но распределенные неизвестные величины могут изменить результаты анализа.

47. В пересмотренной таблице 3 показаны неизвестные величины, распределенные или пропорционально, или путем применения того или иного метода вменения. Здесь намного проще увидеть изменения как в числовом, так и в процентном отношении, а также распределение возрастных групп в двух переписях. Несомненно, для того чтобы получить точные, надежные результаты группы редактирования должны удостовериться в том, что отредактированные данные согласуются между двумя переписями и/или обследованиями, а также внутренне совместимы. Строка «не зарегистрировано» опущена.

Таблица 3

Население и изменения в его составе по возрастным группам с интервалом 15 лет без неизвестных данных: 2000 и 2010 годы

Возрастная группа	Число		Изменение в числовом выражении	Изменение в процентном выражении	Проценты	
	2010 год	2000 год			2010 год	2000 год
Всего	4 147	3 319	828	24,9	100,0	100,0
Младше 15 лет	1 743	1 408	335	23,8	42,0	42,4
15–29 лет	1 217	952	265	27,8	29,3	28,7
30–44 года	695	578	117	20,2	16,8	17,4
45–59 лет	341	230	111	48,3	8,2	6,9
60–74 года	114	109	5	4,6	2,7	3,3
75 лет и старше	37	42	–5	–11,9	0,9	1,3

D. Основы редактирования

48. Редактирование — это систематическая проверка недостоверных и несогласующихся ответов и последующая ручная или автоматизированная корректировка (с использованием «неизвестных данных» или динамического вменения) в соответствии с заранее установленными правилами. Некоторые операции редактирования включают ручную корректировку, что означает корректировку вручную сотрудниками в учреждении. Другие операции редактирования заключаются в электронном редактировании с использованием компьютеров. В публикациях результатов переписи может содержаться какое-то количество незначительной информации в случае, если национальные бюро переписи/статистические управления не редактируют результаты переписи или обследования. Редактирование снижает количество искаженных оценок, облегчает обработку данных и повышает доверие пользователя. Кроме того, как отмечали Пуллам, Харфам и Ожевер (Pullum, Harpham and Ozsever, 1986), «главная задача успешного редактирования или чистки данных состоит в том, чтобы в первую очередь определить, согласуются ли различные ответы один с другим и с основным форматом инструментария обследования».

49. Файлы первичных данных переписи содержат самые разнообразные ошибки. При обработке данных ошибки подразделяются на две категории. Ошибки, которые могут заблокировать дальнейшую обработку данных, и ошибки, которые дают недостоверные или несогласующиеся результаты без нарушения логической последовательности дальнейших операций по обработке данных. Как отмечается во втором пересмотренном варианте *Принципов и рекомендаций в отношении переписей населения и жилого фонда* (United Nations, 2008, para. 1.311), исправляться должны все ошибки первого рода, а также максимальное количество ошибок второго рода. Таким образом, основная цель редактирования результатов переписи на этапе обработки заключается в том, чтобы выявить как можно больше ошибок и внести изменения в набор данных, с тем чтобы позиции данных были достоверными и согласовывались между собой. Однако в процессе обработки нельзя исправить все ошибки переписи, в том числе ответы на вопросы переписного листа, которые представляются внутренне согласованными, но фактически являются примерами неверного ответа со стороны респондентов или неправильной регистрации со стороны счетчиков.

50. Редактирование, как правило, разделяется на две категории: 1) фатальное редактирование, при котором ошибки выявляются с уверенностью, и 2) редактирование под вопросом, которое указывает на сомнительные элементы данных (Granquist and Kovar, 1997, p. 420). При фатальном редактировании выявляются элементы данных, которые безусловно содержат ошибки, а при редактировании под вопросом отмечаются данные, которые могут быть недостоверными или несогласующимися. К фатальным ошибкам, то есть ошибкам, выявляемым при фатальном редактировании, относятся недостоверные или отсутствующие записи, а также ошибки, обусловленные несогласованностью. И наоборот, при редактировании под вопросом выявляются элементы данных, которые не подпадают под преимущественно субъективные пределы редактирования, элементы данных относительно высокие или низкие по своему значению, если сравнивать их с другими данными в том же вопроснике, и другие сомнительные записи. Для того чтобы поддерживать доверие к переписи, особенно когда национальное бюро переписи/статистическое управление принимает решение о распространении микроданных, процесс редактирования должен быть направлен на выявление и обработку фатальных ошибок. Сомнительные ошибки исправлять труднее, при этом преимуществ меньше, чем при выявлении и исправлении фатальных ошибок, а стоимость всего процесса возрастает.

51. Поскольку все позиции данных включаются в перепись исключительно потому, что они нужны планирующим и директивным органам, то во время редактирования и вменения данных переписи необходимо исправлять больше сомнительных ошибок, чем при обследовании. Тем не менее, определяя окончательные результаты редактирования данных переписи, сотрудники, занимающиеся содержательной частью переписи, должны тщательно изучить приемы редактирования, разработанные для экспериментальных переписей, и методы, выработанные во время обработки данных, с тем чтобы убедиться в том, что отдельные приемы редактирования соответствуют ожидаемым затратам и результатам. Такое исследование должно проводиться в рамках оценки результатов переписи. Как отмечают Гранкист и Ковар (Granquist and Kovar, 1997, p. 422), данные «по коэффициентам совпадений, которые входят в число признаков, ведущих к изменениям в первоначальных данных, редко сообщаются при оценке или изучении процессов редактирования».

52. Еще один набор методов и терминологии относится к микроредактированию и к макроредактированию. Как уже отмечалось, при редактировании данных переписи и обследования выявляются ошибки в записях данных и между ними. В настоящем *Руководстве* описывается процесс микроредактирования, который направлен на то, чтобы обеспечить достоверность и согласованность записей отдельных данных и взаимосвязей между записями в том или ином домохозяйстве. Еще один метод, макроредактирование, заключается в том, чтобы проверять агрегированные данные, с тем чтобы также убедиться в их обоснованности. В рамках этого метода таблицы составляются из отредактированных данных и проверяются на соответствие по прогнозируемым показателям частоты и допуска в целях выявления различных проблем с данными; при обнаружении «ошибок» с помощью макроредактирования можно внести масштабные изменения в агрегированные данные, направить ту или иную единичную запись данных на повторную обработку или вновь провести микроредактирование для устранения возникшей проблемы. Например, в стране может быть очень большой процент лиц с незарегистрированным возрастом. После

вменения данных о возрасте, для того чтобы получить полный набор данных, проверки на макро- или агрегированном уровне могут дать уверенность в том, что выборочные данные, не сообщенные лицами старшего возраста, не искажают вмененные величины. Группа редактирования в зависимости от результатов анализа может предпочесть принять меры к тому, чтобы снизить риск потенциального искажения. Как макро-, так и микроредактирование требует тщательного тестирования перед применением.

53. Как отмечалось выше, при редактировании следует в максимальной степени сохранять первоначальные данные. Группа редактирования должна располагать достоверными данными высокого качества, но при этом сохранять те данные, которые организация собирает на местах. Первоначальные данные необходимо сохранять на всех этапах компьютерной обработки на тот случай, если группа редактирования примет решение о необходимости пересмотра самого процесса редактирования. Иногда возникает необходимость возврата к первоначальным данным, если группа обнаружит, что в процессе редактирования появилась систематическая ошибка. Иногда пересмотр обусловлен тем, что часть показателей в наборе данных отсутствует или дублируется и тогда этот набор данных приходится пересматривать и вновь редактировать.

54. Иногда источник ошибок находится за пределами того учреждения, где ведется обработка данных. Банистер (Banister, 1980, p. 2) отмечает, что «если известно, что большая доля какой-либо подгруппы лиц не ответила на конкретный вопрос переписи, это означает, что эти лица или не поняли вопроса, или не хотят отвечать на него и сотрудничать в проведении переписи». При этом автор утверждает, что коэффициент неполучения ответов для подгрупп следует записывать на носители данных переписи и в публикуемые таблицы. Сейчас национальные бюро переписи/статистические управления больше предпочитают сохранять эти данные для исследователей на компакт-дисках или других носителях.

55. Появляется все больше подтверждений тому, что никакое компьютерное редактирование не может заменить собой сбор данных переписей более высокого качества. Национальные бюро переписи/статистические управления знают, что в какой-то момент автоматизированное редактирование не только ограничено, но и приводит к обратным результатам: редактирование больше добавляет ошибок в набор данных, чем исправляет их. Замена элемента переписи — это не то же самое, что его корректировка. В связи с этим группа редактирования должна совместно определить начало, середину и конец процесса редактирования.

56. Редактирование и вменение могут улучшить, а могут и не улучшить качество данных, но достоверный набор данных намного облегчает проведение анализа. Процесс начинается с составления переписного листа. Демографы и другие специалисты, занимающиеся содержательной частью переписи, как правило, определяют, что нужно включать в него, зачастую консультируясь с группами пользователей. В конечном счете данные переписи выпускаются «в первую очередь не конкретно для демографов, а для намного более широкой аудитории ученых, политических деятелей и простых людей» (Banister, 1980, p. 17). Однако улучшение результатов переписи без недостоверных и несогласующихся данных необходимо, если на карту поставлено доверие к переписи и репутация национального бюро переписи/статистического управления. Как

отмечает Банистер, «занимающиеся переписью организации могут ссылаться на примеры журналистов, пишущих юмористические статьи, или граждан, с негодованием пишущих официальным лицам, занимающимся переписью, об опубликованных таблицах, где указаны трехлетние дедушки и пассажиры, пользующиеся несуществующими поездами».

57. Проблема заключается в том, чтобы определить, насколько далеко следует заходить, чтобы получить набор данных хорошего качества. Как отмечалось выше, появление компьютеров, сначала больших компьютеров, а затем микрокомпьютеров, позволило практически полностью автоматизировать процесс редактирования. Во многих национальных бюро переписи/статистических управлениях специалисты по содержательной части переписи стали, по существу, энтузиастами редактирования. В связи с этим сейчас бюро проводят много проверок на согласованность, что раньше трудно было сделать, в частности перекрестные проверки записей и перекрестные проверки домохозяйств. К сожалению, это свойство микрокомпьютеров вызвало и множество проблем, самой большой из которых является чрезмерное редактирование.

1. Каким образом чрезмерное редактирование оказывает пагубное воздействие

58. Чрезмерное редактирование оказывает отрицательное воздействие на весь процесс редактирования несколькими способами, в том числе удлинняя сроки, повышая затраты и искажая истинные значения. Оно также дает ложное чувство уверенности в качестве данных. Обо всем этом говорится ниже.

a) Своевременность

59. Чем больше национальное бюро переписи/статистическое управление занимается редактированием, тем больше времени занимает весь процесс. Главный вопрос заключается в том, чтобы определить, сколько добавленное время прибавляет к ценности продукта переписи. Каждая группа редактирования должна оценивать на постоянной основе и постфактум чистую выгоду от дополнительного времени и ресурсов, затраченных на весь продукт переписи. Зачастую выгоды с точки зрения затраченного времени настолько ничтожны, что лучше иметь небольшие проблемы с данными, чем лишать основных пользователей возможности получать своевременную информацию.

b) Финансы

60. Аналогичным образом, с увеличением времени растут и затраты на проведение переписи. Каждое национальное бюро переписи/статистическое управление должно решить, если оно увеличивает объем и сложность редактирования своих данных, стоит ли увеличение затрат таких дополнительных усилий и может ли оно себе позволить эти дополнительные затраты.

c) Искажение истинных значений

61. Хотя в целом процесс редактирования направлен на повышение качества данных, увеличение объема и сложности редактирования может повлечь за собой и отрицательные последствия. Иногда группы редактирования ошибочно меняют элементы данных по самым разным причинам: из-за недоста-

точного общения между специалистами по обработке данных и специалистами по содержательной части переписи; ошибок в очень сложной и насыщенной программе; или из-за многократной обработки того или иного элемента переписи во время редактирования. Национальные бюро переписи/статистические управления намерены всячески избегать такого рода проблем. Например, Гранквист и Ковар (Granquist and Kovar, 1997) указывают на то, что вменение возраста мужа или жены по установленной возрастной разнице между ними может быть полезным, но может искусственно исказить данные, если таких случаев много.

d) Ложное чувство уверенности

62. Чрезмерное редактирование дает сотрудникам национального бюро переписи/статистического управления и другим пользователям ложное чувство уверенности, особенно когда бюро не осуществляют и не документируют меры по обеспечению гарантии качества. Кроме того, в таблицах с данными переписи случайные результаты появляются независимо от того, какой объем работы выполняет группа редактирования, поэтому необходимо предупреждать пользователей о том, что могут возникать небольшие ошибки. Особенно это актуально сейчас, когда многие страны публикуют выборочные микроданные. Национальные бюро переписи/статистические управления в принципе не имеют желания публиковать данные, наносящие ущерб процессу планирования, поэтому необходимо тщательно следить за тем, чтобы все наиболее важные переменные редактировались надлежащим образом и могли быть использованы при планировании. Например, ни одно национальное бюро переписи/статистическое управление не хотело бы публиковать микроданные или таблицы с неизвестными показателями пола или возраста. Вместе с тем такими переменными, как нетрудоспособность или грамотность, можно оперировать, особо не подвергая их редактированию. Хотя в сводных таблицах могут появляться некоторые несоответствия, поскольку национальные бюро переписи/статистические управления не могут отредактировать все пары переменных, группы редактирования должны проверять наиболее важные комбинации. Если при редактировании обнаруживаются несогласующиеся величины, то группой редактирования должны быть предусмотрены процедуры корректирования.

2. Обработка неизвестных величин

63. Группа редактирования на самом раннем этапе планирования переписи должна решить, как обращаться с «неуказанными» или неизвестными величинами. Как отмечалось выше, колонки или ряды неизвестных величин в таблицах не несут никакой информации и, значит, бесполезны, поэтому планирующие органы в большинстве стран предпочитают, чтобы эти данные вменялись. Без обработки неизвестных величин многие пользователи распределяют их в выпускаемых таблицах в том же соотношении, что и известные данные, таким образом исчисляя неизвестные величины постфактум. Кроме того, некоторые информативные, но недостоверные величины нельзя рассматривать как пропуски — их необходимо исправлять. Например, запись «Южная Америка» в качестве места рождения должна исправляться на конкретную страну (например, Перу). Группа редактирования должна решить, по какой системе обрабатывать неизвестные величины.

3. Кажущиеся изменения

64. Национальные бюро переписи/статистические управления, даже если имеют такую возможность, как правило, не работают с моделями, если они разрабатывают собственные правила редактирования. Группы редактирования должны разрабатывать правила, которые соответствовали бы фактическим характеристикам населения или жилищ. Все данные должны редактироваться по правилам. Например, свод правил может предусматривать, чтобы ребенок главы домохозяйства был по меньшей мере на 15 лет младше главы. Однако ребенок главы домохозяйства может быть фактически социальным, а не биологическим ребенком: он или она может быть биологическим ребенком супруга, который при этом не является главой домохозяйства. В результате разница в возрасте может быть меньше 15 лет. Поскольку планирующие органы в большинстве стран не проводят различий между детьми и пасынками или падчерицами, то при вышеупомянутых обстоятельствах, если по правилам редактирования меняется возраст ребенка, могут возникать несоответствия в уровне образования, показателях участия населения в экономической деятельности и других областях. С учетом этого такое правило следует протестировать, прежде чем применять его в полном масштабе, чтобы посмотреть, какие будут результаты.

4. Определение уровня допустимых отклонений

65. Группа редактирования должна разработать «уровни допустимых отклонений» для каждого элемента, а иногда для сочетания элементов данных. Уровни допустимых отклонений показывают допустимое количество недостоверных и несогласующихся ответов, прежде чем группы редактирования приступят к устранению ошибок. Например, в отношении большинства позиций переписи какой-то небольшой процент респондентов не даст «приемлемых» ответов по той или иной причине. Для некоторых элементов, например таких как возраст и пол, которые используются в сочетании со многими другими элементами в целях планирования, уровень допустимых отклонений может быть весьма низким. В случае низкой доли недостающих или несогласующихся ответов (менее одного или двух процентов) любые приемлемые правила редактирования вряд ли повлияют на использование данных. Если эта доля высокая (5–10 процентов или более, в зависимости от ситуации) в простых или даже в сложных случаях, вменение может исказить результаты переписи.

66. Чтобы сократить до минимума количество недостающих ответов, национальные бюро переписи/статистические управления должны добиваться того, чтобы проводящие перепись сотрудники прилагали все усилия для получения информации на месте. Если та или иная страна решит, что ей не нужна столь высокая точность в отношении некоторых элементов, таких как грамотность или нетрудоспособность, то уровень допустимых отклонений для таких элементов может быть значительно выше. Иногда группы редактирования могут корректировать элементы, которые содержат слишком много ошибок, вновь направляя счетчиков на места, проводя повторный опрос по телефону или используя свои знания в той или иной области. Однако зачастую повторное направление счетчиков на места или проведение каких-либо других последующих операций обходится слишком дорого и национальные бюро переписи/статистические управления могут принять решение либо не использовать дан-

ный элемент, либо использовать его, сопровождая предупредительными примечаниями.

67. При этом возникает вопрос, кто должен определять уровень допустимых отклонений для того или иного элемента данных. Установить уровень допустимых отклонений может команда редактирования, включая специалистов по содержательной части переписи и специалистов по обработке данных. Сотрудники, отвечающие за содержательную часть переписи, должны использовать элементы данных в течение продолжительного времени, и поэтому они профессионально заинтересованы в том, чтобы получать данные самого высокого качества. Однако специалисты по обработке данных могут признать, что они фактически не в состоянии разработать соответствующие программы редактирования, чтобы сократить уровень допустимых отклонений до приемлемых размеров, или что сами данные могут препятствовать успешному применению программы с учетом уровня допустимых отклонений.

5. Извлечение уроков из процесса редактирования

68. По мере редактирования данных переписи необходимо проводить подробный анализ положительной и отрицательной обратной связи в целях повышения качества результатов проводимой переписи или обследования с перспективой для дальнейших переписей и обследований. Группа редактирования должна постоянно работать над тем, чтобы определять, что работает должным образом, а что не работает. Она также должна определить, можно ли улучшить и рационализировать те аспекты данного процесса, которые работают должным образом, с тем чтобы эти данные доходили до пользователей еще быстрее. Чем раньше в ходе проведения переписи национальные бюро переписи/статистические управления обнаружат ошибки, тем больше вероятность их исправления.

6. Гарантия качества

69. Гарантия качества играет важную роль во всех операциях переписи. Исходя из этого, несомненно, должны действовать официальные механизмы обеспечения качества, с тем чтобы наблюдать за ходом компьютерного редактирования и этапом вменения. Для анализа качества редактирования и скорости обработки крайне важное значение имеют контрольные журналы, критерии качества и диагностическая статистика (Granquist and Kovar, 1997; Statistics Canada, 1998).

7. Затраты на редактирование

70. Данное *Руководство* может помочь странам сократить большие затраты времени и ресурсов, необходимых для редактирования и вменения данных переписи или обследования. Как отмечают Гранквист и Ковар (Granquist and Kovar 1997, p. 418), даже «в 1990-х годах редактирование, по существу, обходится так же дорого, как и в 1970-х годах, несмотря на то что данный процесс в значительной степени рационализирован благодаря постоянному использованию технологических достижений». В большинстве стран редактирование отнимает несоразмерно много времени и средств, поэтому каждая страна должна определить доходность инвестирования в эту работу. По мнению этих же авторов,

затраты на редактирование данных обследований домашних хозяйств в начале 1990-х годов составляли около 20 процентов всех бюджетных средств в мире, выделяемых на проведение переписей.

71. Чрезмерное редактирование может задержать получение результатов переписи. Хотя сотрудники, занимающиеся у себя в стране переписями/обследованиями, могут иметь лишь эпизодический опыт подобного рода применительно к переписям, исследование Пуллама, Харфама и Ожевера (Pullum, Harpham and Ozsever, 1986) показало, что в результате машинного редактирования результатов Всемирного обследования фертильности публикация результатов задержалась почти на один год. Национальные бюро переписи/статистические управления могли бы лучше распоряжаться своими средствами, чтобы во время переписи или обследования проводить прежде всего регистрацию с более высоким качеством.

8. Вменение

72. Вменение (условное исчисление) — это процесс разрешения проблем, касающихся недостающих, недостоверных или несогласующихся ответов, выявленных во время редактирования. Вменение проводится путем замены одного или более ответов или недостающих величин в записи или нескольких записях, которые редактируются, с тем чтобы получить достоверные, логически согласованные учетные данные. Контакт с респондентом или изучение вопросника вручную устраняет некоторые проблемы в самом начале процесса. Однако решить все проблемы на этих ранних этапах, как правило, невозможно из-за большого количества ответов, значительных затрат и установленных сроков. Таким образом, путем вменения обрабатывается то, что не удалось сделать при редактировании, поскольку желательно получить полный и согласованный файл с условно начисленными данными. Лучше всего вменение проводят те члены группы, которые имеют полный доступ к микроданным и владеют хорошей дополнительной информацией.

- a) Вмененная запись должна быть очень похожа на запись, которую не удалось отредактировать. Лучше всего осуществлять вменение минимального числа переменных, сохраняя, таким образом, как можно больше данных, представленных респондентом. При этом исходят из предположения (которое на практике не всегда верно) о том, что у респондента больше вероятность сделать всего одну или две ошибки, а не несколько.
- b) Вмененная запись должна согласовываться со всеми отредактированными данными.
- c) Группы редактирования должны помечать вмененные величины и четко указывать методы и источники вменения.
- d) Группа редактирования должна сохранять невмененные и вмененные величины в полях записи, с тем чтобы можно было оценить степень и последствия вменения.

9. Архивирование

73. Процедура обеспечения качества данных переписи или обследования заключается, в частности, в том, чтобы документировать все операции, а

затем эту документацию архивировать. Национальные бюро переписи/статистические управления должны сохранять как файлы с отредактированными данными, так и файлы с неотредактированными данными для последующего анализа. Некоторые процедуры, например многие виды сканирования, автоматически сохраняют первоначальное отображение. Аналогичным образом, сразу после ввода с клавиатуры пакетов данные следует увязать и сохранить для возможного анализа. Однако при любой процедуре важно архивировать оригинальные копии неотредактированных файлов. На практике копии неотредактированных данных должны храниться в нескольких местах в пределах статистического управления, а также в других частях страны и за ее пределами.

74. Документация должна быть достаточно полной, для того чтобы органы, планирующие проведение переписи или обследования, могли восстановить те же процессы позднее, с тем чтобы обеспечить сопоставимость с проводимой переписью или обследованием. И процессы, и результаты должны быть воспроизводимы. И наконец, неотредактированные, так же как и отредактированные данные необходимо хранить в нескольких местах, приняв надлежащие меры по обеспечению их постоянной доступности с течением времени.

75. Как отмечено в других разделах, часть документации включает два типа отчетов по редактированию. В первом отчете приводится краткая статистика с указанием числа и процентной доли ошибок (исходя из надлежащих критериев, таких как общее число жилищных единиц, общей численности населения, населения трудоспособного возраста, взрослых женщин и т. д.). Второй отчет содержит по крайней мере образец «профильной» структуры неотредактированных записей по домохозяйствам или по жилищам, список ошибок и методы их устранения в отношении жилищной единицы или лиц, проживающих в данной жилищной единице, а также отредактированные данные по жилищной единице или домохозяйству.

76. Необходимо предоставить два набора ошибок на логических территориальных уровнях и, конечно, для основных административных единиц; с другой стороны, предоставление списка ошибок на более низких территориальных уровнях могут помочь в выявлении проблем с подготовкой счетчиков, контролем качества и других проблем, связанных с регистрацией.

Глава II

Прикладные программы редактирования

77. В настоящей главе дается общий обзор прикладных программ для процесса редактирования и вменения данных. В ней рассматривается схема общего потока редактирования данных переписей или обследований — от необработанных данных, которые сканируются или вводятся с клавиатуры, через редактирование структуры и содержания данных вплоть до получения отредактированного набора данных⁴. Были отобраны примеры, показывающие, какого рода проблемы могут возникнуть у пользователя в связи с необработанными данными и почему отредактированные данные более полезны. Здесь рассматриваются вопросы введения данных с клавиатуры и кодирования в качестве предварительного этапа процесса редактирования. В данной главе также представлены общие вопросы компьютерного редактирования наряду с руководящими принципами по таким вопросам, как проверка на достоверность и согласованность. Подробно рассматриваются два характерных для компьютерного редактирования способа — статическое вменение («cold deck») и динамическое вменение («hot deck»).

78. Вне зависимости от того, вводятся ли данные переписи путем сканирования или с клавиатуры, сохраняется некая общая схема потока данных. Редакционная группа переписи начинает работу с неотредактированных данных. В большинстве случаев все данные заранее закодированы счетчиком переписи или сотрудником бюро переписи, следовательно, такой набор данных готов для структурного кодирования. В некоторых случаях для проведения редактирования требуется операция для перевода отсканированных данных в иной машиночитаемый формат в зависимости от предполагаемого к использованию текста редактирования. В других случаях отсканированные данные потребуют также повторного автоматизированного кодирования для заполнения таких ячеек данных, как место рождения, отрасль и род деятельности.

79. В любом случае неотредактированные данные должны находиться в формате, который позволяет компьютерным программам производить структурное редактирование (которое более подробно описано в главе III). Структурное редактирование предусматривает проверку в целях обеспечения, чтобы все крупные административно-территориальные единицы были представлены в географической или цифровой последовательности и чтобы в рамках каждой крупной административно-территориальной единицы была обозначена каждая мелкая административная единица, причем также в географической или цифровой последовательности. Затем в рамках каждой мелкой административной единицы должен быть обозначен каждый населенный пункт, также в географической или цифровой последовательности. Эта процедура продолжается вплоть до самого низкого территориального уровня. Как указано в следующей главе, необходимо проводить надлежащие процедуры,

⁴ При составлении первоначального издания данного *Руководства* для переписей 2000 года практически во всех странах данные вводились с клавиатуры. Сегодня в большинстве стран данные сканируются, иногда с последующим вводом с клавиатуры дополнительных данных. В настоящем издании *Руководства* делается попытка учесть сканирование при редактировании структуры и содержательной части данных. Даже в период подготовки настоящего издания *Руководства* появились новые технологии, включающие применение как карманных персональных компьютеров (КПК), так и Интернета для сбора и интерактивного редактирования данных (см., например, шведскую программу Ireback (2000) для использования Интернета). Так же как многие развивающиеся в технологическом отношении страны испытывали проблемы со сканированием в начале 2000-х годов, в настоящее время многие страны обнаружили, что использование КПК все еще требует дополнительного изучения.

чтобы обеспечить включение в набор данных каждой жилищной единицы один-единственный раз.

80. Редактирование структуры данных должно также обеспечивать присутствие при необходимости всех надлежащих видов записей и отсутствие повторения видов записей, когда это не требуется. Следовательно, при проведении переписей населения и жилого фонда сначала необходимо включать записи данных о населении и жилом фонде, и этот порядок должен соблюдаться по всему набору данных. В большинстве случаев должны присутствовать только записи по жилому фонду, с тем чтобы избавиться от лишних записей; программисты должны также представлять записи по жилому фонду в отношении домохозяйств, по которым отсутствуют такие записи. Аналогичным образом, демографические записи должны присутствовать по заселенным жилым единицам (которые обычно определяются в качестве таковых в соответствующей записи по жилой единице) и отсутствовать по незаселенным единицам.

81. Применительно к структуре данных важно отметить, что неминуемо возникнет необходимость возврата к структурному редактированию в ходе редактирования содержательной части данных, а подчас и после него, поскольку ошибки появляются в ходе различных процедур переписи. Это нормальный и ожидаемый процесс переписи; следовательно, соответствующие требования по затратам времени, персоналу и оборудованию должны быть учтены в рамках общей системы.

82. Затем можно начинать редактирование содержательной части данных. Каждая позиция данных по населению и жилому фонду должна рассматриваться по отдельности, а также обычно в комбинации с другими элементами данных для определения достоверности каждого элемента и оптимального соотношения между элементами. Главы IV и V посвящены различным элементам, относящимся к населению и жилому фонду, как описано во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда Организации Объединенных Наций*.

83. После завершения редактирования содержательной части необходимо сформировать полностью отредактированный набор данных. Неотредактированные данные должны храниться в нескольких надежных местах, при этом важные неотредактированные элементы данных (или все неотредактированные элементы) должны также указываться в конце различных видов записей. И вновь следует подчеркнуть, что по мере составления таблиц также может возникнуть необходимость повторного редактирования содержания данных для решения любых специфических проблем, которые возникают в результате составления конкретных перекрестных таблиц.

84. Цель редактирования данных переписей и обследований заключается в обнаружении пропусков и несоответствий в записях данных, а для их устранения применяется метод вменения. Для обработки пропусков и различных видов неприемлемых записей вводимых данных применяются конкретные процедуры редактирования. В процессе вменения заменяются недостоверные записи и устраняются несоответствия, обнаруженные в наборе данных. Результатом этой работы является отредактированный файл микроданных для составления таблиц с указанием приемлемых и согласующихся величин по всем применяемым элементам данных для каждой зарегистрированной жилищной единицы или лица.

85. Опять же необходимо отметить, что никакое редактирование не может заменить собой высококачественную регистрацию. Процесс редактирования хорошо работает, когда метод вменения применяется по отношению к произвольным пропускам и несоответствиям. Однако, если во время сбора данных возникают систематические ошибки, качество данных нельзя повысить путем редактирования, какими бы совершенными ни были соответствующие процедуры. С точки зрения качества получаемых данных основополагающее значение имеет выбор тем для исследования. Во время опроса респонденты должны проявлять желание и иметь возможность предоставить адекватную и нужную информацию. Таким образом, может возникнуть необходимость избегать тех тем, которые могут вызвать опасения, пробудить какие-либо местные предрассудки или суеверия, а также вопросов, слишком сложных и трудных, для того чтобы средний респондент мог легко ответить на них в контексте переписи населения. Точная формулировка каждого вопроса, которая необходима для получения наиболее достоверного ответа, неизбежно будет зависеть от национальных условий и должна быть проверена до проведения переписи. В связи с этим крайне важно, чтобы национальные бюро переписи/статистические управления выделяли достаточно ресурсов для получения данных переписи самого высокого качества.

86. Для проведения этапа компьютерного редактирования группа редактирования готовит письменные инструкции или спецификации по редактированию, таблицы решений, графические схемы и символический код. Символический код — это набор письменных инструкций или спецификаций по редактированию, как показано на рисунке 8.

87. Графические схемы помогают специалистам по содержательной части переписи понимать различные взаимосвязи между переменными и облегчают написание инструкций по редактированию. Выборочные графические схемы приводятся в приложении IV. Специалисты по содержательной части переписи готовят инструкции по редактированию совместно со специалистами по компьютерам, описывая, что нужно делать по каждому элементу данных. Инструкции по редактированию должны быть ясными, лаконичными и четко выраженными, поскольку они служат основой для всего пакета программ редактирования.

88. Вся группа редактирования данных переписи — и специалисты по содержательной части, и обработчики данных — должна обладать обширным опытом в вопросах обработки и анализа демографических данных. Неквалифицированный персонал может непреднамеренно внести дополнительные ошибки и погрешности в результаты переписи.

А. Вопросы кодирования

89. Как отмечалось выше, в течение большей части второй половины двадцатого века данные во всех странах вводились с помощью клавиатуры. Хотя в настоящее время в большинстве стран данные при проведении переписи сканируются, однако зачастую данные обследований по-прежнему вводятся с клавиатуры. Даже при сканировании бланков анкет, определенные переменные все еще необходимо переводить из текстовой в цифровую форму. Процесс составления машиночитаемых цифровых и буквенно-цифровых данных называется кодированием.

90. Некоторые пакеты программ редактирования легко воспринимают буквенно-цифровые данные и обрабатывают их. При этом, однако, большинство таких пакетов программ испытывают определенные сложности в разбивке данных на категории, их суммировании, получении процентных показателей, медианных значений и т. д. при включении в них неких текстовых данных.

91. Необходимо по возможности избегать кодов, которые полностью состоят из букв или комбинации букв и цифр (так называемые буквенно-цифровые коды). При сканировании заполненных формуляров буквенно-цифровые данные не представляют серьезной проблемы; с другой стороны, в отношении многих компьютерных пакетов программ использование таких данных требует серьезных манипуляций или по крайней мере повышенного внимания. Многие программы редактирования способны обрабатывать буквенные символы только, если они взяты в кавычки или помечены каким-либо иным образом.

92. При разработке схемы кодирования сотрудники по проведению переписи или обследования должны рассмотреть преимущества каждой схемы с точки зрения затрат времени, энергии и средств. Проблемы кодирования довольно незначительны для малых стран или небольших исследований, поскольку в этом случае требуется значительно меньшая по объемам обработка данных, чем для переписи. Кроме того, сканируемые данные практически не подвержены проблемам, связанным с дополнительными колонками информации.

93. С другой стороны, когда в рамках переписи или обследования используются, например, две колонки данных для элемента «родство» вместо одной, сканирование даст ошибки, которые отсутствовали бы при единственной колонке; иными словами, в присутствии кодов с 1 по 9 сканер может воспринять их как буквенный символ, или пропуск, или пометку, превращенную в некий читаемый символ. Тем не менее такие проблемы легко решаются в процессе редактирования, как описывается далее в настоящей публикации.

94. Однако при наличии двух колонок, скажем, с кодами от 1 до 10, существует риск внесения новой разновидности ошибок. Вместо истинных значений от 1 до 9 в этом случае могут получиться исходные величины от 0 до 99, а также вышеупомянутые буквенные символы, пропуски и пометки. Когда редакторы получают величину 13, они должны принять стратегическое решение о том, что делать дальше с этой величиной. Означает ли она 3 и, таким образом, 1 является ошибкой? Означает ли она 10 и, следовательно, 3 — неправильная величина? В большинстве случаев специалисты по содержательной части предоставляют спецификации по редактированию данного элемента, однако присутствие таких величин автоматически увеличивает время и сложность редактирования и может снизить качество конечного набора данных.

95. Одна из наиболее распространенных проблем, которая в конкретном плане обсуждается ниже в данной публикации, возникает применительно к элементам данных из категории рождаемости. В настоящее время во многих странах ведется сбор информации о детях в составе домохозяйства, детях, проживающих в других местах, и умерших детях, а иногда — по суммарному числу этих детей в разбивке по полу ребенка. Следовательно, в таких странах может присутствовать до 12 элементов таких данных. Проблема состоит в том, сколько знаков можно использовать для каждого из этих элементов. При

использовании двух колонок величины для мальчиков в домохозяйстве могут варьироваться от 0 до 99; при использовании только одной колонки величины могут варьироваться от 0 до 9. Тем не менее, поскольку весьма маловероятно, что какая-либо женщина в домохозяйстве может иметь более 9 мальчиков, использование двух знаков дает высокую вероятность появления помарок или неправильного считывания при сканировании — например считывание цифры 9 вместо 0, что в результате даст суммарное значение, равное 91 ребенку вместо 01.

96. Таким образом, для данных по мальчикам и девочкам, находящимся дома, в других местах или умершим, наиболее оправданным является применение одной колонки. Однако для показателя общего числа детей, находящихся дома, общего числа детей в других местах, общего числа умерших детей и общего числа детей более обоснованным может быть применение двух колонок. Многое зависит от уровней рождаемости в конкретной стране. Эпизодически может встретиться необычное домохозяйство, в котором фактически будет более девяти человек в той или иной категории; однако, как и всегда в ходе проведения переписи, статистическое управление должно придерживаться некоего баланса между ошибками и правильными данными.

97. Для порядковых переменных рекомендуется следующая серия кодов для степени родства:

- 1 Глава домохозяйства (или домохозяин)
- 2 Супруг(а)
- 3 Ребенок
- 4 Брат/сестра
- 5 Родитель
- 6 Внук/внучка
- 7 Другой родственник
- 8 Лицо без родственных связей

Для большинства стран такой набор стандартных кодов охватывает большую часть родственных связей. В некоторых странах для категории главы домохозяйства добавляется код «0», при этом к другим родственникам может добавиться десятая категория.

98. Эти коды могут использоваться для получения данных о составе домохозяйства, как указано в приложении I по производным переменным. Тем не менее во многих странах, особенно в тех, которые переживают эпидемию ВИЧ/СПИДа, существует потребность в более подробной информации, которая может быть получена с помощью таких кодов. В этих странах может потребоваться особая информация по зятям/невесткам, тестям/тетям, свекрам/свекровям, бабушкам/дедушкам, племянникам/племянницам и т.д. В такой ситуации статистическому управлению для выполнения этой задачи потребуются дополнительные двузначные коды.

99. Когда в той или иной стране принимается решение о применении нескольких колонок, необходимо также решить, как их использовать. Приведенный выше пример исходит из того, что коды для указания родства являются последовательными. Однако в случае принятия решения о применении двух

колонок специалисты по содержательной части данного элемента могут пойти по пути присвоения значимости для самих колонок. Например:

- 10 Глава домохозяйства
- 11 Супруг(а)
- 12 Брат/сестра
- 13 Супруг(а) сестры/брата
- 21 Ребенок
- 22 Приемный ребенок
- 23 Пасынок/падчерица
- 24 Племянник/племянница
- 31 Родитель
- 32 Родитель супруга(и)
- 33 Дядя/тетя
- 41 Внук/внучка
- 77 Другой родственник
- 88 Лицо без родственных связей
- 90 Лица, проживающие в общественных учреждениях

100. При такой схеме в первой колонке коды присваиваются по поколениям: 1 — для поколения главы домохозяйства, 2 — на одно поколение младше, 3 — на одно поколение старше, 4 — на два поколения младше и т. д., а затем кодируются виды родственников в рамках каждой из этих категорий. Хотя такие величины могут быть полезными для определения состава семьи, сотрудники управления и широкие пользователи могут счесть их слишком громоздкими.

101. Такой тип кодирования, однако, следует рассматривать применительно к определенным социально-экономическим переменным. Применительно к этнической принадлежности, например, первый знак кода может означать принадлежность к крупному племени или крупной этнической группе, небольшому племени или небольшой этнической группе. При наличии более 10 небольших групп потребуются, конечно, использовать двузначные коды в первой колонке.

102. Аналогичным образом, для элементов данных, включающих три или четыре знака, таких как род деятельности или отрасль, первый знак будет обозначать крупную категорию рода деятельности или отрасли, второй знак — мелкую категорию рода деятельности или отрасли, а третий знак — конкретный род деятельности или отрасль. Поскольку большинство международных схем кодирования уже имеют встроенные уровни кодов, у статистического управления отпадает необходимость осуществления какой-либо дополнительной работы.

103. Когда национальные бюро переписи/статистические управления составляют перечни кодов для программ редактирования и для последующего составления таблиц, они могут по своему усмотрению устанавливать общие коды для некоторых элементов данных. Например, во многих странах коды места (место рождения, место рождения родителей, прежнее место жительства, место работы), языка, этнической/расовой принадлежности и гражданства очень похожи. В связи с этим можно было бы разработать общую схему кодирования «места» в виде трехзначных цифровых кодов, где первый знак означал бы континент, второй — регион и третий — конкретную страну. Национальные бюро переписи/статистические управления могут также использовать

цифровые коды стран, разработанные международными организациями, например статистическим отделом Организации Объединенных Наций (United Nations, 1999). Набор общих кодов для тесно взаимосвязанных переменных может сократить ошибки кодирования и помочь обработчикам данных во время редактирования. Общие коды позволяют также обработчикам данных в необходимых случаях использовать записи из одного элемента для определения значения другого элемента.

104. Структура кодирования может облегчить процесс кодирования, а также дальнейшую обработку данных во время редактирования, составления таблиц и проведения анализа. Для больших стран с большой численностью иммигрантов или этнических групп, коды, составленные по континентам, регионам и странам с разными кодами или цифрами по каждому показателю предпочтительнее простого списка.

105. На рисунке 1 показаны примеры общих кодов для таких элементов данных, как место рождения, гражданство, язык и этническая принадлежность. Для Филиппин коды говорящих на илоканском и тагальском языках отличаются от общего кода филиппинских языков в зависимости от конкретной ситуации в стране и, кроме того, эти коды могут различаться и между собой. В зависимости от конкретной ситуации в той или иной стране эти коды могут также отличаться и друг от друга. Хотя английский язык имеет один код, на нем говорит не одна группа населения. Таким образом, коды для места рождения, гражданства и этнической принадлежности в Канаде и Соединенных Штатах Америки отличаются незначительно. Для лиц, родившихся во Франции, имеющих французское гражданство, говорящих на французском языке и являющихся французами по этнической принадлежности используется один и тот же код. Таким образом, если недостает одного из этих элементов и если группа редактирования сочтет такое решение приемлемым, обработчик данных может взять код в одной из других граф.

106. Если та или иная группа вопросов в переписном листе взаимосвязана, то, вероятно, сотрудники, занимающиеся переписью/обследованием, не должны задавать все эти вопросы. Группа редактирования должна решить в каждом конкретном случае, когда использовать элементы непосредственно для присвоения кода, а когда использовать другие имеющиеся переменные.

Рисунок 1

Примеры общих кодов для отдельных элементов данных

Группа	Место рождения	Гражданство	Язык	Этническая принадлежность
Франция/французский	10	10	10	10
Испания/испанский	20	20	20	20
Латинская Америка	25	25	20	25
Филиппины/филиппино	30	30	30	
илоканский			32	
тагальский			32	
Англия/английский	40	40	40	40
Канада	50	50	40	50
США	52	52	40	52

107. Еще одна проблема возникает, когда определения между отдельными переписями (или между переписью и обследованием) отличаются по таким переменным, как работа или этническая принадлежность. Национальное бюро переписи/статистическое управление должно решить, как учитывать такие изменения и в отношении текущих редактируемых данных проводимой переписи, и в отношении наборов данных из предыдущей переписи, для того чтобы показать тенденции. Если есть исходные неотредактированные данные, обработчики данных могут вносить изменения в соответствующие отредактированные элементы и сделать для всех повторный прогон.

108. Например, какая-либо европейская страна может использовать единый код страны происхождения для всех южноазиатских стран, если отмечено только несколько случаев. Однако в связи с изменениями характера миграции при следующем обследовании или переписи могут потребоваться отдельные коды для Индии, Бангладеш, Пакистана, Шри-Ланки и других южно-азиатских стран на протяжении всего процесса обработки.

В. Ручная и автоматизированная коррекция данных

109. Ручное редактирование данных переписи может занять месяцы или годы, предоставляя множество возможностей для внесения ошибок по вине человека. Ручное редактирование является слабой альтернативой автоматизированному редактированию отчасти потому, что нельзя создать или воссоздать журнал редактирования для ручной коррекции. Компьютерное, или автоматизированное редактирование, сокращает время, необходимое для обработки данных, и количество ошибок по вине человека. Как при компьютерном, так и при ручном редактировании производится проверка достоверности той или иной величины путем поиска приемлемого значения, но с помощью компьютерных программ производится также проверка этого значения по сравнению со смежными величинами с точки зрения их согласованности. И наконец, что наиболее важно, автоматизированное редактирование позволяет создать журнал редактирования, и поэтому его можно воспроизвести, в то время как ручное редактирование не позволяет делать этого.

110. В первые годы внедрения компьютерного ввода данных, редактирование данных в ходе их ввода было невозможно; иными словами, любая коррекция должна была осуществляться вручную либо в рамках офисных процедур кодирования и проверки, либо в рамках компьютерных процедур после ввода данных с клавиатуры. Более новые программные пакеты имеют встроенные функции редактирования, которые обеспечивают блокирование ввода недостоверных записей, если только такой ввод не инициируется самими операторами, и дают возможность видеть отмеченные несоответствия для внесения коррекции операторами вручную или программистом. По мере более широкого распространения сканирования эта эволюция повторилась: в первые годы внедрения сканирования редактирование во время ввода данных было невозможно; однако в последнее время функции редактирования с проверкой на достоверность и пересчета и перекодирования данных встраиваются в системы сканирования.

111. Когда при переписях и обследованиях набираются огромные объемы данных, сотрудники не всегда могут обращаться к исходным документам для

исправления ошибок. Даже если имеются исходные вопросники, записанные в них данные иногда могут быть неправильными или не согласовываться между собой. Система компьютерного редактирования и вменения способна незамедлительно исправить или изменить ошибочные данные и составить отчет по всем обнаруженным ошибкам и всем произведенным изменениям. Компьютерное редактирование следует тщательно планировать, с тем чтобы сэкономить сотрудникам время для других операций по обработке данных. Хотя прогон большого количества данных через компьютерную систему может потребовать много времени, это все же не отнимает столько времени, сколько при ручной коррекции.

112. Ручная коррекция имеет несколько форм. Рассмотрим простой пример ошибки в ответе касательно пола: инспектор проверяет работу счетчика и находит очевидную ошибку, например «мужской» приписан какому-то лицу по имени «Мэри». При изменении пола на «женский» инспектор производит ручное редактирование. Если инспектор не исправляет вопросник, а вместо этого направляет его в местное отделение, то сотрудники этого отделения могут увидеть проблему и вручную исправить ошибку. В центральном отделении во время кодирования кодировщики могут заметить несоответствие между именем и полом и тоже произвести ручную коррекцию. Или же кодировщики могут и не заметить эту проблему, но несогласованность между именем и полом могут заметить операторы, вводя данные вопросников с клавиатуры, и в этом случае они могут произвести ручную коррекцию до ввода данных.

113. Однако если ошибка не замечена, и оператор вводит код для ответа «мужской», тогда можно использовать целый ряд различных процедур. В отношении элементов данных, связанных с полом, например в блоке «рождаемость», программа-редактор может отметить тот факт, что это — мужчина с информацией о рождаемости и сообщить об этом, когда оператор вводит данные. Оператор может после этого посмотреть на вопросник, увидеть, что это действительно женщина, и произвести исправление вручную. Другой вариант состоит в том, что, если национальное бюро переписи/статистическое управление пользуется программой-редактором, независимой от клавиатуры, компьютерная программа может пометить этого человека как мужчину с информацией о рождаемости. Затем, используя географическую информацию, работники управления могут найти в корзине исходный вопросник, достать его и определить, что респондент по имени «Мэри» был ошибочно зарегистрирован не как женщина, а как мужчина. В этом случае сотрудник управления может снова передать информацию оператору, а тот может вывести эту запись и произвести ручную коррекцию.

114. Этот пример показывает как преимущества, так и недостатки ручного редактирования. На любом из вышеуказанных этапов тот или иной сотрудник переписи может обнаружить ошибку — несоответствие между именем и полом — и исправить ее. Однако в национальных бюро переписи/статистических управлениях, использующих ручное редактирование, возможно, есть сотрудники, проверяющие эту взаимосвязь на каждом этапе. На эту работу уходит огромное количество энергии, а результаты, возможно, мало отличаются, особенно в агрегированной форме, от тех, которые получаются, если сотрудникам дано указание не производить ручное редактирование.

115. Первоначально единственным способом внесения исправлений в набор данных было их изменение вручную. Некоторые страны до сих пор испытывают неудобства, используя автоматизированную коррекцию, и поэтому они

делают это вручную на одном из описанных выше этапов. Если набор данных небольшой, сроки приемлемы и работники действуют весьма эффективно, то во многих случаях ручная коррекция срабатывает. Преимущество здесь заключается в том, что если информация в вопроснике представлена полно и точно, а несогласованность можно фактически устранить, глядя на заполненный бланк вопросника, то качество переписи или обследования, вероятно, улучшится лишь незначительно (группа редактирования должна исходить из того, например, что «Мэри» — это не «Гари»; что если речь идет о рождаемости, то данные, видимо, собраны в отношении этого лица, а не по ошибке). Фактически, процедуры редактирования и вменения редко повышают качество собранных данных. Они только меняют некоторые элементы.

116. Иногда просматривать переписной лист в целях проведения ручной коррекции бесполезно. Информация там по какой-то причине отсутствует. Иногда то или иное лицо не хочет указывать свой возраст, поэтому данный элемент в вопроснике не заполнен. В этом случае изучение вопросника не разрешит проблему. Тогда группа редактирования должна решить, как выйти из этой ситуации. Для ручной коррекции национальное бюро переписи/статистическое управление должно вписать «неизвестно» или применить какой-то набор величин для указания данных по элементу возраста.

117. Если не обращаться к респонденту, ручная коррекция неизбежно снижает качество и согласованность данных. Она требует больше времени и затрат. Компьютеры не устают, поэтому они работают быстрее; у них нет личных проблем, которые могли бы помешать сохранению качества или согласованности; и в большинстве случаев они снижают себестоимость обработки. Большинство стран используют сейчас тот или иной вид автоматизированной коррекции.

118. Недостающие и несогласующиеся ответы снижают качество данных и затрудняют представление таблиц по результатам переписи в легко понимаемой форме. Некоторые пользователи предпочитают указывать в таблицах недостающие или несогласующиеся ответы в категории «не сообщается», в то время как другие предпочитают распределять их пропорционально между сообщенными согласующимися данными. Еще одна группа пользователей рекомендует устанавливать правила по вменению «вероятных» ответов для недостающих или несогласующихся данных. Использование компьютеров позволяет эффективно вменять ответы на основе другой информации в вопроснике или указанной информации в отношении какого-то лица или какой-то жилищной единицы с аналогичными характеристиками.

119. Поскольку компьютер позволяет увидеть многие характеристики, из этого следует извлекать пользу в процессе редактирования. Таким образом, редактирование с использованием многих смежных характеристик может позволить произвести вменение более приемлемых ответов, чем при простом редактировании. С другой стороны, слабо разработанная процедура редактирования может привести к получению данных переписи плохого качества. В группу редактирования должны входить опытные специалисты по содержательной части переписи из самых разных имеющих к этому отношению областей, а также обработчики данных. Члены группы редактирования должны тщательно отобрать переменные для их проверки на согласованность, с тем чтобы определить спецификации редактирования и вменения. Программная продукция должна содержать определенную про-

центную долю ответов, которые были изменены или вменены. В этом случае аналитикам легче будет судить о качестве данных; так, высокая доля вмененных данных может служить предупреждением к тому, чтобы использовать эти данные с осторожностью.

120. Журнал редактирования, или контрольный журнал, фиксирует изменения каждой переменной. Журнал используется для того, чтобы проследить историю ответов, начиная с получения данных до процесса редактирования и вменения.

С. Руководящие принципы коррекции данных

121. Редактирование, будь то ручное или автоматизированное, должно быть направлено на то, чтобы данные как можно ближе отображали реальную ситуацию путем устранения пропусков и неправильных записей или путем изменения несогласующихся данных.

Вставка 2

Основные руководящие принципы коррекции данных

После определения процедуры редактирования может оказаться полезным учитывать следующие рекомендации по коррекции данных:

- ✎ вносить как можно меньше изменений в первоначальные зарегистрированные данные;
- ✎ устранять очевидные несоответствия между ответами;
- ✎ вводить данные для ошибочных или недостающих элементов, используя другие введенные данные о жилищной единице, лице или других лицах в домохозяйстве или в сопоставимой группе в качестве руководства, всегда соблюдая при этом конкретные процедуры. В некоторых случаях для определенных элементов подходит категория «не сообщается».

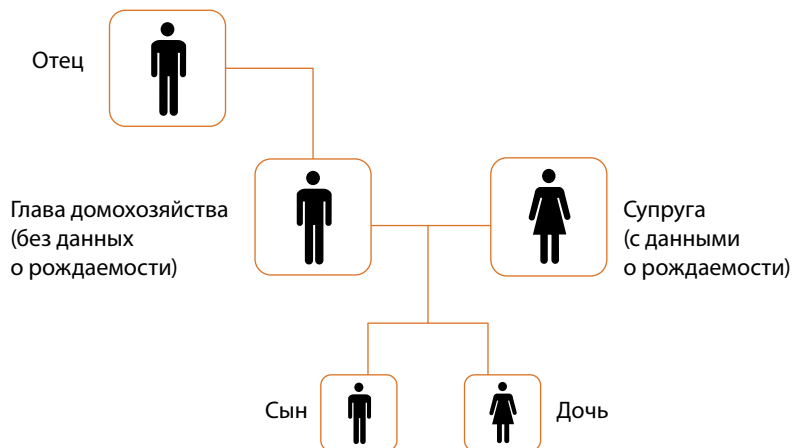
122. Рассмотрим следующую диаграмму (рисунок 2, ниже) для конкретного домохозяйства. Диаграмма показывает домохозяйство с согласующимися родственными связями и данными касательно пола. Глава домохозяйства — мужчина, и по нему нет информации о рождаемости; его супруга — женщина, и о ней есть соответствующая информация.

123. Однако во многих случаях информация непоследовательна. И тогда возникают следующие вопросы: каким должен быть процесс редактирования в отношении домохозяйства с несогласующимися данными? Как должна группа редактирования проводить свою работу, если глава домашнего хозяйства и его супруга зарегистрированы как лица мужского пола, как на рисунке 3 (ниже)? Раньше типичное правило редактирования содержало предпосылку о том, что первым лицом в паре является мужчина, особенно если это лицо является главой домохозяйства, а второе лицо, или супруга, является женщиной.

124. Если главой домохозяйства в этом случае будет не муж, а жена, то принятое правило редактирования будет неправильным, и национальное бюро переписи/статистическое управление совершит в конечном счете четыре ошибки:

Рисунок 2

Типовое гипотетическое домохозяйство с указанием родства, пола и данных о рождаемости его членов



- пол главы домохозяйства будет указан неправильно;
- пол супруга будет указан неправильно;
- глава домохозяйства потеряет информацию о рождаемости;
- супругу мужского пола будет по ошибке приписана информация о рождаемости.

Таким образом, совершенно очевидно, что подобная процедура редактирования не является удовлетворительной.

125. И наоборот, когда при удовлетворительной процедуре редактирования обнаруживается, что и глава и супруг одного пола, тогда производится проверка данных обоих лиц на фертильность. Поскольку выясняется, что репродуктивной способностью обладает только глава домохозяйства, то глава рассматривается как лицо женского пола. В этом случае правила редактирования таких элементов удовлетворительны.

Рисунок 3

Пример домохозяйства, где глава и супруг одного пола

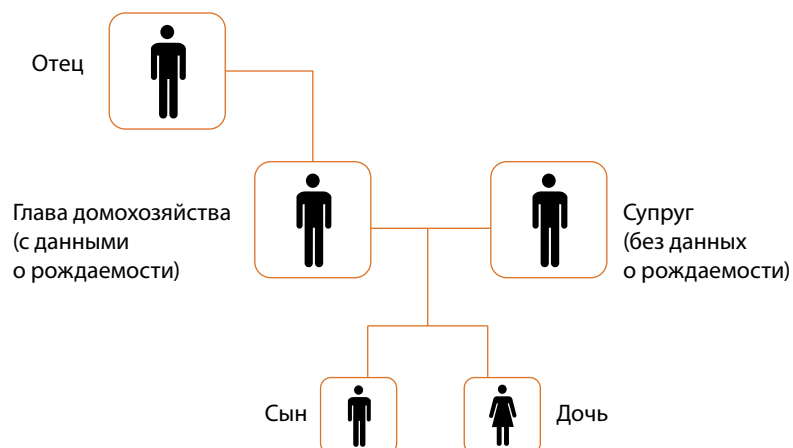
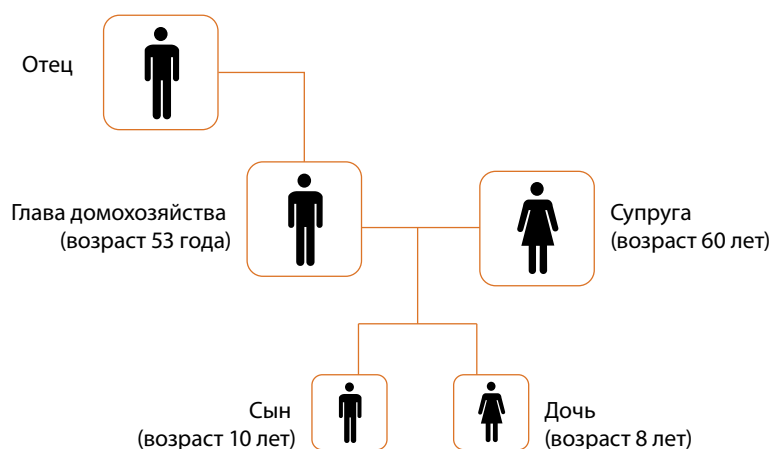


Рисунок 4

Пример домохозяйства с указанием возраста некоторых членов домохозяйства

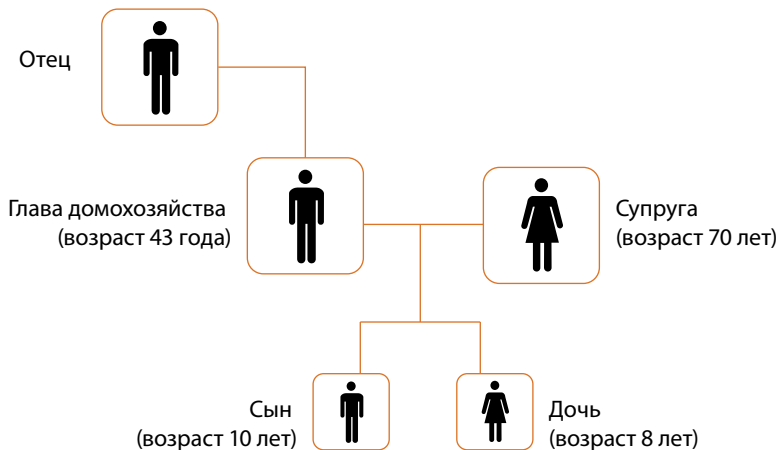


126. Еще один пример, также иллюстрирующий этот вопрос, показан на рисунке 4. В большинстве стран считается, что детородный возраст составляет от 15 до 49 лет. Предположим, что женщина сообщает о рождении ребенка в 52 года, что подтверждается прямым свидетельством через номер строки, указанной для матери ребенка, или расчетной разницей в возрасте (разница в возрасте между матерью и ее биологическим ребенком, вероятно, не может превышать 50 лет, хотя для приемных детей разница в возрасте может быть и больше). Группа редактирования должна решить, приемлема указанная разница в возрасте или ее надо заменить, поменяв при редактировании один или другой возраст. Если увеличивается возрастной диапазон деторождения при редактировании, и другие женщины заявляют о рождении детей в более позднем возрасте, то в набор данных могут быть внесены другие отклонения, если сам возраст сообщен неправильно. В этом случае группа редактирования снова должна решить вопрос о пригодности сообщаемых возрастов для конкретных переменных.

127. На рисунке 5, ниже, предлагается еще один возможный сценарий. Предположим, что при редактировании обнаруживается, что женщина в возрасте 70 лет имеет детей 10 и 8 лет, как на рисунке 4. Такая ситуация возможна, потому что муж мог бы иметь детей от предыдущей жены. При таких обстоятельствах дети относятся к главе домохозяйства, а не к супруге как таковой, даже если более вероятно, что операторы сделали ошибку, набрав «7» вместо «4», имея ввиду 40 лет. Независимо от причины предположим, что специалисты по содержанию данных переписи требуют от обработчика данных заменить возраст матери или ребенка, если разница между ними составляет более 50 лет. Такое требование ведет к еще более сложному редактированию. Если женщине 70 лет, а первому ребенку 10, то группа редактирования должна решить, чей возраст заменить. Группа редактирования может принять решение заменить возраст первого ребенка на 20 лет, и тогда проблема для этого первого ребенка разрешится, или группа может заменить возраст супруги, и тогда остается проблема с возрастом второго ребенка, который также требует редактирования.

128. Если учитывать только возраст матери и одного ребенка, то при вменении возраст будет приписываться произвольно, и это будет правильно

Рисунок 5

Пример домохозяйства с возможной несогласованностью указанных возрастных данных

примерно в половине случаев. Однако если редактор посмотрит еще и на возраст мужа, то более вероятно, что он изменит возраст супруги, руководствуясь этой дополнительной информацией. Такое изменение сделает возрастные показатели всей семьи более совместимыми.

D. Проверка на достоверность и согласованность

129. Одно из основных требований к редактированию заключается в том, чтобы ни в одном элементе данных не было недостоверных величин. Кроме того, ответы по всем связанным друг с другом элементам в рамках отдельных записей и между ними должны согласовываться между собой. Недостоверными записями называются те записи, которые неприемлемы по техническим или эстетическим причинам. Например, коды только для мужчин или только для женщин допустимы для указания пола. Любая другая величина будет неприемлема и потребует изменения на «неизвестно» или на запись, обозначающую один из двух допустимых полов; поскольку большинство стран планируют и формируют свою политику для множества переменных величин исходя из признака пола, наличие неизвестных величин в наборе данных осложнит усилия по получению необходимых для такой работы однозначных величин. Аналогичным образом, таблицы с таким несоответствием данных, как «тростниковые стены и бетонная кровля», «женщина в возрасте 13 лет с 20 детьми», «трехлетний ребенок со степенью доктора наук» покажут некомпетентность статистического управления, даже если несколько случаев несоответствия не повлияли на фактический процесс планирования в стране.

130. При вменении данных нужно в максимальной мере учитывать всю информацию о связанных переменных одновременно, а не обязательно последовательно по отношению к связанным величинам. Однако в некоторых случаях редактор может провести проверку на согласованность до определения достоверности той или иной записи. Если в процессе вменения величина присваивается на основе проверки на согласованность, то необходимо сравнить эту

величину с исходной записью, для того чтобы определить, есть ли фактическое изменение. Если изменения нет, то исходная запись остается в прежнем виде.

1. Нисходящий принцип редактирования

131. Эта процедура начинается с редактирования первого элемента («верхнего»), который обычно является первой переменной в вопроснике, с последовательным переходом к другим элементам вплоть до завершения редактирования всех элементов данных. Обычный подход предусматривает, что сначала принимаются во внимание показатели доли ответов на вопросы и сравнительная важность различных элементов данных. Редактирование обычно начинается с данных относительно пола и возраста в силу их важности, особенно при динамическом вменении. Хотя нисходящий принцип не отслеживает взаимосвязи между элементами данных, он тем не менее обеспечивает приемлемые рамки для завершения редактирования.

132. В процессе редактирования некоторые поправки изменяют значение того или иного элемента данных более одного раза. При такой процедуре в набор данных могут быть внесены одна или более ошибок. Вмененное значение может не согласовываться с другими данными. Даже если переменные обрабатываются последовательно, конкретную переменную следует по возможности редактировать одновременно со всеми другими величинами. Например, возраст ребенка, исчисленный по возрасту матери, может не согласовываться с указанными годами посещения школы или годами проживания ребенка на данном счетном участке. В этом случае возраст будет повторно вменяться до тех пор, пока не исчезнет противоречие. Вмененный возраст будет промежуточной переменной до окончательного закрепления. При редактировании вмененные промежуточные переменные не следует регистрировать как изменения до окончательного закрепления.

133. Хотя в отношении некоторых элементов данных и условий программа редактирования может воспринять незаполненные графы или запись «не сообщается», связанная с ними информация поможет воспроизвести большинство неуказанных или ошибочно указанных данных. Полученные таким образом показатели могут быть правильными или неправильными в зависимости от конкретного случая. Однако широкие возможности и скорость компьютера при сопоставлении самых различных хранящихся данных позволяют определить замещенные значения, которые достаточно четко характеризуют ситуацию. Полученные в результате этого таблицы в большинстве случаев будут содержать меньше противоречий, чем те, которые составлены по неотредактированным записям или записям, в которых все неприемлемые значения при вменении переводятся в категорию «не сообщается».

134. Программа редактирования может производить также структурные проверки (см. [главу III](#)). Она должна проверять элементы данных, относящиеся к населению (см. [главу IV](#)), и элементы, касающиеся жилья (см. [главу V](#)). Кроме того, в процессе редактирования, вероятно, следует создавать одну или несколько повторно закодированных переменных по отдельной записи, необходимой для составления таблицы, как отмечается в приложении I.

135. Чрезвычайно важно избегать кругового редактирования, сначала меняя один или несколько элементов, а затем на каком-то более позднем этапе возвращая их к прежнему значению. В других разделах данного *Руководства* от-

мечается, что сотрудники должны делать несколько прогонов, с тем чтобы быть уверенными в том, что все элементы данных полностью отредактированы. Можно выработать критерии редактирования для изменения данных в ходе первого прогона, но при этом, если их применять к измененным данным при втором прогоне, произойдет возвращение к первоначальной конфигурации. Такая процедура может продолжаться в течение многих прогонов. Группа редактирования должна избегать введения таких критериев в процесс редактирования.

2. Редактирование по принципу групповых переменных

136. «Нисходящий» принцип редактирования данных переписи и обследования, представленный в разделе 1, выше, не всегда дает лучшие результаты, то есть те, которые ближе всего соответствуют реальному распределению переменных. Как уже говорилось, если нисходящий принцип применяется без надлежащих предосторожностей, он часто вызывает проблемы при редактировании.

137. Другой принцип заключается в редактировании групповых переменных и основан на системе Феллджи–Холта. Этот принцип требует больше специальных знаний по применению компьютеров и более высокой производительности компьютеров, но может принести результаты, более близкие к «реальности». Различные виды редактирования групповых переменных приводятся в приложении V «Методы вменения». В системе редактирования групповых переменных необходимо определить набор позитивных заявлений и взаимосвязь между переменными. Затем редактор проверяет каждое заявление по отношению к данным о домашнем хозяйстве, чтобы посмотреть, все ли заявления верны. По каждому ложному заявлению редактор отслеживает элемент за элементом все недостоверные записи или несоответствия. После всех проверок система редактирования и вменения должна оценить, как лучше всего изменить запись, с тем чтобы она прошла все редакционные проверки. Группы редактирования, как правило, применяют принцип наименьших изменений и меняют минимально возможное количество переменных для получения приемлемой записи.

138. Одиннадцать декларативных заявлений на рисунке 6 служат примером правил, которые можно применять при редактировании отдельных характеристик населения по групповому принципу. В этом примере глава домашнего хозяйства должен быть в возрасте 15 лет или старше. Для обобщенных отредактированных данных было бы лучше использовать «X» лет, где X — установленный минимум для данной страны. Такие приведенные в этом примере заявления, как степень родства, пол, возраст, семейное положение и рождаемость, сосредоточены на других важных первичных переменных. Эти переменные тесно взаимосвязаны, и поэтому группы редактирования должны рассматривать их вместе, с тем чтобы отредактировать данные наиболее эффективным образом. Здесь следует отметить, что, хотя все переменные важны, некоторые из них имеют большее значение для представления данных, чем другие.

139. Рисунок 3 иллюстрирует простой случай, когда по какой-то причине и глава домохозяйства, и его супруга имеют один и тот же пол и, как показывают данные, оба они — мужчины, при этом один из них — мужчина с данными о рождаемости. Совершенно очевидно, что признак пола указан неправильно (что отражено в примечании ниже), и что пол мужчины с данными о рождаемости должен быть изменен на женский.

Рисунок 6

Пример правил редактирования отдельных характеристик населения по принципу групповых переменных

№	Правило	Степень родства	Пол	Возраст	Семейное положение	Рождаемость
1	Глава домохозяйства должен быть в возрасте 15 лет или старше					
2	Супруг(а) должен(на) быть в возрасте 15 лет или старше					
3	Супруг(а) должен(на) состоять в браке					
4	Если есть супруг(а), глава домохозяйства должен состоять в браке					
5	Если есть супруг(а), глава домохозяйства и супруг(а) должны быть противоположного пола	1	1			
6	Лицо младше 15 лет никогда не должно состоять в браке					
7	Лицо мужского пола не должно иметь данных о рождаемости		1			1
8	Лицо женского пола младше 15 лет не должно иметь данных о рождаемости					
9	Для лица женского пола в возрасте 15 лет или старше графа «рождаемость» должна быть заполнена					
10	Ребенок должен быть младше главы домохозяйства					
11	Родитель должен быть старше главы домохозяйства					
	Всего	1	2			1

Примечание: Значения «1» указывают, что один или более элементов данных не согласованы. Например, в позиции 5, если глава домохозяйства или супруг(а) относятся к одному полу, программа редактирования отмечает несоответствие степени родства и пола, и в этих ячейках появляется «1».

140. В примере на рисунке 7 оба супруга относятся к той же совокупности, что и на рисунке 6. Оба зарегистрированы как лица мужского пола. Здесь процедура редактирования простая и понятная. Переменная с наибольшим количеством дублированных ошибок должна быть исправлена в первую очередь. На рисунке 7 программа-редактор осуществляет процедуру вменения для графы «пол», поскольку, если основываться на данных рисунка 6, эта переменная содержит больше всего ошибок по отношению к 1) родству и полу и 2) рождаемости и полу. Когда программа-редактор проверяет рождаемость и обнаруживает, что в отношении главы домохозяйства информация о рождаемости присутствует, а в отношении супруга нет, то методом вменения признак «женский пол» присваивается главе домохозяйства. И наконец, когда группа редактирования вновь проверяет ряды повторений и все позитивные заявления верны, дальнейшего редактирования не требуется.

Рисунок 7

Пример с главой и супругом одного пола в неотредактированном наборе данных и его разрешение

Лицо	Степень родства	Пол	Дети, родившиеся живыми
Неотредактированные данные			
1	Глава домохозяйства	Мужской	03
2	Супруг	Мужской	Пропуск
Данные после редактирования графы «пол»			
1	Глава домохозяйства	Женский	03
2	Супруг	Мужской	Пропуск

Рисунок 8

Выборочные спецификации редактирования для исправления переменной пола с символическим кодом

If ПОЛ ГЛАВЫ ДОМОХОЗЯЙСТВА = ПОЛ СУПРУГА

Если графа «Рождаемость» ГЛАВЫ ДОМОХОЗЯЙСТВА заполнена

Если графа «Рождаемость» СУПРУГА не заполнена

(если ПОЛ главы домохозяйства оказался не женским) сделать ПОЛ = женский; endif

(если ПОЛ супруга оказался не мужским) сделать ПОЛ = мужской; endif

Else Произвести некоторые дополнительные действия, потому что у них один и тот же пол и у обоих указан рождаемость!!! [«Произвести некоторые дополнительные действия» может означать использование пола предыдущего главы домохозяйства, или попеременное указание пола главы домохозяйства, или использование соотношения полов всех глав домохозяйств для надлежащего ответа и т.д.]

End-if

End-if

Else Это тот случай, когда графа «Рождаемость» главы домохозяйства не заполнена

Если графа «Рождаемость» СУПРУГА заполнена

(если ПОЛ главы домохозяйства оказался не мужским) сделать ПОЛ = мужской; endif

(если ПОЛ супруга оказался не женским) сделать ПОЛ = женский; endif

Else Произвести некоторые дополнительные действия, если у ОБОИХ не указана рождаемость!!! [«Произвести некоторые дополнительные действия» может означать использование пола предыдущего главы домохозяйства, или попеременное указание пола главы домохозяйства, или использование соотношения полов всех глав домохозяйств для надлежащего ответа и т.д.]

End-if

End-if

End-if

Рисунок 9

Пример анализа данных об очень молодой вдове с тремя детьми, отредактированных по принципу групповых переменных

№	Правило	Степень родства	Пол	Возраст	Семейное положение	Рождаемость
1	Глава домохозяйства должен быть в возрасте 15 лет или старше	1		1		
2	Супруг(а) должен(на) быть в возрасте 15 лет или старше					
3	Супруг(а) должен(на) состоять в браке А					
4	Если есть супруг(а), глава домохозяйства должен состоять в браке					
5	Если есть супруг(а), глава домохозяйства и супруг(а) должны быть противоположного пола					
6	Лицо младше 15 лет никогда не должно состоять в браке			1	1	
7	Лицо мужского пола не должно иметь данных о рождаемости					
8	Лицо женского пола младше 15 лет не должно иметь данных о рождаемости		1	1		1
9	Для лица женского пола в возрасте 15 лет или старше графа «рождаемость» должна быть заполнена					
10	Ребенок должен быть младше главы домохозяйства					
11	Родитель должен быть старше главы домохозяйства					
	Всего	1	1	3	1	1

141. Спецификации редактирования для этой программы можно записать, как показано на рисунке 8, выше. Если графа рождаемости заполнена для обоих супругов, то редактор сработает. Однако очевидно, что редактирование будет неполным, поскольку программа просматривает только те случаи, где точно указана рождаемость как для главы домохозяйства, так и для супруга.

142. На рисунке 9, выше, показан пример, когда при редактировании согласно введенной с клавиатуры информации получается, что главе домашнего хозяйства женского пола 13 лет, она вдова, но с тремя детьми. Такие данные получаются, когда программа прочитывает правила редактирования.

143. Мы продолжаем рассматривать пример 13-летней вдовы — главы домохозяйства с тремя детьми. Первая проверка, описанная в правиле 1 — глава домохозяйства должен иметь возраст 15 лет или старше — оказывается неудачной, поскольку глава домохозяйства моложе 15 лет. Поскольку ее возраст 13 лет, то ставятся метки в графах «степень родства» и «возраст» в силу несоответствия между этими двумя переменными. Она не является супругой, следовательно, не применяется ни правило 2, ни правило 3. Правила 4 и 5 не задействуются по той же причине: они применяются только к супругам. Тем не менее, поскольку согласно правилу 6 лицо младше 15 лет никогда не должно состоять в браке (в данном случае речь идет о возрасте 13 лет), а наша 13-летняя девочка является «вдовой», это правило нарушается. Поскольку правило 7 применяется к лицам мужского пола, оно не срабатывает. Согласно правилу 8 лица женского пола младше 15 лет не должны иметь данных о рождаемости; ввиду того что данное лицо имеет рождаемость, нарушается и это правило. Правила 9, 10 и 11 к данному лицу не применяются.

144. Исходя из ряда положительных ответов переменная возраста в наибольшей степени ошибочна и должна быть исправлена в первую очередь. После исправления возраста редактирование будет закончено, если это устраняет все случаи несогласованности. В противном случае программа будет редактировать эту переменную в отношении последующего наибольшего количества несоответствий.

Е. Методы коррекции и вменения данных

145. Как отмечалось выше, незаполненные графы в записях данных, такие как «не сообщается», «неизвестно» или иная недостающая информация встречаются во всех переписях и обследованиях. Недостоверные данные вносятся также респондентом, счетчиком или возникают в результате ошибок при вводе данных. Методы коррекции бывают самыми разными и зависят от конкретного элемента данных. В большинстве случаев элементам данных можно присвоить достоверные коды с достаточной уверенностью, что они правильные, используя при этом ответы из других элементов данных в записях о данном лице или домохозяйстве или в записях о других домохозяйствах или лицах.

146. В данном *Руководстве* представлены два компьютерных метода коррекции ошибочных данных. Один из них — метод статического вменения, или метод «cold deck», — используется главным образом для недостающих или неизвестных элементов. Другой метод — метод динамического вменения, или метод «hot deck», — можно применять для недостающих данных, а также для несогласующихся или недостоверных элементов. В различных компьютерных пакетах и различных программах в этих пакетах используются разные методо-

логии и применяются методы «cold deck» и «hot deck» различными путями, как показано в приложениях.

1. Метод статического вменения («cold deck»)

147. При статическом вменении («cold deck») программа редактирования присваивает конкретный ответ по недостающему элементу данных из предварительно установленного набора, или ответ вменяется пропорционально из распределения достоверных ответов. При использовании метода «cold deck» программа не обновляет первоначальный набор переменных. Значения не меняются по сравнению с теми, которые содержатся в исходной статической матрице после обработки записей для первого, второго, десятого или любого иного лица. Исходные значения позволяют производить вменение недостающих данных.

148. Статическое вменение, как и динамическое вменение, представляет собой стохастический метод, но получаемые с его помощью значения со временем не меняются. Эта концепция описана в приложении V.

149. Иногда при статическом вменении используется метод соотношений, когда ответы присваиваются исходя из заранее установленных пропорций. В качестве примера пропорционального распределения ответов представим себе таблицу достоверных данных, то есть данных, заполненных для недостающих элементов об отработанном времени в неделю для мужчин в возрасте 33 лет, занятых в сельском хозяйстве, где показано, что 25 процентов таких мужчин работали 50 часов в неделю; 40 процентов — 60 часов в неделю; и 35 процентов — 70 часов в неделю. Недостающие или недостоверные ответы об отработанном времени мужчин в возрасте 33 лет в сельском хозяйстве были заменены: 25 процентов времени на 50 часов, 40 процентов времени на 60 часов и 35 процентов времени на 70 часов. Однако, если нет достоверных данных из предыдущих переписей, обследований или других источников, этот метод требует составления предварительных таблиц достоверных ответов из проводимой переписи, что может быть нереально с экономической и оперативной точек зрения.

2. Метод динамического вменения («hot deck»)

150. Еще одним способом избавления набора данных от неизвестных величин является динамическое вменение, или метод «hot deck», который используется при распределении значений для отсутствующих, неизвестных, неправильных или несогласующихся элементов. Первоначально этот метод был разработан бюро переписи США, но впоследствии другие агентства его усовершенствовали. При динамическом вменении используется одна или более переменных для оценки вероятного ответа, когда в наборе данных появляется неизвестная величина (или в некоторых случаях несколько неизвестных величин). Динамическое вменение становится все более популярным при редактировании данных переписи, поскольку его легко использовать и получать при этом выверенные, воспроизводимые результаты. Кроме того, устраняя неизвестные величины, легче проследить тенденции между переписями и обследованиями, поскольку аналитику не надо иметь дело с неизвестными величинами в каждом конкретном случае.

151. При динамическом вменении известные данные об отдельных лицах с аналогичными характеристиками дают наиболее приемлемую информацию,

которую можно использовать, если неизвестна какая-то часть информации о другом человеке. К этим характеристикам относятся пол, возраст, степень родства с главой домохозяйства, экономический статус и образование. Сама матрица вменения представляет собой набор величин, аналогичный колоде карт. На этих матрицах хранится, а затем выдается информация, используемая в тех случаях, когда встречаются неизвестные величины. Эта колода постоянно меняется путем обновления данных и/или логического «тасования колоды», с тем чтобы изменять вмененные величины во время обработки данных; отсюда и термин «hot deck» — «горячая колода».

152. Величины, хранящиеся в наборе «hot deck», представляют собой информацию о «ближайших соседях» с аналогичными данными. Следует отметить, что ближайший сосед — это обычно *предыдущий* сосед, поскольку, особенно в рамках описанного ранее нисходящего метода, жилищные единицы и проживающие в них лица рассматриваются только один раз, а затем программа двигается вперед. Таким образом, в пределах той или иной деревни, например когда неизвестны данные о сиротстве некоего лица по материнской линии, набор «hot deck» будет содержать информацию о самом последнем по времени лице того же пола и возраста с подтвержденными данными о сиротстве по материнской линии. Такой подход играет особо важную роль в странах со сравнительно большими потоками миграции или высокой распространенностью ВИЧ/СПИДа или иными явлениями, требующими неординарной статистической работы. Аналогичным образом, характеристики жилого фонда с большей вероятностью будут схожими в рамках одной деревни или группы деревень, чем между деревней или группой деревень и другими регионами страны.

153. Это можно просто проиллюстрировать тем фактом, что единственная величина может храниться в виде «колоды». Например, если пол того или иного лица по какой-то причине указан неправильно, «колоде» присваивается произвольное исходное значение пола (мужской или женский), определяя, таким образом, исходную величину. Это случайное значение становится показателем пола первого лица, которое встретится с пометкой «пол неизвестен». Однако если пол первого лица указан правильно, то пол первого лица заменяет случайное значение. Если неизвестен пол второго лица, то тогда матрица вменения присваивает пол, который в ней хранится. В этом случае вмененный пол — это пол первого лица. В сущности, если программа-редактор находит приемлемую величину для того или иного элемента, она закладывает ее в матрицу вменения. При обнаружении неприемлемой величины она заменяется вмененным достоверным значением из матрицы вменения.

154. Одна из проблем, которая встречается при описываемом здесь динамическом вменении («hot deck»), заключается в том, что если два различных элемента имеют неизвестные значения, то один и тот же «донор» не может быть использован для присвоения достоверных ответов. Каждая величина может исходить от «реального» лица, но это могут быть разные люди. Лучше всего было бы присваивать обе переменные одновременно от одного и того же лица. Однако программирование таких сложных матриц может представлять некоторые трудности.

155. Приводимые ниже данные (рисунок 10) представляют собой набор из десяти отдельных лиц. Незаполненные графы, отмеченные символами «X» и «XX», указывают на отсутствующие данные. Часто недостающую информацию

Рисунок 10
Выборочные домохозяйства в качестве примера входных данных для динамического вменения

Идентификационный номер	Степень родства	Пол	Возраст
1	1	1	39
2	2	2	35
3	3	1	13
4	3	X	10
5	4	2	40
6	4	1	XX
7	4	2	13
8	5	X	XX
9	5	1	44
10	5	2	36

Примечание: X и XX = отсутствующая информация.

обозначают числа 9 и 99, в данном случае для пола (9 для одного знака) и для возраста (99 для двух знаков). Иногда, однако, возникает необходимость, чтобы цифра 9 замещала другую, реальную величину, например при ограниченном числе кодов степени родства; таким образом, эти величины должны использоваться очень экономно и на практике, если есть возможность использования другой величины типа «X», «.» или «..», ее, вероятно, следует использовать. Следует отметить, что хотя для использования при вменении имеются и другие переменные, такие как образование и род занятий, они в этот краткий пример не включены.

156. Если исходным значением для матрицы вменения под названием SEXARRAY является «мужской» (код = 1), матрица вменения будет выглядеть примерно следующим образом: SEX = 1.

157. После обработки информации о лице 1 значение останется 1. Однако после обработки информации о втором лице значение изменится на 2, поскольку это лицо женского пола. Тогда переменная будет выглядеть следующим образом: SEXARRAY = 2.

158. Для каждого правильного значения пола того лица, данные о котором обрабатываются в данный момент, код для пола этого лица заменяет значение в матрице вменения. Когда обрабатываются данные о третьем лице, значение при вменении снова меняется на 1 или «мужской».

159. Пол четвертого лица неизвестен, поэтому редактор смотрит на значение матрицы вменения и заменяет неизвестное значение значением с матрицы вменения. Лицо 5 женского рода, поэтому редактор заменяет предыдущее значение в матрице вменения по данным лица 3 (мужской). Такой процесс продолжается вплоть до лица 8.

160. При редактировании опять используется вменение, и лицо 8 становится лицом женского пола, поскольку значение в матрице вменения в отношении лица 7 указывает на женский пол. Матрица вменения используется при редактировании для получения величин дважды: один раз для получения значения «мужской» и один раз для получения значения «женский». Поскольку показатели пола появляются примерно с одинаковой частотой, то во время

длинного прогона каждый пол используется при вменении в пропорции примерно 50 на 50. После обработки данных по всем десяти лицам переменная будет выглядеть следующим образом: SEXARRAY = 2.

161. Матрица вменения приписывает пол именно таким образом, но при этом существуют и другие, более сложные способы применения такой процедуры. Например, программа-редактор может использовать степень родства к главе домохозяйства и пол для того, чтобы было легче определить возраст отдельного лица. Рассмотрим следующий частичный список кодов степени родства:

- 1 = Глава домохозяйства
- 2 = Супруг(а)
- 3 = Ребенок
- 4 = Другой родственник
- 5 = Не родственник

162. Обработчик данных может создать исходные значения возраста, которые могут приближаться к реальной ситуации, касающейся степени родства по признаку пола. Эти значения не столь важны, поскольку редактор почти наверняка заменит их, прежде чем использовать. Кроме того, редакция требует вменения многих значений, поэтому лишь немногие исходные величины влияют на окончательные таблицы. Эти величины могут быть такими, как показано на рисунке 11.

Рисунок 11

Исходная статическая матрица для возраста, исчисляемого исходя из пола и степени родства

	Степень родства				
	Глава домохозяйства (1)	Супруг(а) (2)	Сын/дочь (3)	Другой родственник (4)	Не родственник (5)
Мужской (1)	35	35	12	40	40
Женский (2)	32	32	12	37	37

163. Вновь рассмотрим десять лиц, представленных на рисунке 10. Поскольку первое лицо в нашей выборке записано как глава домохозяйства (код = 1) и он — мужчина (код = 1), его возраст (39) заменяет при вменении первый элемент (координаты 1,1). При этом в «колоде» находятся величины, указанные на рисунке 12.

Рисунок 12

Пример матрицы динамического вменения после одного изменения

	Степень родства				
	Глава домохозяйства (1)	Супруг(а) (2)	Сын/дочь (3)	Другой родственник (4)	Не родственник (5)
Мужской (1)	39	35	12	40	40
Женский (2)	32	32	12	37	37

164. Второе лицо является супругой (код = 2) и женщиной (код = 2), поэтому ее возраст (35) заменяет значение во втором ряду во второй колонке с заменой «колоды» на эти значения. Аналогичным образом, возраст других лиц в домохозяйстве заменяет через пятое лицо значения в матрице вменения.

165. Отметим, что во время предыдущей процедуры вменения пола пол 1 был приписан лицу 4. Ввиду того что программа-редактор требует вменения значения для пола, она не вносит в массив возраст этого лица в качестве обновления. Программа будет обновлять массив только теми значениями из записей, где пол и степень родства первоначально указаны правильно. Однако когда программа доходит до лица 6, она обнаруживает, что возраст неизвестен. Данное лицо относится к мужскому полу и является «другим родственником» главы домохозяйства. Исходя из этого, программа-редактор использует элемент матрицы вменения для лиц мужского пола, чья степень родства входит в группу «другой родственник» (четвертая колонка в первом ряду) и приписывает значение возраста для этой категории («другой родственник мужского пола» — в данном случае — 40).

166. Ни пол, ни возраст восьмого лица не указаны. Редактор вменяет пол как женский, а затем определяет возраст, исходя из этого выделенного пола и кода степени родства (5). В этом случае возраст составляет 37 лет.

167. Хотя программа редактирования вменила значение для возраста по известной степени родства, она пользовалась заранее выделенным значением пола для других переменных. В данном случае применение распределенных величин для дальнейшего вменения является примером слабой процедуры редактирования (см. раздел 3d), ниже). Здесь для вменения лучше поискать другие известные элементы данных, например семейное положение.

168. После проверки десятого лица значения матрицы вменения даны на рисунке 13. В этом примере в обоих случаях вмененных значений использовалась исходная статическая матрица. Как правило, при вменении используется лишь небольшое число исходных величин или вообще не используется. В большинстве же случаев используются величины, взятые из данных о зарегистрированном населении.

Рисунок 13

Пример матрицы динамического вменения после многократных изменений

	Степень родства				
	Глава домохозяйства (1)	Супруг(а) (2)	Сын/дочь (3)	Другой родственник (4)	Не родственник (5)
Мужской (1)	39	35	13	40	44
Женский (2)	32	35	12	13	36

3. Вопросы динамического вменения («hot deck»)

а) Географические соображения

169. Если в программе редактирования для условного начисления недостающих величин используется динамическое вменение, то следует попытаться использовать данные, отсортированные по наименьшему географически опре-

деленному району. С такой процедурой должна увеличиться вероятность получения правильного ответа, поскольку те, кто проживает в одном и том же небольшом географическом районе, как правило, имеют более или менее однородные характеристики по жилому фонду, демографические и прочие характеристики. Если совокупность населения неоднородна, то никакой корреляции не будет, и поэтому группа редактирования должна рассматривать переменные в каждом конкретном случае. Кроме того, как будет рассмотрено ниже, существуют переменные, которые неприменимы в определенных областях — например, центральное отопление в районах с очень теплым климатом — и это необходимо учитывать при редактировании.

b) Использование смежных элементов данных

170. Прежде чем прибегать к динамическому вменению для получения недостающих величин следует попытаться использовать смежные элементы данных, чтобы присвоить значение, которое, вероятно, будет правильным. Например, если не указано семейное положение того или иного лица, то программа-редактор определит, есть ли у этого лица супруг(а) в домохозяйстве. Если есть, то программа присвоит код для лица, состоящего в браке, не прибегая к матрице вменения. Однако если такого признака нет, то программа, возможно, будет полагаться на значение из матрицы вменения.

c) Как порядок расположения переменных влияет на матрицы

171. Национальные бюро переписи/статистические управления, пользующиеся матрицами вменения, должны определить, какие переменные им нужны, поскольку они устанавливают порядок их редактирования. Для элементов данных, касающихся населения, они захотят в первую очередь отредактировать данные относительно пола и возраста, с тем чтобы их можно было использовать в других матрицах вменения. При общем редактировании не следует использовать в матрицах вменения неотредактированные переменные, хотя большинство пакетов компьютерных программ воспримет «неизвестные» ряды или колонки. Доли ответивших и распределение признаков в рамках переменных величин помогут определить наилучшие переменные и наиболее полезные признаки в рамках этих переменных, чтобы содействовать в разработке значений «hot deck». В последующих матрицах вменения можно использовать эти элементы данных после редактирования. Однако по возможности следует исключать отредактированные данные из матрицы вменения.

172. Например, если программа-редактор исчисляет возраст по полу и степени родства, то ячейки массива для этой матрицы вменения (пол по степени родства) не следует обновлять, если пол или степень родства были вменены. Как правило, только в том случае, если возраст, пол и степень родства достоверны и согласуются между собой, пакет редактирования должен вводить возраст в ячейку для соответствующего пола и степени родства. Однако иногда в силу других факторов использование отредактированных данных неизбежно. Важно отметить, что в большинстве стран специалисты игнорируют это предложение и исчисляют данные из ранее вмененных величин. Возможным решением является использование флажков вменения, с тем чтобы препятствовать использованию вмененных данных для сопоставления той или иной донорской информации с несоответствующей единицей.

d) Сложность матриц вменения

173. Делая матрицу вменения более детализированной, национальное бюро переписи/статистическое управление повышает вероятность получения согласующейся, «правильной» величины в матрице вменения. Например, программа может исчислять семейное положение, используя только степень родства. Однако с возрастом увеличивается вероятность вдовства или развода. В связи с этим имеет смысл исчислять семейное положение по возрасту и степени родства. Программа-редактор, которая использует возраст и степень родства, берет данные о семейном положении от лица с теми же характеристиками в непосредственно предшествующей этому достоверной записи, хранящейся в матрице вменения.

174. Тем не менее описанная выше процедура может создать новые проблемы. Национальное бюро переписи/статистическое управление, как правило, редактирует элементы переписного листа в определенной последовательности, когда при применении нисходящего принципа возраст редактируется после семейного положения. В этом случае, если в записи нет данных о семейном положении и возрасте, то нельзя брать значение семейного положения непосредственно из предшествующей записи с такими же данными о возрасте и степени родства.⁵ В результате программа может не определить возрастную категорию для этой записи. Данную проблему можно решить по-другому, если в матрице вменения будет ряд или колонка для «не сообщенных» элементов. Такая процедура позволит программе устанавливать значение для семейного положения, используя категорию семейного положения из непосредственно предшествующей записи с теми же категориями родства и возраста с указанием «не сообщается». Однако против такого подхода можно привести два фактора. Во-первых, случай «не сообщается» в одной и той же комбинации настолько редкий, что будет трудно обновлять матрицу вменения для недостающего элемента. Во-вторых, по существу, нельзя получить методом «cold deck» соответствующие величины для этих комбинаций значений «неизвестно» в наборе «hot deck», поскольку они не существуют в «реальном» мире.

175. Разрешение вышеописанной проблемы создает больше работы для обработчика данных, но дает в результате более достоверный со статистической точки зрения продукт. Программа редактирования сначала ведет проверку, с тем чтобы определить, есть ли у этих элементов правильные коды. Если запись для данного лица не имеет правильного кода для этого элемента данных, то такой элемент для данной записи в матрице вменения не используется. Обработчики данных могут облегчить этот процесс, создав более простую матрицу вменения. В продолжение ранее приведенного примера можно отметить, что, если программа должна исчислить семейное положение, из-за того, что данная величина отсутствует, матрица вменения будет, как правило, иметь два измерения: возраст и степень родства. Если после проверки программа не найдет правильного кода для возраста, она будет исчислять семейное положение только по степени родства. Поскольку данные о степени родства редактируются до информации о семейном положении, код родства будет достоверным. Те же принципы используются в программе для всех процедур динамического вменения.

e) Разработка матрицы вменения

176. Сотрудники, занимающиеся содержанием переписи, совместно с обработчиками данных должны подготовить соответствующие матрицы вмене-

⁵ Наилучшая практика редактирования не предполагает использование отредактированных величин в рамках метода «hot deck». Иногда такой практике следовать довольно сложно в силу ограниченности сроков публикации результатов или сложности компьютерного программирования. В таких случаях исчисляется одна из нескольких переменных и ее значение помещается в надлежащие матрицы («hot decks»), а после этого используется для вменения последующих переменных.

ния (некоторые группы редактирования пользуются несколькими матрицами вменения). Матрицы вменения обновляются только достоверными ответами; группы редактирования не пользуются распределенными или вмененными величинами. Специалисты по содержательной части переписи и обработчики данных должны проверять спецификации редактирования и полученные методом «hot deck» данные на согласованность и полноту.

177. На разработку матрицы вменения нужно потратить немало времени и умственных усилий, включая исследования возможности использования административной отчетности и результатов предыдущих переписей или обследований, особенно для величин «cold deck». Но даже после научных исследований и разработок редакторам не следует применять матрицы вменения произвольно. Когда в матрицах вменения присутствует внутренняя несогласованность, требуются значительные усилия для их согласования. Если в матрицах вменения не используются стандартные условные обозначения, то персонал должен рассматривать каждую из них отдельно.

178. В примерах, приводимых в данном *Руководстве*, каждый элемент в матрицах вменения имеет одно значение, однако некоторые группы редактирования оставляют более одного варианта для каждой ячейки. Вы можете представить это в виде двухразмерной матрицы с третьим измерением, идущим обратно в информационную доску. Эти ячейки дают дополнительное измерение. Например, если возраст всех детей в семье неизвестен, скажем, в семье с четырьмя мальчиками компьютер не будет присваивать одно и то же значение четыре раза, создавая четверых близнецов. Вместо этого будут указаны четыре разных возраста. Однако даже здесь одно и то же значение может быть присвоено более одного раза в зависимости от того, что хранится в матрицах.

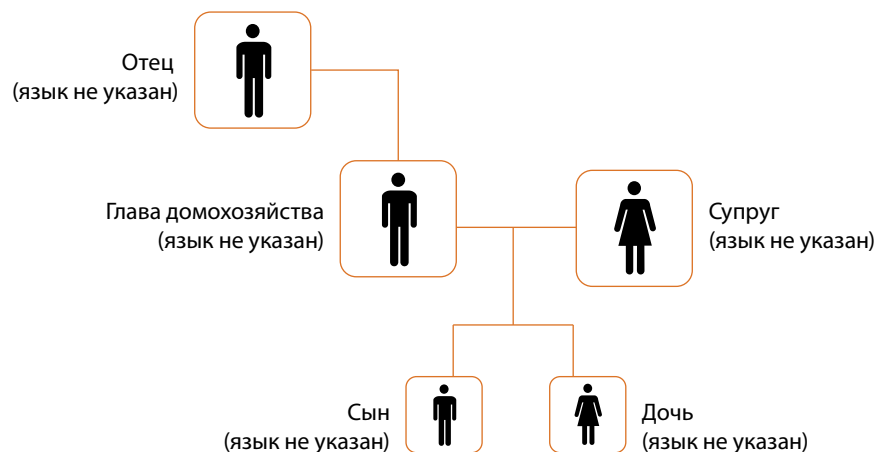
f) Стандартизованные матрицы вменения

179. С помощью стандартизованных матриц вменения можно рационализировать процесс редактирования. Матрицы вменения со стандартными измерениями различных социальных и экономических переменных, таких как возрастная группа и пол, можно быстро проверить и затем использовать.

180. Например, национальное бюро переписи/статистическое управление может решить разработать матрицу вменения с тем, чтобы определить код языка при отсутствии соответствующих данных. Почти наверняка программа-редактор проверит домашнее хозяйство, где проживает другое лицо, зарегистрированное как говорящее на данном языке. В случае неудачи программа может выбрать язык предыдущего лица того же пола из той же возрастной группы (обновив матрицу вменения, если все три элемента достоверны). Такая процедура покажет вероятный язык, поскольку лица, говорящие на одном и том же или аналогичном языке, как правило, территориально находятся близко друг от друга.

181. На рисунке 14 переменная «язык» не содержит никакой информации в отношении главы домохозяйства. По той или иной причине сканер или оператор могут не уловить информацию о языке или код или что-то другое может пойти неправильно. Однако поскольку супруг и дети все говорят на суахили, этот язык можно присвоить главе домохозяйства и отцу главы домохозяйства, чьи данные о языке тоже отсутствуют. Следует отметить, что главой домохозяйства на рисунке 14 является женщина.

Рисунок 16
Пример членов домохозяйства без указания языка



185. Исключения из правил редактирования будут делаться в самом начале редакторского прогона. Сотрудники должны внимательно следить за изменениями языка, которые могут происходить, когда они перемещаются из одного географического района в другой. Некоторые страны должны также следить за смешением языков среди местного населения. Однако даже в этом случае, если нет выборочного недоучета некоторых языков, процентная доля распределенных и нераспределенных величин после вменения должна быть примерно одной и той же.

186. В ходе редактирования еще можно обращать внимание на религию. Опять же ответы на вопросы о религии можно исчислять по возрасту и полу. Программа редактирования будет продолжать обновление, когда имеется вся информация, и будет извлекать ответы из матрицы вменения для информации из категории «неизвестно». Такая матрица вменения будет выглядеть как матрица для языка, но в ячейках матрицы вместо языка будет указана религия.

187. Заметим, что такое разъяснение предполагает нисходящий последовательный подход. Группы редактирования, использующие сложные методы, такие как метод Фелледжи–Холта или методология вменения по ближайшему соседу (МВБС) (см. [приложение V](#)), применяют все смежные приемы редактирования одновременно. Настоящая процедура также предполагает наличие определенного порядка редактирования.

188. Многие экономические характеристики, такие как уровень экономической активности, время, отработанное в последнюю неделю или недели, и время, отработанное за последний год, можно исчислить, используя аналогичные характеристики. Используя аналогичные матрицы вменения, программа редактирования может быстро проверить значение для характеристик переменных и процесс редактирования должен в целом пойти быстрее.

189. Иногда трудно получить надлежащим образом отредактированные характеристики для первых матриц вменения в серии. Как правило, статистическое управление не хочет использовать неотредактированные элементы в качестве измерений для матрицы вменения; редактор не будет использовать ни пол, ни возраст в качестве измерений матрицы вменения до того, как эти пока-

затели не будут отредактированы. Следовательно, в первых нескольких матрицах вменения будут использоваться разные переменные, которые не нуждаются в редактировании, или переменные, значение которых изменить нельзя. Для самой первой матрицы вменения в отношении элементов, касающихся населения, редактор может использовать число лиц в жилищной единице, используя значение «0» для незаселенных единиц.

190. Для редактирования данных о жилье в целом, в первой матрице вменения можно также использовать число лиц в жилищных единицах в качестве исходного измерения, но группа редактирования может изменить приемы для элементов, касающихся жилья, с тем чтобы учесть незаселенные единицы. Например, в отношении жилого фонда, если вначале редактируются такие элементы данных, как «строительный материал наружных стен» или «тип стен», исходные величины могут быть основаны на числе лиц в жилищной единице, включая значение для случая, когда жилищная единица является незаселенной.

191. Если в случае незаселенной единицы информация о «типе стен» достоверна, редактор обновляет первую ячейку о типе наружных стен. Если тип стен известен, то для заселенной жилищной единицы редактор обновляет ячейку, сообразуясь с числом лиц в данной единице. Однако если строительный материал внешних стен неизвестен, матрица вменения выдаст значение для строительного материала наружных стен на основе числа лиц в данной единице.

192. После первоначального использования этой матрицы вменения группа редактирования может пожелать переключиться на некоторые другие характеристики жилого фонда, такие как «тип крыши» или «условия владения». Независимо от выбранной характеристики должно проводиться четкое различие между единицами и обеспечиваться достаточное разнообразие, чтобы один и тот же критерий не выбирался повторно. Повторный отбор одного и того же критерия может дать значение, соответствующее квази «cold deck», а не то, которое получается при динамическом вменении («hot deck»). Например, применение динамического вменения в отношении армейских казарм («помещений, предназначенных для коллективного проживания») может привести к тому, что одна и та же величина будет использоваться повторно, если единственными отобранными характеристиками будут возраст и пол. В этом случае все проживающие будут, вероятно, мужского пола и большинство будет подпадать под ограниченный возрастной диапазон. Следовательно, эта конкретная матрица может и не дать наилучшие результаты. Если «условия владения» будет иметь достаточно разновидностей с достаточной долей владельцев и арендаторов, то такая переменная может сработать. В противном случае страна может использовать различные критерии для типов крыши.

193. В целом многие группы редактирования считают, что используя сопоставимые измерения для матриц вменения, они меньше занимаются проверкой, получают результаты более быстро и, возможно, более точные.

g) Когда динамическое вменение не используется

194. Если группа редактирования решает совсем не использовать динамическое вменение, важным моментом остается тем не менее последовательность редактирования. Например, возраст связан со многими элементами данных, включая степень родства с главой домохозяйства, уровень образования, занятость и рождаемость (для женщин). Рассмотрим данные о членах домохозяйства на рисунке 17:

Рисунок 17

Пример главы домохозяйства с ребенком при отсутствии данных о возрасте ребенка

Лицо	Степень родства	Возраст	Уровень образования	Работа	Род занятий	Дети, родившиеся живыми
1	1	40	12	1	33	ПРОПУСК
3	3	X	7	ПРОПУСК	ПРОПУСК	ПРОПУСК

Примечание: X = отсутствующие данные о возрасте; ПРОПУСК = не применяется.

195. Запись для лица 3 — степень родства 3 (ребенок), но без указания возраста. Чтобы определить возраст, программа редактирования может использовать разницу в возрасте между главой домохозяйства и ребенком (или величину, полученную методом «cold deck», или величину, взятую от предыдущей единицы путем вменения). Если, например, разница составляет 25 лет, то возраст ребенка получается 15 лет (возраст главы домохозяйства 40 лет минус разница в возрасте 25).

196. Количество лет обучения также известно и составляет в данном случае 7 лет. Возраст 15 лет вполне может соответствовать этому уровню образования. Поскольку количество соответствующих лет обучения для того или иного возраста меньше, чем возрастной диапазон для разницы в возрасте между главой домохозяйства и ребенком, то лучше проверить прежде всего, соответствует ли данным показателям уровень образования. И если данный уровень указан, то разница в возрасте, определенная путем статического («cold deck») или динамического («hot deck») вменения, может быть использована для получения соответствующего возраста. Если уровень образования неизвестен, то тогда разницу в возрасте между главой домохозяйства и ребенком можно использовать для присвоения возраста.

197. Однако даже информация о возрастной разнице может отсутствовать. По существу, для большинства стран более вероятно отсутствие данных об уровне образования, чем о возрасте. Но в следующем примере показано, что может предпринять группа редактирования, если отсутствуют данные о возрасте и об уровне образования.

198. На рисунке 18 нет данных ни о возрасте, ни об уровне образования, но другая информация имеется. Лицо 3 недостаточно взрослое, чтобы иметь работу и слишком молодое, чтобы иметь детей (или оно мужского пола). Используя информацию о занятости из набора величин «cold deck», можно получить возраст, но это значение будет ниже, чем наименьший возраст, приемлемый для работы. В качестве варианта, если группа редактирования использует динамическое вменение, значение в матрице вменения дает величину для возраста. Для выбранного возраста следует, вероятно, использовать возраст главы домохо-

Рисунок 18

Пример главы домохозяйства с ребенком при отсутствии данных о возрасте и уровне образования ребенка

Лицо	Степень родства	Возраст	Уровень образования	Работа	Род занятий	Дети, родившиеся живыми
1	1	40	12	1	33	ПРОПУСК
3	3	X	X	ПРОПУСК	ПРОПУСК	ПРОПУСК

зайства как одну из переменных для сохранения согласованности. Например, если главе домохозяйства не 40 лет, а 20, то, очевидно, было бы неправильным обозначать возраст 14 лет для лица 3. Когда возраст установлен, то можно определить и уровень образования, и он тем самым должен согласовываться и с возрастом, и со статусом работающего лица.

199. Если группа редактирования решит вменить все или большинство элементов данных, она должна решить вопрос о том, как логически построить процесс редактирования. Для элементов данных, касающихся населения, редактирование должно начинаться с рассмотрения всех элементов, где потенциально могут быть неизвестные величины. Группы редактирования должны использовать информацию из обследований, административной отчетности, предыдущих переписей пилотного этапа текущей переписи и другую имеющуюся информацию для того, чтобы было легче принимать решение о включении каждого элемента данных в первую и последующие матрицы вменения. Детали разработки матриц вменения в весьма значительной мере зависят от того, к какой стране они относятся, однако все национальные бюро переписи/статистические управления, скорее всего, имеют для этого какую-то информацию. Тестирование различных наборов переменных в значениях «hot decks» поможет в получении наиболее приемлемого набора данных для конкретной страны.

200. Многие пакеты программ редактирования отслеживают число лиц в конкретной жилищной единице по мере их обработки. Например, матрица вменения для неизвестного пола может позволить присвоить мужской или женский пол в зависимости от числа лиц, проживающих в жилищной единице. Следовательно, исходное значение, которое может быть выбрано для лица, пол которого неизвестен или недостоверен, в случае жилища, где проживает один человек, может быть «мужской». Для жилища, где проживают два человека, исходное значение может быть «женский». Для жилища, где проживают три человека, значение может быть «мужской» и т. д. Матрица в этом случае может быть использована только в качестве последнего средства, после того как все данные, такие как пол главы домохозяйства, супруг(а) и информация о рождаемости, проверены, произведена проверка на согласованность и все несоответствия устранены.

h) Какого размера должны быть матрицы вменения?

201. Большинство пакетов компьютерных программ способны воспринимать многомерные матрицы вменения. Прежде чем формировать матрицы вменения, следует учитывать следующие моменты.

i) Проблемы, возникающие в случае слишком большого размера матрицы вменения

202. Одна из наиболее серьезных проблем, с которой сталкиваются национальные бюро переписи/статистические управления, где совместно работают специалисты по содержательной части переписи и специалисты по обработке данных, — это слишком активные программы редактирования. Здесь очень легко увлечься разработкой пакетов редактирования, и в этом случае программирование занимает намного больше времени, чем это необходимо, и замедляет обработку данных переписи или обследования. Группа редактирования может, например, решить, что при определении возраста в дополнение к таким элемен-

там, как «пол», «уровень образования» и «уровень экономической активности», для женщин необходимо включать такой показатель, как «общее число родившихся детей». Добавление элемента «общее число родившихся детей» может сделать оценку возраста несколько более точной, но это может быть не оправдано из-за повышения сложности программирования. В этом случае группы редактирования должны решить, какая размерность матрицы вменения даст наилучшие результаты с точки зрения точности и эффективности. Слишком большие матрицы вменения (где слишком много ячеек) невозможно обновить полностью, и вместо этого могут неправильно использоваться значения «cold deck».

ii) *Осмысление принципа функционирования матрицы вменения*

203. Наряду с тем что матрицы вменения могут быть слишком большими, дезориентировать могут еще и пути доступа к данным. Важно, чтобы и сотрудники по содержательной части переписи, и обработчики данных могли отслеживать все маршруты доступа. Совместно они должны удостовериться, что матрица вменения выполняет поставленную задачу. И вновь специалисты по содержанию и обработчики данных должны совместно удостовериться, что каждая переменная и размер матрицы вменения применяются должным образом. Более того, они должны убедиться, что надлежащим образом работают все комбинации.

iii) *Проблемы, возникающие в случае слишком маленького размера матрицы вменения*

204. Матрица вменения является слишком маленькой, если у нее очень мало размерностей или если по причине группирования данных (когда слишком мало возрастных групп или уровней образования) одно и то же значение матрицы вменения используется неоднократно до обновления. Например, без определения пола в спектре данных о возрасте всем детям в семье, скорее всего, будет присвоен один и тот же возраст, если он неизвестен. Сотрудники, занимающиеся содержательной частью переписи, должны вместе с обработчиками данных проверить матрицы вменения на все возможные комбинации, с тем чтобы ни одна из них не повторялась слишком часто.

iv) *Элементы данных, представляющие трудности для матрицы вменения*

205. Некоторые элементы, такие как «род занятий» и «отрасль», оказались, как выяснилось, весьма трудными для редактирования. Раздельные матрицы вменения по роду занятий и отрасли могут дать несогласующиеся результаты, а попытки провести перекрестную проверку всех парных данных о роде занятий и отрасли могут оказаться дорогостоящими и сложными для исполнения. Например, если окажется, что парикмахеры работают на рыбоперерабатывающих предприятиях, то здесь потребуются какой-то иной вид редактирования. Кроме того, большое количество категорий рода занятости и отрасли может существенно затруднить проведение динамического вменения. В отношении некоторых элементов данных группа редактирования может решить, что в этом случае редактирование приведет к обратным результатам и предпочесть использовать показатели «не указано» или «не сообщается». В других случаях достаточным может оказаться подход статического вменения («cold deck»).

4. Проверка матриц вменения

206. Базовая структура матрицы вменения в программном пакете редактирования должна выглядеть примерно так, как показано на рисунке 19. В спецификациях редактирования должны быть указаны массивы, используемые для вменения, и значения «cold deck» для исходного набора значений.

a) Создание исходной статической матрицы

207. При выполнении нижеописанной процедуры матрица вменения обновляется каждый раз, когда обнаруживается лицо с достоверными данными по всем трем элементам — в данном случае «степень родства», «пол» и «возраст». Однако если программа редактирования находит недостоверное (или неуказанное) значение пола, то матрица вменения выбирает величину на основе достоверных кодов степени родства и пола (переменных, которые уже отредактированы).

b) Сообщения об ошибках

208. Пакеты редактирования должны предоставлять несколько методов, чтобы удостовериться, что редактирование и вменение идут должным образом. Два таких свойства команды «message» (сообщение) и «write» (писать) рассматриваются ниже.

209. Одним из источников информации является отображение сообщения, как показано на рисунке 19, ниже. Эта команда выдает конкретные со-

Рисунок 19

Выборочный набор величин для матрицы «cold deck» и выборочный код вменения

Исходный набор значений для матрицы вменения A01-AGE-FM-SEXRL (2,6)

Глава домохозяйства	Супруг(а)	Ребенок	Прочие родственники	Родители	Не сообщается	Пол
40	40	10	20	65	20	Мужской
40	40	10	20	65	20	Женский

.

Код вменения

If AGE = 0:98

допустим A01-AGE-FM-SEXRL (SEX,RELATIONSHIP) = AGE

Else

сообщение 'Возраст неизвестен, поэтому вменяется' AGE

написать 'Возраст неизвестен, поэтому вменяется, Age = ' AGE

вменить AGE = A01-AGE-FM-SEXRL (SEX,RELATIONSHIP)

сообщение 'AGE is now known' AGE

endif

.

общения и общий подсчет (общее число раз, когда появляется сообщение) для географических категорий (например, счетный участок, мелкое административно-территориальное подразделение, крупное административно-территориальное подразделение), а также для каждого переписного листа. Для всех переписных листов сводный отчет может выглядеть примерно так, как показано на рисунке 20:

Рисунок 20

Пример сводного отчета для ряда вменений на одну ошибку

Подсчет	Количество ошибок	Сообщение	Номер строки
–	14–1	Слишком много детей на одну женщину	2629
–	14–2	Слишком много детей на одну женщину	2645
2	14–3	Присутствующие мальчики не указаны	2669
2	14–4	Присутствующие девочки не указаны	2678
33	14–5	Месяц последнего рождения не указан	2723
7	15–6	Нет родившихся детей; разница в возрасте между матерью и ребенком подтверждена	2892

Примечание: Здесь «14» просто относится к элементу 14 в данном ряду; ошибки пронумерованы по порядку.

210. В отчете по переписному листу (рисунок 21) может быть указан номер переписного листа, включая все обозначенные географические коды. Затем в отчете могут быть перечислены ошибки, найденные в программе по элементам данных (в данном случае «возраст») и по номеру строки в программе пакета программного обеспечения, указанном внизу справа. В данном примере графа «возраст» была незаполнена, но матрица вменения выдала возраст 48 лет на основе степени родства и пола данного лица. В этом случае конкретный возраст был неизвестен, но команда «message» могла по желанию указать и эту информацию.

Рисунок 21

Выборочное сообщение об ошибках в переписном листе

Идентификационный номер переписного листа: 01 01 017		Номер строки
AGE (1) =	Возраст неизвестен, поэтому вменяется	46
AGE (1) = 48	Возраст уже известен	

211. Хотя представляется целесообразным составить список всех отдельных ошибок по выборочным тестам или небольшим отдельным наборам данных, очевидно, что в ходе фактических прогонов размер массива выходных данных будет чрезвычайно большим и неудобным для работы (и приведет через некоторое время к бессмысленным результатам). Во избежание такого исхода необходимо установить триггер для «отключения» всех или некоторых проблем отдельных опросных листов для завершения переписи. Сводные статистические данные должны, безусловно, сохраняться.

с) Заказная распечатка ошибок

212. В программном обеспечении может быть еще одна команда, позволяющая проводить более подробный анализ спецификаций редактирования и

потока редактируемых данных. Данную команду можно использовать для того, чтобы показать информацию до внесения изменений, а затем все внесенные изменения. И наконец, она показывает запись или записи уже после произведенных изменений. Таким образом, аналитик может удостовериться, что программа редактирования надлежащим образом проходит по всем путям доступа. Результаты могут быть такими, как показано на рисунке 22. В первой строке выходных данных даются переменные (например, провинция, степень родства, пол, возраст). Затем показаны поступающие данные, потом ошибки (в данном случае отсутствие категорий «возраст»), а затем данные после изменения.

Рисунок 22

Пример дополнительной распечатки ошибок по переписному листу, включая множественные переменные

	Провинция	Округ	Глава домохозяйства	Степень родства	Пол	Возраст
Входящие данные	01	01	17	1	1	
Ошибка	Возраст неизвестен, поэтому вмененный возраст = ПРОПУСК					
Отредактированные данные	01	01	17	1	1	48

213. Такая процедура помогает группе редактирования определить, правильно ли проходит по путям доступа программа редактирования.

214. Такой вид проверки является важной частью редактирования данных переписи и обследования. Описанный ниже метод представляет собой один из возможных способов проверки процедур редактирования. Этот процесс можно начать с проведения специалистами систематического анализа путем создания «идеального» домохозяйства. Идеальное домохозяйство — это полное домохозяйство — глава домохозяйства, супруг(а), дети, прочие родственники и лица, не являющиеся родственниками, — со всеми своими характеристиками. Идеальное домохозяйство должно пройти все стадии редактирования без всяких ошибок. Процедура совершается следующим образом:

- a) обработчики данных вводят ошибки в каждое домохозяйство последовательно в соответствии с порядком спецификаций редактирования и программы редактирования;
- b) затем аналитик проверяет все пути доступа на ранней стадии редактирования;
- c) после того как программа редактирования пройдет надлежащим образом все пути доступа, обработчики данных прогоняют выборку из всего набора данных в поиске отклонений в фактическом наборе данных и с внесением изменений, если это необходимо;
- d) в заключение обработчики данных прогоняют весь набор данных.

215. Убедившись, что сообщения работают нормально и соответствующие изменения сделаны, обработчик данных может по своему усмотрению отключить их на более низких уровнях (например, на уровне каждого переписного листа). Если в больших странах прогонять все наборы данных целиком, оставляя функции оператора сообщений для каждого переписного листа, то получающееся в результате количество строк и перфоленты будет запретительным. Тем не менее сводный отчет по этим сообщениям следует продолжать, поскольку он дает полезную информацию по различным уровням территори-

ального распределения. При этом выходные данные будут выглядеть примерно так, как показано на рисунке 22.

216. Компьютерное редактирование, как правило, предусматривает процедуру сохранения. В журнале редактирования показаны все изменения данных и единицы счета для случаев изменения и заменяемых величин. Обращение к журналу редактирования помогает определить, является ли количество изменений достаточно низким для группы записей, которые надо учитывать.

217. Если в конкретном элементе данных слишком много ошибок, то этот элемент, возможно, не был должным образом проверен либо сам по себе, либо в связи с другими элементами, и это указывает на то, что счетчики или респонденты не поняли данный элемент. Иногда счетчики путаются и, например, собирают информацию о рождаемости только у взрослых мужчин, а не у женщин. Если данные будут собираться таким образом систематически, группа редактирования может попросить программистов переместить данные о рождаемости из рубрики «мужской» в рубрику «женский», если речь идет о супружеской паре. В противном случае группа редактирования вряд ли сможет исправить ошибку на этом этапе.

218. Как правило, программу редактирования необходимо прогнать через несколько разных файлов, с тем чтобы охватить все ситуации. Кроме того, обработчикам данных потребуется внести изменения в связи с синтаксическими или логическими ошибками. Даже наиболее опытные специалисты по обработке данных иногда вводят с клавиатуры знак «более» вместо «менее», и эта ошибка обнаруживается только после нескольких прогонов, поскольку конкретная проблема может сразу и не проявиться. Также могут сразу не проявиться и незначительные дефекты логики. Здесь опять специалисты по содержательной части переписи и специалисты по обработке данных должны совместно решать эти проблемы по возможности в самом начале процесса редактирования.

d) Сколько раз следует прогонять программу редактирования?

219. Как отмечалось, сразу же после составления опросного листа необходимо приступить к разработке и тестированию спецификаций и программ редактирования. При использовании нисходящего подхода различные элементы данных должны составляться по отдельности, однако, даже когда возникает необходимость в одновременном редактировании нескольких переменных, схему редактирования отдельных элементов данных следует протестировать на небольших частях общего набора данных. Спецификации редактирования должны разрабатываться специалистами по содержательной части, а затем отдельные программы редактирования должны составляться программистами. В этом случае можно осуществлять полную схему редактирования и прогонять ее по все более крупным частям набора данных с параллельной доработкой.

220. Как разъясняется ниже, и отдельные части программы редактирования, и всю программу в целом целесообразно прогонять три раза.

221. При первом прогоне программы редактирования в матрицы вменения вводятся реальные значения, а не те величины, которые созданы в первоначальной статической матрице. В некоторых странах используются данные из других источников — из предыдущей переписи или обследования, или из

административной отчетности — чтобы ввести в массив значения «cold deck». Обработчик данных прогоняет весь набор данных или значительную его часть, чтобы получить значения для матрицы вменения. Вероятно, такие значения «cold deck» из фактического набора данных будут более точными и обновленными. При редактировании используется лишь около двух процентов этой первоначальной статической матрицы: остальные значения выводятся путем динамического вменения.

222. При втором прогоне программы редактирования осуществляется собственно редактирование. Второй прогон состоит из нескольких повторных прогонов, для того чтобы охватить все ситуации. В это время обработчикам данных необходимо внести изменения, для того чтобы исправить погрешности, появившиеся в результате синтаксических или логических ошибок. Кроме того, даже наиболее опытные специалисты по обработке данных могут совершать ошибки и, поскольку конкретная проблема может проявиться не сразу, ошибка может быть обнаружена только после нескольких прогонов. Точно так же могут сразу не проявиться и незначительные дефекты в логике.

223. Третий прогон программы редактирования делается для того, чтобы убедиться, что 1) в наборе данных не осталось ошибок и что 2) программа редактирования не внесла новых ошибок. После того как обработчики сделают последний редакционный прогон, в выходных результатах ошибок уже не должно быть. Если они остаются, то, вероятно, нарушена логика редактирования, поэтому обработчик данных должен ее изменить. Кроме того, этот прогон обычно подсказывает обработчику данных, не внесены ли случайно при редактировании новые ошибки логикой схемы редактирования.

5. Флажки вменения

224. Флажки вменения представляют собой метод, используемый, в частности, для того, чтобы сохранять информацию о неотредактированных данных. Как отмечалось ранее, многие группы редактирования беспокоит потеря потенциальной информации при замене неотредактированных ответов. В тех случаях, когда значение меняется из-за несогласованности данных, группы редактирования могут по своему усмотрению сохранить первоначальное значение или значения для проведения после переписи дополнительного демографического анализа или анализа ошибок. Как специалисты по содержательной части переписи, так и программисты захотят проанализировать различные аспекты недостающих, недостоверных или несогласующихся данных. Члены группы редактирования должны удостовериться, что вмененные и невмененные распределения непротиворечивы, и посмотреть, есть ли в плане редактирования и вменения какая-либо систематическая ошибка. Например, иногда специалисты по обработке данных случайно используют только значения «cold deck», потому что программа не проводит обновления матрицы вменения. Если в стране проводится предварительное тестирование переписи, то группе редактирования, возможно, потребуется изучить взаимосвязи между некоторыми переменными после предварительного тестирования, для того чтобы завершить составление переписного листа. Во время предыдущих переписей, еще до широкого распространения микрокомпьютеров с жесткими дисками большой емкости, многие статистические управления не имели достаточно места на своих лентах или других носителях информации для сохранения дополнительных

данных. Сейчас же для большинства стран хранение информации о неотредактированных данных уже не представляет никакой проблемы.

225. Некоторые страны в качестве флажка для каждого элемента данных предпочитают использовать простую, двоичную учетную переменную. Этот метод прост и занимает один байт для каждой переменной. Например, во время переписи 1990 года Бюро переписи США разместило флажки вменения для каждой переменной в конце каждой записи и в сообщениях о жилье и о населении. Например, для каждой переменной по жилому фонду величина для признака была сначала «0», а затем менялась на «1», если каким-то образом менялся исходный элемент данных. Программа не сохраняла первоначальные значения, хотя статистические управления иногда их компилировали или для каждой записи, или в совокупности.

226. Для сохранения неотредактированных ответов существуют и другие методы. В примере, приведенном на рисунке 23, национальное бюро переписи/статистическое управление изменило возраст супруга с 70 на 40 лет, используя матрицу вменения. Национальное бюро переписи/статистическое управление может легко поставить значение, которое было до вменения (в данном случае 70 лет), там, где зарезервировано место для флажков вменения, и зарезервировать переменную, используемую в публикуемых таблицах для распределенного значения, в данном случае 40. Для изучения изменений в наборе данных статистическое управление может составить плотность распределения вероятности или комбинационные таблицы распределенных и нераспределенных величин. Если после такого анализа усилий по редактированию набора данных таблицы, составленные с учетом редакционных изменений, окажутся подозрительными или неверными, группы редактирования могут изменить схему редактирования или часть редакционного потока; кроме того, в связи со значительным увеличением емкости жестких дисков в последние годы все первичные значения могут сохраняться в архиве для последующего использования. Статистические управления, вероятно, захотят сохранить по крайней мере два файла, поскольку файл, содержащий только отредактированные данные, может быстрее переводиться в таблицы.

Рисунок 23

Выборочные демографические записи с признаками для вменяемых значений

Лицо	Пол	Возраст	Дети, родившиеся живыми (СЕВ)	Флажок пола	Флажок возраста	Флажок СЕВ
1	1	40	ПРОПУСК			1
2	2	40	7		70	

227. На рисунке 24, ниже, показан случай, когда девочка 13 лет зарегистрирована как родившая ребенка (в графе «дети, родившиеся живыми» стоит «1»). Однако группа редактирования решила, что минимальный возраст для первого рождения будет 14 лет и что рождение детей лицами женского пола младше 14 лет, вероятнее всего, ошибка, а не факт. Как всегда, при этом возникает вопрос, не представляет ли этот случай искажение в наборе данных по сравнению с реальным значением.

228. В соответствии с правилами редактирования информации о родившихся детях при вменении ставятся «пропуски». Отметим, что установка флажка СЕВ несколько сложнее, поскольку он должен учитывать вмененную

Рисунок 24

Пример флажка для молодой женщины с пробелом в графе «рождаемость» и добавлением флажка

Лицо	Пол	Возраст	Дети, родившиеся живыми (СЕВ)	Флажок пола	Флажок возраста	Флажок СЕВ
С пропуском в графе «рождаемость»						
4	2	13	1			
С пропуском в графе «рождаемость» и добавлением флажка						
4	2	13	ПРОПУСК			1

позицию *BLANK* (*ПРОПУСК*), а также цифровые записи. Предположим, что специалисты по содержательной части переписи хотят изучить числовые показатели и характеристики лиц в возрасте 13 лет, о которых сообщается, что у них есть ребенок. Обработчики данных могут зарегистрировать исходную информацию на участке записи, оставленном для флажков, как правило, в конце записи. После этого в публикуемых таблицах в отношении этой женщины не будет информации о всех родившихся детях, но такая информация тем не менее будет доступна для последующих исследований. На более позднем этапе, в частности при планировании последующего обследования или следующей переписи, группы редактирования могут использовать информацию о детях, родившихся у 13-летних женщин, чтобы решить, можно ли им снижать регистрационный возраст.

229. Одна из проблем, связанных с использованием флажков вменения, заключается в том, что вышеописанная процедура занимает достаточно много места в компьютере. Если флажки повторяют каждую переменную, то набор отредактированных данных будет приблизительно в два раза больше набора неотредактированной информации. Для многих стран это может быть неприемлемым с точки зрения длительного хранения данных. Однако исходные данные и отредактированная информация могут храниться для последующего воспроизведения.

230. Страны с очень большой численностью населения, возможно, предпочтут использовать флажки вменения на выборочной основе для целей научных исследований. Например, какая-либо страна может принять решение создать набор данных с каждой сотой жилищной единицей. Затем начнется редактирование этого уменьшенного набора с флажками вменения, с тем чтобы оценить, как такое редактирование влияет на качество данных, и определить различия между неотредактированными и отредактированными данными.

Г. Другие системы редактирования

231. Данное *Руководство* посвящено в основном описанию нисходящих методов компьютерного редактирования данных переписи и обследования. Некоторые страны используют другую, более сложную процедуру компьютерного редактирования, известную как редактирование переменных по групповому принципу (см. раздел [D.2](#), выше). Впервые этот метод был разработан Фелледжи и Холтом (Fellegi and Holt, 1976), и он обычно применяется к наиболее важным переменным в переписи или обследовании: возрасту, полу, степени родства и семейному положению. Однако его можно применять и к любой

группе переменных или ко всем переменным, указанным в опросном листе переписи или обследования. При такой методике программа редактирования смотрит ответы на эти вопросы одновременно для одного лица или для всех членов домохозяйства, для того чтобы выявить недостающие или несогласующиеся ответы. При обнаружении неизвестных (пропущенных), недостоверных или несоответствующих записей, при помощи серии тестов определяется, какой из отобранных элементов содержит наиболее значительные ошибки, и такой элемент меняется первым. Затем тесты повторяются, для того чтобы определить, не осталось ли недостоверных и несоответствующих значений; если таковые имеются, программа редактирования меняет элемент с наибольшим количеством оставшихся проблем. Эта процедура повторяется до того момента, пока не останется ошибок.

232. Принцип Фелледжи–Холта получил дальнейшее развитие в Статистическом управлении Канады, которое использовало его при проведении переписей у себя в стране с 1976 по 1991 год. Для переписи 1996 года в Канаде этот метод был еще более усовершенствован и получил название «новая методология вменения» (НМВ). Он впервые позволил «одновременное осуществление вменения цифровых и качественных переменных с минимальными изменениями для решения крупных задач [редактирования и вменения (Bankier, Houle and Luc)].

233. Если в процессе редактирования используется традиционный метод динамического вменения, или метод «hot deck», то вмененная информация для серии вопросов переписного листа может касаться многих различных лиц в зависимости от того, какая информация использовалась для обновления матрицы вменения. Например, если данные, касающиеся пола, степени родства и семейного положения лица А правильные, то соответствующие матрицы вменения будут обновляться этими величинами. Если возраст лица А не указан или недостоверен, он, естественно, не будет использоваться для обновления матриц вменения. Фактически обновление этой величины произойдет за счет других элементов. Таким образом, если у следующего лица показатель пола будет несоответствующим, и элемент данных «пол» вменяется, то данные относительно пола будут браться у лица А. Если еще неизвестен и возраст, то программа редактирования будет использовать возраст какого-либо другого лица.

234. В «новой методологии вменения» для элементов используются доноры в надежде, что вся недостающая или несогласующаяся информация может быть взята у одного или у нескольких доноров. Для того чтобы получить всю или большую часть информации от одного донора, все записи данных должны храниться в памяти компьютера. И тогда, если неизвестны или недостоверны и возраст, и пол, одна и та же занесенная в память переменная даст значения для обоих этих элементов.

235. Цели и задачи автоматизированной методологии вменения «hot deck» должны быть следующими:

- a) домохозяйство, данные по которому вменяются, должно иметь тесное сходство с домохозяйством, данные о котором не удалось отредактировать;
- b) подлежащие вменению данные о домохозяйстве должны поступать по возможности не от двух или более доноров, а от одного. Кроме того, подлежащие вменению домохозяйства должны иметь тесное сходство с этим одним донором;

- с) одинаково хорошие приемы вменения, основанные на данных об имеющихся донорах, должны иметь одинаковые шансы быть отобранными, для того чтобы избежать ложного увеличения численности небольших, но важных групп населения (Bankier, Houle and Luc).

236. По новой методологии вменения по ближайшему соседу эти цели достигаются путем выявления в первую очередь прошедших редактирование домохозяйств, максимально близких по своим характеристикам домохозяйству, данные о котором отредактировать не удалось. Это означает, что два домохозяйства должны быть равными по максимально возможному числу качественных переменных, имея лишь небольшие различия между цифровыми величинами. Домохозяйства с такими характеристиками называются «ближайшие соседи». Следующий шаг заключается в том, чтобы определить для каждого ближайшего соседа наименьшие подмножества несовпадающих переменных (и цифровых, и качественных), которые позволили бы в случае вменения отредактировать данные о домохозяйстве. Затем в произвольном порядке отбирается одна из этих операций по вменению, которая проходит через редактора и напоминает как не прошедшее редактирование домохозяйство, так и прошедшие редактирование домохозяйства (Bankier, Houle and Luc).

237. В данной главе были рассмотрены общие процедуры редактирования и составления таблиц. Глава III охватывает вопросы редактирования структуры данных, включающего первую и наиболее важную задачу компьютерного редактирования, поскольку такое редактирование устанавливает, что каждое домохозяйство упоминается только один раз и в надлежащем месте в рамках иерархии той или иной страны.

Глава III

Редактирование структуры данных

238. При редактировании структуры данных проверяется охват и определяется, как согласуются друг с другом различные записи. Такое структурное редактирование должно обеспечить, чтобы *a)* все записи о домохозяйствах и помещениях для коллективного проживания на счетном участке были указаны, причем в надлежащем порядке; *b)* по всем занятым жилищным единицам имелись записи о проживающих там лицах, а по незаселенным единицам записей о

Вставка 3

Руководящие принципы редактирования структуры данных

Редактирование структуры данных должно выполнить следующие задачи:

- ☞ обеспечить, чтобы каждая группа счетных участков (СУ) имела правильные географические коды (провинция, район, СУ и т. д.) и чтобы для наименования групп СУ использовалась единая практика;
- ☞ обеспечить, чтобы была включена каждая жилищная единица и чтобы все домохозяйства были включены в соответствующие СУ;
- ☞ объединять домохозяйства в соответствующие СУ, а СУ — объединять в соответствующие единицы более высокого территориального уровня;
- ☞ способствовать правильному выбору между страницами, посвященными отдельным лицам, и страницами, посвященными домохозяйствам, в буклетах, переписных листах или вне их, исходя из численности населения и компоновки переписного листа;
- ☞ причислять каждую индивидуальную запись к ее достоверному типу записи;
- ☞ обрабатывать записи о групповом жилье и о многоквартирных домах отдельно от жилищных единиц;
- ☞ обеспечить соответствие между различными видами записей: например, по незаселенным жилищным единицам отсутствуют записи о лицах, по занятым единицам есть по крайней мере запись об одном лице. Обеспечить, чтобы количество записей о лицах для каждого домохозяйства соответствовало общей подсчитанной численности домохозяйств в записях о жилье. Обеспечить правильное количество переписных листов, когда для одного домохозяйства используется много документов, и их надлежащую увязку;
- ☞ устранить двойные записи внутри домохозяйств (двойные лица) и между домохозяйствами (двойные домохозяйства или части домохозяйств), чтобы избежать чрезмерного охвата;
- ☞ обрабатывать записи с пропусками в рамках одного вида записи;
- ☞ обрабатывать недостающие жилищные единицы.

жильцах быть не должно; с) домохозяйства не должны иметь двойных или недостающих записей о лицах; и d) счетные участки не должны иметь ни двойных, ни недостающих записей о жилом фонде. Таким образом, при редактировании структуры проверяется, чтобы переписные листы были в целом заполнены.

239. Конкретные приемы редактирования структуры данных, используемые для одной переписи или обследования, возможно, придется со временем менять, поскольку технология, применяемая для определения и исправления ошибок в структуре, меняется весьма быстрыми темпами. Исходя из этого в данной главе рассматривается более общий вопрос достоверности и взаимосвязи элементов данных в записях и между ними. В главах IV и V говорится о конкретных отдельных элементах переписи населения и жилого фонда.

А. Редактирование географических параметров

1. Местоположение жилых помещений (населенный пункт)

240. Согласно второму пересмотренному изданию *Принципов и рекомендаций в отношении переписи населения и жилого фонда* (United Nations, 2008, para. 2.78), населенный пункт определяется как «четко выделяемая концентрация населения..., где жители проживают в граничащих друг с другом совокупностях жилых помещений и у которой есть свое название или признанный на месте статус». Дополнительную информацию, касающуюся месторасположения жилых помещений, можно найти в определениях «населенный пункт» и «городское и сельское население» в пунктах 2.78–2.88 *Принципов и рекомендаций*. Лицам, занимающимся проведением переписи жилого фонда, необходимо изучить эту информацию, поскольку географические понятия, используемые для описания местонахождения жилых помещений при проведении переписей жилого фонда, чрезвычайно важны как для проведения самой переписи, так и для последующего табулирования полученных результатов переписи (United Nations, 2008, para. 2.455).

241. При редактировании данных о местонахождении географические коды должны быть абсолютно точными. Получение полных, точных кодов для географической иерархии при обработке данных является одной из труднейших задач всей переписи. Если территориальная принадлежность будет закодирована неправильно, то операторы, вводящие данные, могут приписать жилищную единицу или единицы какой-то другой части страны. Зачастую исправить такого рода ошибку очень сложно.

2. Постоянное проживание в городских и сельских районах

242. Традиционные различия между городскими и сельскими районами в рамках одной страны основываются на том утверждении, что городские районы, независимо от того, как они определяются, предполагают другой образ жизни и обычно более высокий уровень жизни по сравнению с сельскими районами. Во многих промышленно развитых странах это различие стирается, и основное различие в условиях жизни между городскими и сельскими районами все больше выражается в степени концентрации населения. Несмотря на то что разница в образе и уровне жизни между городскими и сельскими районами в развивающихся странах остается значительной, быстрая урбанизация в этих

странах вызвала большую потребность в информации, связанной с различиями в размерах городских районов (United Nations, 2008, para. 2.82).

243. Большинство стран еще до проведения переписи определяют, какие географические районы считать «городскими», а какие — «сельскими», а после сбора данных переписи вводят необходимые коррективы. Если страна присваивает коды при постоянном проживании в городских и сельских районах (например, 1 для городских и 2 для сельских районов), то эти коды можно ввести при наборе данных с клавиатуры или можно определить во время редактирования, исходя из критериев, установленных группой редактирования. Если группа редактирования подготовит список территориальных единиц с разбивкой на городские и сельские, то обработчики данных могут легко ввести соответствующие коды в записи о жилом фонде.

244. Следует приложить усилия к тому, чтобы характеристики населения в целом соответствовали конкретному счетному участку. Например, в ряде стран некоторые категории квалифицированных специалистов, за исключением врачей, учителей и лиц с аналогичным родом занятий, практически не проживают в сельских районах, так же как работники фермерских хозяйств практически никогда не проживают в городских районах. Группа редактирования должна произвести проверку, с тем чтобы убедиться, что географический район классифицирован правильно.

В. Проверка охвата

1. Регистрация де-факто и де-юре

245. *Определение постоянного места жительства.* Как правило, «постоянное место жительства» определяется в целях переписи как место, где проживает лицо на момент проведения переписи и где это лицо находилось в течение некоторого времени или намерено остаться на некоторое время. Как правило, большинство регистрируемых лиц не переезжали в течение определенного времени, и, следовательно, определение их постоянного места жительства является простой задачей. Для других лиц применение такого определения может привести к многочисленным интерпретациям, особенно при частой смене места жительства таким лицом. Странам при определении постоянного места жительства рекомендуется применять пороговый период времени в 12 месяцев, в течение которых постоянным местом жительства является: *a)* место, в котором данное лицо проживало постоянно в течение большей части последних 12 месяцев (то есть по крайней мере шесть месяцев и один день), не считая временного отсутствия в отпуске и командировке, или в котором лицо намерено прожить не менее шести месяцев или *b)* место, в котором лицо проживало постоянно в течение по крайней мере последних 12 месяцев, не считая временного отсутствия в отпуске и командировке, или в котором лицо намерено прожить не менее 12 месяцев (United Nations, 2008, paras. 1.461–463).

246. С учетом этого определения постоянного места жительства, национальные бюро переписи/статистические управления стремятся собирать данные переписи де-факто (в тех местах, где люди находятся в ночь переписи) или де-юре (в тех местах, где они обычно находятся). При проверке взаимосвязи между записями о жилье, в частности, в связи с учетом лиц в жилых помещениях, и записями об отдельных лицах программа редактирования должна учиты-

вать вид переписи. Иногда страны собирают информацию и де-факто, и де-юре. Тот или иной элемент по каждому лицу может означать, проживает ли он/она 1) постоянно; 2) находится там временно, а обычно проживает в другом месте; или 3) обычно проживает в данном домохозяйстве, но временно отсутствует. В таблицах с данными де-факто используются только элементы 1 и 2, если присутствуют все три вида записей; в таблицах с данными де-юре используются только элементы 1 и 3, если присутствуют все три вида.

247. Национальные организации по проведению переписи/статистические управления, пользующиеся этими категориями, должны использовать их весьма осторожно на любом этапе — во время сбора и обработки данных, распространения и анализа. В частности, таблицы по конкретному набору данных необходимо составлять с учетом желаемой категории населения. Пользователи этих трех категорий должны знать об отобранной совокупности, поскольку анализ всего набора данных приведет к двойному включению в записи некоторых лиц. Если требуется совокупность де-факто, из таблицы необходимо исключить категорию 3 — лица, временно отсутствующие; если необходима совокупность де-юре, из таблицы необходимо исключить категорию 2. Во время составления первоначальных таблиц, таблиц для распечатанных отчетов и дополнительных носителей группа редактирования может выбрать для обработки подмножество из всего набора данных. Для последующих таблиц в документации файлов следует четко указывать, каким образом обрабатывать все возможные варианты данных. Более подходящим может стать применение многих файлов.

248. Программа редактирования должна быть составлена таким образом, чтобы обеспечить, что при наличии всех трех видов записей данные надлежащим образом будут согласовываться между собой. Если в записях де-факто есть несколько ответов, это может указывать на то, что на самом деле речь идет об отсутствующих на данный момент постоянно проживающих лицах или возникает другая связанная с регистрацией проблема, которая может потребовать специального решения.

2. Иерархия домохозяйств и жилищных единиц

249. В главе V рассматривается вопрос о взаимосвязях между домохозяйствами, жилищными единицами и жилыми помещениями. Применение этих понятий зависит от каждой отдельной национальной организации по проведению переписи/статистической организации. Однако прежде чем перейти к редактированию отдельных жилищных единиц, группа редактирования должна разработать методы проверки, чтобы выяснить соблюдение иерархии при сборе данных и их введении с клавиатуры.

3. Фрагменты переписных листов

250. Прежде чем редактировать элемент за элементом, компьютерная программа должна проверить достоверные записи, недостающие записи и двойные номера строк в рамках редактирования структуры данных. Она должна также определить, относятся ли редактируемые записи к лицам, проживающим в жилых помещениях группами. Операторы могут сделать ошибку при вводе данных, а иногда они могут забыть исключить обрывочную информацию (части записей). Одна из функций предварительного редактирования должна состоять в том, чтобы посмотреть файл обрывочных записей, для того чтобы

исключить их. Наиболее распространенным случаем будет запись, содержащая географические коды, но без информации о населении или жилом фонде.

С. Структура записей о жилом фонде

251. Одна из тем, которую можно включить в опросный лист для сбора информации при проведении национальных переписей жилого фонда, — это количество жилищ в строении. В данном случае единицей учета является строение и собираются сведения о количестве обычных и основных жилищ в нем (см. United Nations, 2007, para. 2.524).

252. Термин «общее редактирование» относится к практике обеспечения соответствия количества жилищных единиц, входящих в состав строения, общему количеству жилищных единиц в записи о жилом фонде. В случае несоответствия количество жилищных единиц, указанных в качестве одной из характеристик конкретного строения, следует откорректировать, с тем чтобы оно соответствовало количеству записей о жилищных единицах. Если данное строение закодировано как имеющее пять жилищных единиц, а фактически в отдельных записях по этому строению указаны четыре жилищных единицы, группа редактирования должна решить, какую поправку внести: *a*) изменить первую цифру на основе учета отдельных записей (что в большинстве случаев более приемлемо); или *b*) ввести еще одну запись, используя информацию об имеющихся записях (чего следует избегать).

Д. Соответствие между записями о жилом фонде и населении

253. Если перепись или обследование включают записи как о населении, так и о жилом фонде, то при редактировании структуры данных необходимо удостовериться, что эти два вида записей согласуются между собой.

1. Свободные и занятые жилища

254. Данные о свободной жилищной единице не должны содержать записей о заселении, а по занятой жилищной единице такие записи должны быть. Если имеются записи о заселении, а жилище зарегистрировано как незаселенное, то из категории «свободное» оно будет переведено в категорию «занятое». Иногда запись составлена так, что статус свободного помещения и условия владения входят в один и тот же элемент данных, поэтому такую информацию необходимо учитывать и при принятии решения. Кроме того, если есть ответ в графе «стоимость единицы» для единиц, занимаемых владельцем, или «арендная плата» для единиц, занимаемых арендатором, то программы редактирования используют в своих определениях эту информацию; в противном случае может потребоваться матрица вменения.

255. Если там, где предполагается, что жилищная единица занята, никаких записей о заселении нет, то группа редактирования должна решить, считать ли ее за свободную единицу или подставить данные о лицах из другой единицы. Если данная единица свободна, то методом вменения легко заменить эту переменную на статус свободной единицы. Однако, если жилищная единица занята, группа редактирования должна решить, стоит ли и как приписывать

лиц из другой единицы, по возможности с тем же числом жильцов и аналогичными характеристиками. Если нельзя узнать характеристики неуказанных лиц, этот метод следует использовать, только когда группа редактирования решит, что у нее нет другого выхода, или вообще отказаться от него. Три возможных варианта изложены ниже.

a) Оставить жилищную единицу в категории «свободная»

256. В этом случае группа редактирования принимает решение о том, что свободные жилищные единицы, данные о которых поступают с мест, должны оставаться в категории «свободные», и, следовательно, никакие величины не вменяются. О редактировании данных, касающихся свободных жилищных единиц, говорится в главе V.

b) Повторное неоднократное посещение жилищной единицы для заполнения переписных листов

257. Национальное бюро переписи/статистическое управление может решить применить процедуры, требующие от счетчиков возвращения к данным о свободных жилищных единицах, пока они не убедятся в том, что они или свободны или заняты, и пока они не соберут по крайней мере минимальные характеристики. В этом случае группа редактирования должна разработать такие приемы, которые позволят проверить, свободна ли данная жилищная единица и достаточно ли характеристик, чтобы считать ее занятой. В зависимости от того, какую информацию группа редактирования считает «минимальной», применяется обычный метод редактирования, описанный в главе IV, или берутся данные из записей донора для «недостающих» лиц, о чем говорилось выше.

c) Замена данных об отсутствующих лицах данными из другой жилищной единицы

258. Процедуры замены данных о целых домохозяйствах или отдельных неуказанных лицах описываются в других подразделах настоящей главы. В соответствии с этими процедурами необходимо исходить из предположения о том, что неуказанные лица имеют те же характеристики, как и те, которые берутся для замены, что почти наверняка не всегда так, а сами процедуры очень сложны. Тем не менее без этих процедур подсчет общего числа лиц и числа лиц по отдельным характеристикам может дать более низкие показатели.

2. Двойной учет домохозяйств и жилищных единиц

259. Двойной учет жилищных единиц происходит по целому ряду причин. Иногда оператор по вводу конкретных данных может дважды ввести данные об одной и той же жилищной единице. Иногда разные операторы, вводящие данные, могут случайно повторно набрать на клавиатуре данные об одних и тех же жилищных единицах или даже целых счетных участках из-за отсутствия гарантий качества в национальном бюро переписи/статистическом управлении. И наконец, счетчик может неправильно записать территориальный код жилищной единицы, что ведет к двойному учету информации, когда две жилищные единицы имеют один и тот же территориальный идентификационный номер.

260. Если бюро отслеживает введенные пакеты данных, то дублирования не будет. Тем не менее следует разработать такую программу редактирования, которая исключит двойной учет домохозяйств, когда операторы по вводу данных дважды набирают одну и ту же информацию о домохозяйстве или домохозяйствах. Страны не должны сортировать свои данные до окончания структурной проверки и устранения проблем с двойными записями. До сортировки сотрудники могут исправить пакеты данных вручную; после сортировки сотрудники могут уже не обнаружить эту проблему. После сортировки данных программа редактирования может проверить двойные домохозяйства и путем вменения устранить последующие двойные записи.

3. Недостающие домохозяйства и жилищные единицы

261. Аналогичным образом после сортировки могут выявиться и недостающие домашние хозяйства. Например, программа редактирования предполагает получить последовательность домохозяйств в рамках самого низкого территориального уровня, например 1, 2, 3, 4, а получает только 1, 2, 4. Тогда необходимо принять решение изменить нумерацию этих единиц или найти какой-либо «приемлемый» метод замены единицы 3 другой единицей. Существует несколько способов добавления недостающих домохозяйств, когда становится очевидным, что они действительно отсутствуют и требуют добавления. Один метод — это простое дублирование предыдущего домохозяйства. Однако, если известно число жильцов в этом домашнем хозяйстве, что часто имеет место (даже если неизвестны их характеристики), существует возможность вернуться в начало и продублировать предыдущую единицу с таким же числом жильцов. Аналогичным образом, если известен возраст и пол членов домохозяйства, то такая информация может помочь в получении заменяющего домохозяйства. Не рекомендуется использовать вменение по методу «hot deck» для составления информации о членах домохозяйства, поскольку этот метод зачастую дает несоответствующие друг другу переменные.

4. Соответствие между числом жильцов и суммарным числом жильцов

262. Число жильцов, зарегистрированных в записях о данной жилищной единице, должно точно совпадать с общим числом членов домохозяйства. Программа редактирования суммирует число лиц, а затем сравнивает эту величину с числом жильцов в записи о жилищной единице. Если сумма отличается по значению от числа жильцов, необходимо или скорректировать число жильцов так, чтобы оно равнялось числу членов домохозяйства, или скорректировать отдельные записи. В главе V конкретно говорится о редактировании данных жилого фонда по числу жильцов.

а) Когда число жильцов больше суммарного числа жильцов

263. Если значение конкретной переменной в графе «число жильцов» в записи о жилищной единице превышает сумму записей об отдельных лицах, перед группой редактирования встает реальная проблема. Никто не может знать характеристики недостающих лиц. В связи с этим группа редактирования, решившая вменить характеристики недостающих лиц или заменить их данными из аналогичных домохозяйств, может столкнуться с дилеммой. Дан-

ные о недостающих лицах заменять не следует. Однако если данные об общем числе жильцов приемлемы, то в качестве альтернативы можно сократить численность зарегистрированного населения. Группа редактирования должна проанализировать всю картину в целом и затем найти подходящий путь.

264. Существует несколько способов обнаружения и замены недостающих записей, и ни один из них не является полностью удовлетворительным. Можно сохранять в памяти целые домохозяйства с различными важными характеристиками. Если появляется домохозяйство с несколькими, но не со всеми его членами, то можно поискать файл с данными о домохозяйстве, где все или почти все известные характеристики совпадают, а затем скорректировать данные о недостающих лицах на основе данных о других лицах из домохозяйства-донора. Однако программировать такую операцию весьма сложно, поэтому национальные бюро переписи/статистические управления, применяя такой подход, должны приступать к планированию задолго до проведения такой операции.

265. Разновидность этой процедуры состоит в том, чтобы пометить флажками все домохозяйства с недостающими записями и продолжать редактирование остальных данных. В конце процедуры редактирования, после того как исправлены все отдельные записи, группа редактирования может предложить специалистам по обработке данных просмотреть файл и внести добавления и изменения, используя полностью отредактированный набор данных. Используя такой нисходящий подход, группа редактирования может найти приемлемых доноров.

b) Проверка числа лиц по признаку пола

266. Иногда число жильцов в записи о жилищных единицах регистрируется по признаку пола. В этом случае программа редактирования должна суммировать число лиц по каждому полу отдельно. И опять же, если суммы отличаются от числа жильцов, одну из величин в каждом случае необходимо корректировать. Как правило, вместо добавления «недостающих» записей или исключения записей с полезной информацией суммарные данные о жилищных единицах корректируются, поскольку в этом случае ошибку, скорее всего, сделал счетчик в бланке опросного листа о жилищной единице.

c) Порядковая нумерация

267. Записи о населении должны быть расположены и пронумерованы по порядку. Такая нумерация должна указываться как переменная, например номер строки или порядковый номер в переписном листе. Кроме того, порядковые номера должны указываться в цифровой последовательности. Ошибки могут возникать из-за того, что иногда переписные листы или личные бланки оказываются испорченными, потому что счетчики собирают информацию в неправильном порядке или могут пропускать страницы, непреднамеренно оставляя пустые страницы в наборе данных. Хотя отсутствие последовательности, как правило, не влияет на процесс редактирования или составления таблиц, многие национальные бюро переписи/статистические управления предпочитают переставлять личные данные в правильном порядке. В связи с этим программа редактирования должна обладать способностью обнаруживать приводимые в неправильном порядке личные данные и вновь устанавливать их

последовательность. Поскольку иногда повторное упорядочение может повлиять на степень родства с главой домохозяйства, это необходимо предусмотреть в спецификациях редактирования. Повторное упорядочение определенно повлияет на такие переменные, как номер строки матери или номер строки мужа.

5. Соответствие между жильцами и типом строения/ домохозяйства

268. Вид родственных отношений между членами домохозяйства должен согласовываться с типом жилищной единицы. Иногда данные о членах домохозяйства появляются в данных о жилище, зарегистрированном как помещения для коллективного проживания или наоборот. В этих случаях вид родственных отношений или тип жилищной единицы должен указываться с учетом размера домохозяйства и других переменных.

Е. Двойные записи

269. Двойные номера строк вряд ли появятся при оптическом считывании или ином сканировании переписных листов. В отношении бланков, вводимых с клавиатуры, национальное бюро переписи/статистическое управление может по своему усмотрению проверить вручную соответствие между списком домашних хозяйств и номерами строк для домашнего хозяйства, которые вводятся с клавиатуры. Такая ручная проверка может повысить качество вводимых данных, особенно при сопоставлении 1) имен лиц на странице, где перечислены все члены домохозяйства, с 2) данными о лицах по колонкам, рядам или страницам. Два лица, данные о которых первоначально кажутся продублированными, могут на деле оказаться близнецами, если посмотреть их имена.

270. Во вводимых с клавиатуры переписных листах не должно быть дублирования номеров строки, если экраны данных или схемы пропусков настроены должным образом. Большинство современных пакетов программ создают порядковые номера автоматически при вводе данных. Ошибка может быть внесена, когда сотрудники вводят двойные данные для одного лица или когда ошибочный номер строки может создать двойную запись. По мере обработки каждой записи программа редактирования сопоставляет ее с предыдущими записями о заселении для конкретной жилищной единицы. При редактировании необходимо убедиться, что каждый номер строки записан правильно. Двойные номера строки являются ошибкой и подлежат изменению.

271. Страны могут по своему усмотрению разработать собственные схемы ввода данных с клавиатуры, не используя готовые пакеты. В этом случае группа редактирования должна выбрать приемлемый уровень ошибок. Для таких решений есть множество методов. Один из них заключается в следовании приведенным ниже руководящим принципам:

- a) если две разные записи имеют одинаковый номер строки, а число различающихся характеристик не превышает двух, программа редактирования устранил одну из записей, исходя из того, что она, вероятнее всего, двойная;
- b) если различаются три или более характеристик, номер строки будет изменен⁶.

⁶ Традиционно двойные записи отслеживались и корректировались вручную, однако этот процесс все больше и больше автоматизируется, по крайней мере частично. Винклер в своей недавней публикации (Winkler, 2006) начал совместно рассматривать вопросы автоматизации редактирования структуры и содержания данных.

F. Специальные совокупности

1. Коллективно проживающие лица

272. При редактировании структуры должны обрабатываться данные о лицах, проживающих в коллективах, таких как учреждения, казармы, дома для престарелых, в отличие от тех, кто живет в обычных жилищных единицах. Поскольку в коллективах, как правило, нет главы домохозяйства, страны должны определить, как лучше проводить различие между типами жилищных единиц. Один из методов состоит в том, чтобы использовать другой тип записи для коллективов. Еще можно присвоить конкретный код для категории «родственные отношения», что будет означать помещение для «группового» или «коллективного» проживания.

a) Когда для коллективов используется другой тип записи

273. Если национальное бюро переписи/статистическое управление принимает решение использовать отдельный тип записи, группа редактирования без труда сможет определить, какие записи относятся к коллективам или являются коллективными записями. Таблицы с данными о коллективах легко составить, используя непосредственно только эти записи. Переменные, которые характерны исключительно для коллективных записей, например тип коллектива, можно редактировать и вносить отдельно. Переменные, исключенные из коллективных записей, можно легко проверить, чтобы удостовериться, что они действительно отсутствуют. Однако файл получается более объемным, поскольку эти записи, вероятно, должны быть короче, чем обычные записи о заселении жильцов, но они занимают столько же места, сколько в прямоугольном файле. Кроме того, во время редактирования и внесения некоторые программы могут проверять определенные элементы в записях о заселении и в коллективных записях.

b) Когда переменная отличает коллективы от других записей

274. При использовании отдельной переменной вместо отдельного типа записи группе редактирования, возможно, будет труднее определить, какие записи относятся к коллективам или являются коллективными записями. В этом случае можно легко составить таблицы с данными о коллективах, опираясь лишь на саму переменную, которая отмечает, какие записи относятся к коллективно проживающим лицам. При этом переменные, относящиеся исключительно к коллективам, например тип коллектива, можно все еще редактировать и вносить отдельно. Переменные, которые исключаются из коллективных записей, можно легко проверить, чтобы удостовериться, что они фактически пусты, используя код для коллективов. Файл при этом получается более компактным, поскольку дополнительных записей о лицах, проживающих коллективно, не требуется, а они просто включаются в качестве записей о заселении с другим кодом для переменной домохозяйства/коллектива. При редактировании и внесении программа проверит только записи о заселении по некоторым элементам данных, а не будет проверять записи, касающиеся заселения и коллективные записи.

с) Когда отсутствует код «тип коллектива»

275. Код, обозначающий коллективы, может отсутствовать или быть недостоверным, или возможно несоответствие между кодом коллектива и кодами родственных отношений. Если код для коллективов отсутствует, а коды родственных отношений указывают на коллектив, то в этом случае предлагается соответствующим образом изменить код коллектива. При наличии кода коллектива в отсутствие кода родственных отношений последний можно определить по типу коллектива.

д) Когда есть код коллектива, но все лица находятся в родственных отношениях

276. Если есть код для коллектива, но все лица в данной жилищной единице состоят в родстве, если судить по кодам родственных отношений, то тогда код следует изменить, чтобы он обозначал жилищную единицу. Вместе с тем, если эта единица зарегистрирована как домохозяйство, но там нет двух лиц, состоящих в родстве, то может быть целесообразно в этом случае заменить ее на помещения для группового или коллективного проживания. В домохозяйстве может быть 5 или 6 не связанных родством лиц, и все равно это не будет коллективом. Как подчеркивалось выше, для разрешения особых, необычных случаев членам группы редактирования, возможно, необходимо проконсультироваться между собой.

е) Проведение различий между разными типами коллективов

277. В большинстве стран есть разные типы коллективов. Часто информация классифицируется еще подробнее по конкретным типам помещений для коллективного проживания. Эту информацию можно или кодировать отдельно в качестве элемента «тип коллективных помещений», или включать в самых разных вариантах в коды родственных отношений в домохозяйстве.

2. Сложные для регистрации группы населения

а) Сезонные мигранты

278. В некоторых странах с сезонной миграцией счетчику надо будет знать, свободна или занята та или иная жилищная единица в указанный срок. Так, если даже о домохозяйстве имеется полная информация, оно может быть также учтено (зарегистрировано) в другом месте. Естественно, может быть и обратный вариант. Домохозяйство, у которого есть два жилища в разных местах (таких жильцов иногда называют «перелетными птицами», поскольку они предпочитают жить в разных более подходящих для них регионах в различные сезоны года), может быть полностью пропущено, если не проследить за этим.

279. Иногда целые домохозяйства регулярно живут одну часть года в одном месте, а остальную часть года — в другом месте. Национальные бюро переписи/статистические управления и группы редактирования должны решить, как поступать в этих разных ситуациях. Например, некоторые лица проводят какое-то время ежегодно в другом месте, такие как жители более холодной части страны в теплое время года, и те, кто живет в теплой части страны в холодное время года. Еще один случай связан с кочевниками, которые часть года переме-

щаются, а часть года ведут оседлый образ жизни, — возможно, ту часть года, когда в стране намечена перепись.

b) Бездомные лица

280. По определению, регистрация какого-либо бездомного лица не дает информации о жилище. Однако «фиктивная» запись (новая запись, которая изначально включает пустые значения для некоторых переменных), облегчает структурную проверку и делает эту запись совместимой со структурой других жилищных единиц. Группа редактирования должна в этом случае решить, стоит ли создавать такую фиктивную запись о жилищах, с тем чтобы содействовать процессам обработки данных и составления таблиц.

c) Кочевники и лица, проживающие в труднодоступных местах

281. И вновь, как и в случае с бездомными лицами, структурная проверка может быть весьма затруднительной. В отдельных странах собирается некая «жилищная» информация, с тем чтобы эту информацию можно было использовать для помощи в редактировании структуры конкретной «единицы». Следовательно, редактирование данных о жилом фонде будет отличаться от того, которое используется для стандартных единиц. Информацию по вопросам заселения жильцов следует собирать так же, как в отношении лиц, проживающих в стандартных жилищных единицах, и редактировать обычными способами, следуя изложенным ниже принципам.

d) Гражданские постоянно проживающие лица, временно отсутствующие в стране

282. При проведении переписи с регистрацией де-юре гражданские лица — постоянные жители, временно отсутствующие в стране, но проживающие в домохозяйствах, которые могут дать информацию о них, должны включаться в стандартное редактирование данных о населении. В ходе переписи с регистрацией де-юре некий показатель должен выявлять временно отсутствующих лиц для возможности определения численности населения как де-юре, так и де-факто. Редактирование данных жилого фонда не будет отличаться по причине отсутствующих лиц. Очевидно, однако, что эти лица не будут включены в перепись де-факто; поэтому данные по ним не появятся в ходе редактирования данных о населении.

e) Гражданские иностранцы, не пересекающие границу ежедневно и находящиеся в стране временно, включая лиц без документов или лиц, временно находящихся на момент переписи на судах в порту

283. Поскольку при проведении переписи с регистрацией де-факто в нее должно включаться любое лицо-резидент, находящееся в стране на момент переписи, указанных выше лиц также необходимо включать в перепись. Лица должны переписываться по месту их проживания на момент переписи, и данные по ним должны редактироваться с применением стандартных приемов редактирования элементов данных по населению. Если данные о жилище не собираются для коллектива или любой нестандартной жилищной единицы, тогда для данных по этим лицам не должно проводиться и редактирование.

Если находящиеся в порту суда рассматриваются как жилищные единицы, тогда следует описать их жилые характеристики и отредактировать, используя информацию по другим судам для включения в данные «hot decks».

284. Иностранцы, находящиеся в стране лишь временно, по-видимому, не должны включаться в перепись с регистрацией де-юре. Лица без документов будут включаться в перепись, особенно в тех странах, где в рамках переписи не ведется отдельный учет лиц с документами и без документов (несмотря на то, что такой отдельный учет обычно дает более качественные результаты переписи). Лица, временно находящиеся в стране, не включаются в перепись с регистрацией де-юре после редактирования, если только они не направляются в конкретный местный район, но при этом, как правило, проживают в данной стране. Если какое-либо судно обычно находится в порту в этой стране, тогда находящиеся на таком судне лица должны, по-видимому, включаться в перепись как обычные резиденты и данные по ним должны редактироваться соответствующим образом.

f) Беженцы

285. Беженцы тоже могут находиться во временных помещениях, что может потребовать введения конкретной переменной, то есть отдельного типа записи или фиктивной записи о жилищах, чтобы учесть их состояние. Группа редактирования должна будет для этого разработать и применить соответствующие процедуры. Обычно в отношении элементов данных по населению и жилому фонду используются стандартные методы редактирования с помощью данных «hot decks», включая «жилые помещения беженцев» в качестве показателя.

g) Военнослужащие армии и флота, дипломатические сотрудники и их семьи, находящиеся за пределами страны, и иностранные военнослужащие армии и флота, дипломатические сотрудники и их семьи, находящиеся в стране

286. Для целей переписи с регистрацией де-юре в перепись, как правило, включаются военнослужащие армии и флота, дипломатические сотрудники и их семьи, находящиеся как в стране, так и за ее пределами. Во многих странах информация о военнослужащих в рамках переписи не собирается, и национальные статистические управления таких стран должны проводить простой подсчет или несколько подсчетов, располагая минимальной дополнительной информацией. При ограниченности информации использование данных «hot decks» затруднительно, и существует большая вероятность внесения ошибок в набор данных; поэтому обычно более предпочтительно не включать в перепись такого рода записи о домохозяйствах, в составе которых есть военнослужащие. Аналогичные проблемы могут возникнуть с дипломатическими сотрудниками. Тем не менее регистрация в рамках страны может дать хорошие результаты при использовании стандартных опросных листов и процедур; вследствие этого такие жилищные единицы должны включаться в обычное редактирование, хотя в данных должен иметься показатель особого статуса жилищной единицы. Поскольку находящиеся за пределами страны жилищные единицы стандартным образом регистрировать нельзя, следует позаботиться об оценке необходимости включения или невключения этих единиц в редактирование. (Их, тем не менее, можно включать в некоторые таблицы.)

287. Для целей переписи с регистрацией де-факто в перепись обычно включаются только находящиеся в стране жилищные единицы. Проживающие за пределами страны домохозяйства, в составе которых присутствуют военнослужащие армии и флота и дипломатические сотрудники, как правило, не включаются в перепись. Жилищные условия таких сотрудников обычно сообщаются лицами, проживающими в их жилищных единицах в стране, из которой они были направлены; включаются только фактически проживающие лица.

h) Гражданские иностранцы, пересекающие границу ежедневно для работы в стране

288. Гражданские иностранцы, пересекающие границу ежедневно для работы в стране, обычно не включаются ни в перепись де-юре, ни в перепись де-факто, поскольку они не проживают в данной стране на дату переписи. Как правило, они регистрируются в направляющей их стране в переписях и де-юре, и де-факто.

и) Гражданские постоянно проживающие лица, пересекающие границу ежедневно для работы в другой стране

289. Гражданские постоянно проживающие лица, пересекающие границу ежедневно для работы в другой стране, являются резидентами проводящей перепись страны и должны включаться в регистрацию и де-юре, и де-факто. Данные и по заселению и по жилому фонду по таким лицам редактируются стандартным образом.

j) Моряки торгового и рыболовного флота, постоянно проживающие в стране, но находящиеся в открытом море на момент переписи (включая лиц, не имеющих иного места жительства помимо своих помещений на борту судна)

290. Моряки торгового флота должны учитываться в переписи с регистрацией чистого де-юре (с коррекцией данных переписи для включения лиц, не имеющих другого места жительства), но не в переписи с регистрацией де-факто. При включении в перепись их жилищные единицы должны содержать ссылку на особый тип жилья, при этом должна быть возможность применения стандартных методов редактирования для элементов по заселению при использовании на судах стандартного национального опросного листа.

Г. Определение главы домохозяйства и супруги(а)

1. Редактирование переменной, касающейся главы домохозяйства

291. При определении членов домохозяйства сначала по традиции определяется глава домохозяйства или основное лицо, а затем остальные члены домохозяйства в соответствии со степенью родства с главой домохозяйства или основным лицом. Глава домохозяйства определяется как лицо, которое признается таковым другими лицами, входящими в это домохозяйство. Страны могут использовать тот термин, который они считают наиболее приемлемым для определения этого лица (в том числе «глава домохозяйства», «основное лицо

домохозяйства»), пока выявляемое таким образом лицо используется исключительно для определения степени родства между членами домохозяйства. Каждой стране рекомендуется представлять в публикуемых отчетах используемые понятия и определения (United Nations, 2007, para. 2.114).

a) Порядок родственных связей

292. Порядок родственных связей в жилищной единице влияет на результаты редактирования, поскольку многие редактируемые данные построены на том, что глава домохозяйства — это первое лицо, и его/ее данные будут редактироваться в первую очередь. Например, при редактировании данных о главе домохозяйства в первую очередь проверяются такие переменные, как язык, этническая принадлежность и религия. Если о главе домохозяйства имеется достоверная информация, выраженная любой из этих переменных, то она вменяется для получения данных о любом другом лице в этом домохозяйстве, которые отсутствуют, неправильно закодированы или неправильно введены с клавиатуры (см. главу IV). Данные о главе домохозяйства должны редактироваться в первую очередь, поскольку его/ее характеристики используются для назначения или вменения данных о других членах домохозяйства.

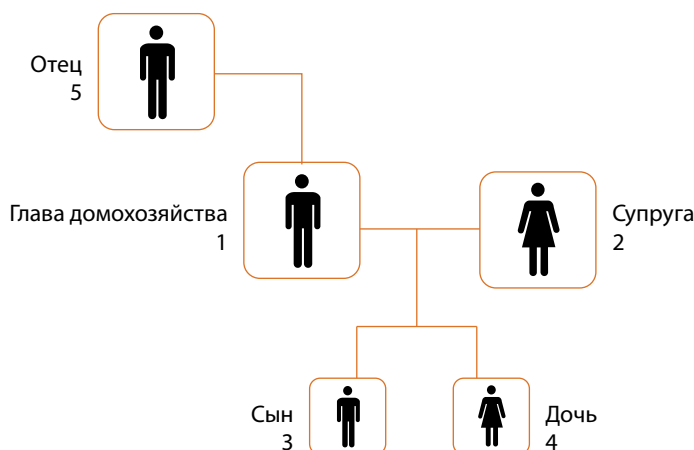
b) Когда глава не является первым лицом

293. На процесс редактирования влияют действия счетчиков на местах, которые они предпринимают исходя из различных ситуаций, возникающих в связи с определением главы домохозяйства. Для лучшего понимания этой проблемы рассмотрим сначала домохозяйство, показанное на рисунке 25.

294. На примере данного домохозяйства показана типичная ситуация, встречающаяся при работе на местах: глава домохозяйства и супруга, их дети и отец главы домохозяйства. Если счетчик собирает информацию именно таким образом, то редактирование, основанное на том, что глава домохозяйства занимает первую позицию в домохозяйстве, пройдет без проблем.

Рисунок 25

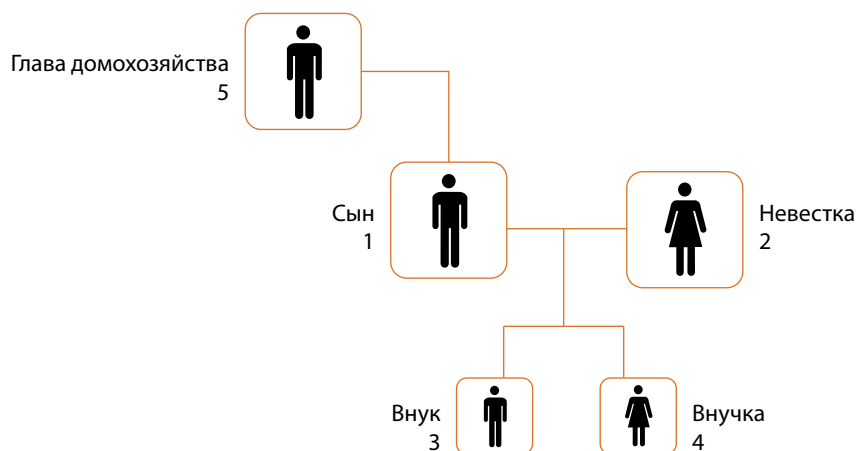
Пример домохозяйства, где глава домохозяйства зарегистрирован как первое лицо



295. Однако если регистрация проводится таким образом, что дедушка определяется как глава домохозяйства, то конфигурация родственных отношений меняется, как показано на рисунке 26. Такая ситуация может возникнуть, если счетчик, посетив дом, обнаруживает основной тип семьи в составе мужа, жены и двух детей и во время опроса отец главы домохозяйства входит в комнату и заявляет, что главой домохозяйства является он. Если взять за основу этого предполагаемого главы домохозяйства, то лицо 5 становится главой домохозяйства, лицо 1 — сыном, лицо 2 — невесткой и т. д.

Рисунок 26

Пример домохозяйства, где глава домохозяйства зарегистрирован как пятое лицо



296. Как становится ясно на примере этих двух домохозяйств, траектории редактирования, основанные на различных вышеуказанных главах домохозяйства, будут разными. Чтобы определить фактического главу домохозяйства для последующего редактирования и составления таблиц, есть три возможности: *a)* можно использовать указатель, чтобы отметить, кто является главой, и этот указатель можно использовать для дальнейшего редактирования и составления таблиц; *b)* если глава не зарегистрирован как первое лицо, его/ее можно переместить на первую позицию, а каждого из тех, кто находится в списке выше, можно переместить на одну позицию ниже; или *c)* можно изменить коды родственных отношений и сделать главой первое лицо, невзирая на другие родственные отношения.

i) Установление указателя для записи о главе домохозяйства

297. При редактировании данных о главе домохозяйства используется указатель для определения номера строки главы домохозяйства в жилищной единице. Если глава остается на выбранной позиции, указатель можно поставить на данную позицию, в этом случае всегда можно легко найти главу, когда это необходимо в конкретных целях редактирования или составления таблиц. Переменную «глава-указатель» («head-pointer») можно установить на номер строки главы домохозяйства и использовать во время редактирования, чтобы приписать или вменить недостающие или недостоверные характеристики других лиц в этой единице. Если глава является первым лицом в домохозяйстве, то значение переменной «глава-указатель» будет «1».

ii) *Как сделать главой первое лицо*

298. Группа редактирования может решить переместить главу на первую позицию в домохозяйстве. В этом случае программирование несколько сложнее того, которое требуется для операции, описанной в подпункте i, выше. Специалист по обработке данных должен разработать программу, которая переместит главу на первую позицию в списке, после него пойдет лицо, которое до этого было на позиции 1, затем лицо, которое было на позиции 2, и т. д., до тех пор пока не дойдет до лица, указанного непосредственно перед тем, кто был главой. Таким образом, если глава находится на позиции 5, порядок следования изменится с 1, 2, 3, 4, 5 на 5, 1, 2, 3, 4. После такого изменения глава будет на позиции 1, и дальнейшее редактирование пойдет легче, поскольку глава всегда будет на этой позиции. Тем не менее при выполнении такой операции целостности набора данных будет нанесен определенный «ущерб». Поскольку порядок лиц изменился, то аналитикам, возможно, будет трудно определить фактический порядок лиц, данные о которых были собраны на местах, и возможное воздействие такого порядка на толкование результатов.

iii) *Изменение кодов родства, чтобы первое лицо сделать главой домохозяйства*

299. Если группа редактирования решит, что первое указанное лицо должно быть главой домохозяйства, то при редактировании необходимо выполнять процедуры, указанные в подпунктах *a* и *b*:

- a) первому лицу присваивается значение главы домохозяйства;
- b) затем для корректировки данных о домохозяйстве применяется программа в целях изменения значений для других членов домохозяйства.

300. Например, на рисунке 26 родитель выступает в качестве главы домохозяйства. Когда лицо 1 становится главой домохозяйства, лицо 2 нужно указать как «супругу», лицам 3 и 4 нужно дать название «ребенок», а лицу 5 — «родитель» (как показано на рисунке 25). При этом понадобится еще стандартная подпрограмма для хранения матрицы с исходными и измененными значениями.

301. При выполнении такой процедуры целостность набора данных нарушается еще больше. Порядок лиц не меняется, как в предыдущем примере, и аналитикам будет трудно определить фактический порядок лиц, данные о которых собирались на местах. Тем не менее в данном случае все родственные отношения изменятся, и аналитики не будут знать, кто был указан первоначально в качестве главы домохозяйства. Кроме того, если в ходе переписи или обследования подбирались личные номера матери, отца или супруги, этот факт необходимо учитывать при проведении любой процедуры изменения нумерации. Вместе с тем если глава стоит на первой позиции, то таблицы номинально будет составлять легче. В отличие от предыдущего примера, программистам для этой процедуры не надо физически перемещать записи.

c) *Более одного главы*

302. Если обнаруживается более одного главы домохозяйства, то группа редактирования должна определить, кого назначить главой домохозяйства. Ре-

дактирование должно производиться на основе характеристик, установленных специалистами по содержательной части переписи, и самим ходом редактирования. Затем программа редактирования должна изменить степень родства другого лица (лиц), определенного(ых) в качестве главы домохозяйства.

303. Особая ситуация складывается в тех странах, которые допускают наличие «совместных глав домохозяйства» либо в силу определенных социально-экономических условий (например, частые отъезды мужчин — глав домохозяйств на горные разработки или для выполнения других видов деятельности, что предполагает оставление супруги в качестве главы домохозяйства), либо в силу того, что респонденты настаивают на «равноправии». Традиционно для целей редактирования важно назначить одного-единственного главу домохозяйства с сохранением в этом случае первичных данных в записи. При этом, однако, публикация Организации Объединенных Наций (United Nations, 2008, параг. 2.117) содержит положения о совместных главах домохозяйств. Если та или иная страна идет по пути включения совместных глав домохозяйств, эти данные должны сохраняться при редактировании; тем не менее при этом потребуются последующее изменение многих приемов редактирования, предлагаемых в настоящем *Руководстве*: в случаях, когда совместные главы домохозяйств придерживаются разного вероисповедания, принадлежат к различным племенам или имеют различные другие демографические и социальные характеристики, одно-единственное лицо более нельзя использовать для процедур вменения.

d) Если глава отсутствует

304. Аналогичным образом, если глава домохозяйства не выявлен, то при редактировании необходимо определить, кого назначить главой домохозяйства. В этом случае при редактировании, видимо, придется откорректировать степень родства других членов домохозяйства. При таком методе определения главы домохозяйства необходимо принимать во внимание такие переменные, как возраст, уровень образования и степень экономической активности, для того чтобы выделить наиболее приемлемую кандидатуру. Блок-схема типового редактирования данных о главе домохозяйства приводится в приложении IV.

2. Редактирование данных о супруге

a) Когда присутствует точно один супруг в моногамном обществе

305. Если есть точно один супруг, то переменная «супруг–указатель» («spouse–pointer») отслеживает номер строки супруга для дальнейшего редактирования. Программа редактирования может включать поиск противоположного пола главы домохозяйства и супруги(а), соответствующих возрастным различиям или других относящихся к делу характеристик. (В странах, где разрешены однополые браки, необходимо адаптировать программу редактирования.)

b) Когда обнаруживается более одного супруга в моногамном обществе

306. В моногамном обществе, если в наборе данных обнаруживается более одного супруга, то программа редактирования должна определить, кто является супругом(ой), и изменить соответственно степень родства других лиц, которые были обозначены как супруги. Кроме того, специалисты по содержа-

тельной части переписи должны определить, какими должны быть характеристики и по какой схеме пойдет редактирование.

c) Супруги в полигамном обществе

307. Если в полигамном обществе обнаруживается более одного супруга, группа редактирования, возможно, захочет оставить информацию в первоначальном виде или провести какую-либо проверку на согласованность. Например, как минимум для каждого из супругов в полигамном браке глава домохозяйства должен быть противоположного пола. А если супруги оказываются одного пола, то необходимо отредактировать ранее встречавшиеся данные о супругах одного и того же пола.

d) Прочие характеристики глав домохозяйств и их супругов

308. Надлежащая практика редактирования предполагает вменение других важных элементов данных о главе домохозяйства и супруге(у), когда они выявлены на этом этапе общего процесса редактирования. Эти элементы включают возраст главы домохозяйства и супруги(а) и семейное положение, которые могут понадобиться позднее в файлах вменения и для других целей редактирования. Кроме того, хорошей идеей является получение с самого начала «социальных» элементов данных, таких как вероисповедание, этническая принадлежность и язык главы домохозяйства, особенно в том случае, если глава не вошел в список как первое лицо. Поскольку большинство программных пакетов начинаются с первого лица и работают по нисходящей схеме, важно располагать информацией о главе домохозяйства до начала редактирования данных о других лицах в данной жилищной единице. Пример блок-схемы редактирования данных, касающихся супруга(и), приводится в приложении IV.

H. Возраст и дата рождения

1. Когда имеется дата рождения, но не указан возраст

309. Когда получена дата рождения, а возраст не указан, то эту информацию можно получить путем вычитания даты рождения из даты проведения переписи или обследования. Некоторые национальные бюро переписи/статистические управления предпочитают получать данные о возрасте, основываясь только на годе проведения переписи и годе рождения, что дает значение с вероятным отклонением. Если используются год и месяц, то возраст будет вычислен более точно, но самые точные результаты дадут день, месяц и год.

2. Когда возраст и дата рождения не согласуются между собой

310. Когда при переписи или обследовании собираются данные и о возрасте, и о дате рождения, «вычисленный» возраст получается путем вычитания даты рождения из даты переписи. Если это значение отличается более чем на один год от зарегистрированного возраста, то группа редактирования, возможно, пожелает произвести корректировку. Как правило, дата рождения имеет преимущество над зарегистрированным возрастом, и зарегистрированный возраст заменяется вычисленным возрастом.

I. Учет недостоверных данных

311. Некоторые группы редактирования, возможно, предпочтут выполнить процедуры подсчета количества недостоверных и несогласующихся записей для основных переменных (или всех переменных), таких как возраст и пол, прежде чем приступить к фактическому редактированию. Если группа редактирования готовится заранее или проводит периодические обследования с использованием этих самых элементов данных, то в их распоряжении могут быть несколько разных матриц динамического вменения. Если доля недостоверных или несогласующихся данных очень мала, группа редактирования может принять решение использовать для вменения только несколько переменных. Если доля ошибок больше, группе редактирования может понадобиться больше переменных, для того чтобы провести большое количество необходимых процедур вменения.

312. Как правило, лучше применять небольшие матрицы вменения, поскольку они легче проверяются при разработке программ редактирования и вменения и их легче использовать во время фактического редактирования. Однако в случае неоднократного использования значений понадобится более крупная и более дифференцированная матрица вменения.

Глава IV

Редактирование элементов данных о населении

313. Глава IV посвящена редактированию элементов данных о населении, включая демографические данные, данные о миграции и социально-экономические характеристики. В спецификациях для этих элементов редактирования учитывается достоверность отдельных элементов данных и согласованность между элементами данных о населении, а также согласованность между элементами данных о населении и жилым фонде. Знание некоторых взаимосвязей между этими элементами позволяет планировать процесс редактирования с точки зрения согласованности, с тем чтобы обеспечить более высокое качество данных для последующего составления таблиц. Например, в демографических записях не должно быть 15-летних женщин с десятью детьми или семилетних детей, посещающих высшую школу.

314. При указании значений для элементов данных о населении группа редактирования должна решить, ставить ли запись «не указывается»; производить ли статическое вменение («cold deck») для значения «неизвестно» или другого значения; или производить динамическое вменение («hot deck») на основе характеристик других лиц или жилищных единиц.

315. Во многих случаях предпочтительнее динамическое вменение, поскольку это исключает редактирование на этапе составления таблиц, когда информация есть только в самих таблицах, для принятия решений о неизвестных величинах. Матрицы вменения позволяют вводить записи для пропусков, недостоверных сведений или устранения несоответствий, когда нет других связанных с ними элементов данных с достоверными ответами. В некоторых странах в общенациональном масштабе наблюдается многообразие характеристик о населении, хотя в большинстве отдельных населенных пунктов такого разброса нет. В других странах среди населенных пунктов могут быть значительные отклонения, особенно если речь идет о проживании в городских и сельских районах. Такие изменения необходимо учитывать при разработке матриц вменения и, в частности, при установлении исходных величин «cold deck». Группа редактирования должна уточнять обстоятельства, при которых следует вычислять данные для того или иного пропуска. Их следует брать из предыдущей жилищной единицы с аналогичными характеристиками.

316. В целях содействия обработке данных все записи о населении должны иметь серийные номера. При редактировании структуры данных, о котором говорилось в главе III, производится проверка на соответствие между порядковым номером и последовательностью серийных номеров.

317. Группа редактирования должна редактировать каждую запись данных о населении только в отношении соответствующих элементов. Редактируе-

мые элементы должны отличаться один от другого в зависимости от городских/сельских, климатических и иных условий. Редактирование желательно производить выборочно в зависимости от этих условий, но практически лишь немногие страны располагают временем и опытом для разработки и применения многомерных массивов данных для замены недостающих или несогласующихся данных. Еще меньше стран применяют эту дополнительную процедуру на практике.

318. Информация, собранная с помощью переписного листа, также применима только к отдельным группам населения. Например, вопрос о рождаемости задается только женщинам, а об экономической активности — только взрослым.

319. Иногда группа редактирования в отношении определенных элементов данных должна допускать запись «не сообщается». В группе редактирования может не быть хорошей основы для вменения ответов по некоторым характеристикам. Решение оставить запись «не сообщается» должно увязываться с требованием выдавать соответствующие характеристики в виде таблиц в целях планирования и формирования политики. До тех пор, пока случаи записей «не сообщается» распределяются так же, как и регистрируемые случаи, распределение записей «не сообщается», когда планирующим органам необходима выборочная информация, не должно вызывать затруднений. Однако, если случаи ответа «не сообщается» в какой-то степени асимметричны, процедура вменения после компиляции может быть проблематичной, особенно в отношении небольших районов или конкретных типов условий. Например, если респонденты из числа молодых девушек отказываются раскрывать личную информацию о рождаемости и данные о рождаемости не указываются, то редактирование не поможет получить эту информацию.

320. Редактировать данные о населении, как правило, сложнее, чем данные о жилом фонде, поскольку сводные таблицы, как правило, намного сложнее. Большинство стран составляют отдельные характеристики жилого фонда только по разным территориальным уровням, но при этом могут иметь многоуровневые сводные таблицы с элементами данных о населении. Как разъяснялось выше, страны, предпочитающие не использовать динамическое вменение, должны определить идентификатор для значений «неизвестно», чтобы пользоваться им в случае недостоверных или несогласующихся ответов.

321. Если страны используют динамическое вменение, группы редактирования должны разработать простые матрицы вменения с размерностями, которые дифференцируют характеристики по населению. Для большинства стран наилучшими первичными переменными для динамического вменения являются возрастная группа и пол, и их следует редактировать в первую очередь. Национальные бюро переписи/статистические управления, использующие метод редактирования переменных по групповому принципу, должны одновременно редактировать данные о возрасте, поле и другие переменные, такие как степень родства и семейное положение. К другим элементам данным, которые могут оказаться полезными при динамическом вменении, относятся уровень образования и статус занятости.

322. Группы редактирования должны работать очень тщательно, чтобы не исказить данные в ходе вменения. Не следует исходить из того, что невмененные и вмененные данные будут обязательно иметь одинаковое распределение. Зачастую неизвестные данные искажаются сами по себе. Например, люди старшего возраста менее расположены сообщать свой возраст, чем молодежь.

А. Демографические характеристики

323. Данные, касающиеся степени родства, пола, возраста и семейного положения каждого человека, являются основными для любой переписи и, вероятно, их следует редактировать вместе. Структуры возраста и пола групп и подгрупп населения имеют основополагающее значение практически при любом планировании, которое ведется на основе данных переписи населения. Эти элементы данных также необходимы для составления содержательных таблиц, поскольку фактически все другие виды анализа основаны на сводных таблицах с другими переменными по возрасту и полу.

324. Принцип использования переменных по групповому принципу (Фелледжи–Холта) при редактировании данных о населении и жилом фонде был рассмотрен в главе II настоящего *Руководства*. Поскольку демографические переменные являются неотъемлемой частью планирования всех переписей, то этот подход следует использовать, если позволяют время и опыт. Качество всего набора данных почти наверняка выиграет от того, что для определения ошибок или несоответствий будет применяться в приоритетном порядке редактирование по возрасту, полу и другим выбранным переменным. В первую очередь подвергаются редактированию элементы данных, содержащие максимальное количество ошибок, а затем элементы, содержащие меньше ошибок или несоответствий.

1. Степень родства

325. Элемент данных о степени родства используется для того, чтобы легче было определить структуру домохозяйства и семьи. Этот элемент указывается где-то в самом начале большинства вопросников переписи и обследования и помогает удостовериться в том, что зарегистрирован каждый, кто проживает в данной жилищной единице. Счетчик и респондент используют информацию о родственных отношениях между членами домохозяйства, чтобы убедиться в том, что никто не пропущен. Элемент данных о степени родства помогает также произвести проверку на согласованность данных, касающихся пола и возраста членов домохозяйства. Об определении единственного главы домохозяйства и не более одного супруга(и) (в неполнородном обществе) говорится в разделах о редактировании структуры данных.

а) Редактирование данных о степени родства

326. Поскольку статистические данные о степени родства приобретают все более важное значение, необходимо более тщательно подходить к разработке приемов редактирования, позволяющих составлять различные виды таблиц с указанием состава семьи и семьи второго поколения. В этом, несомненно, поможет разработка в первую очередь соответствующих кодов родства (см. [приложение I](#) в отношении понятия «тип семьи», и по вопросам перекодирования числа членов семьи второго поколения и степени родства в семье второго поколения).

327. Когда данные о степени родства указать нельзя, а динамическое вменение не применяется, то недостоверные или несогласующиеся ответы необходимо относить к категории «неизвестно». При динамическом вменении родство можно вычислить из матрицы вменения по возрасту и полу или другим соответствующим характеристикам. Матрицы вменения не должны исчислять

такие данные о родстве, которые противоречили бы уже установленным отношениям в домохозяйстве. Например, не следует производить вменение данных о вторых и третьих супругах даже в полигамных домохозяйствах, если только группа редактирования не решит использовать такой прием.

b) Когда главу домохозяйства необходимо указывать первым

328. Если глава домохозяйства не указывается как первое лицо, то при редактировании структуры данных, о чем говорилось в главе III, выяснится, что, для того чтобы отследить положение главы, можно применить указатель. Если группа редактирования захочет указать главу как первое лицо, то его можно переместить на первую позицию или изменив порядок лиц, или оставив домохозяйство на месте, но изменив степень родства, как отмечается в главе о редактировании структуры. Первый метод требует значительного опыта программирования; второй метод может нарушить целостность набора данных, если не принять чрезвычайных мер.

c) Когда степень родства кодируется в обратном порядке

329. Иногда счетчики собирают данные о родстве в обратном порядке: вместо того чтобы собирать данные о родстве каждого члена домохозяйства по отношению к его главе, они собирают эти данные о главе по отношению к каждому члену домохозяйства. Отсюда степень родства третьего лица указывается не как «ребенок», а как «родитель». И тогда домохозяйство может оказаться с четырьмя родителями вместо четырех детей. Когда группа редактирования обнаружит, что такая проблема встречается систематически, она должна выработать решение, чтобы не наносить слишком большого ущерба данным о домохозяйстве.

330. Такая процедура инверсии родственных отношений обычно предусматривает ведение «справочного» файла, содержащего данные о прямой степени родства и инверсной степени родства, принимая во внимание пол респондентов.

d) Когда имеются супруги в полигамном браке

331. Если редактирование структуры будет проводиться так, как указано в главе III, то оно обеспечит проверку моногамных домохозяйств в отношении того, чтобы там был «один-единственный» глава домохозяйства и «не более одной(ого) супруги(а)». Что же касается полигамных домохозяйств, то здесь группа редактирования должна решить, когда полигамные отношения разрешены, а когда нет. Иногда домохозяйство, казалось бы, имеет полигамные отношения, а на самом деле это не так.

332. Например, в домохозяйстве могут быть указаны глава и супруг(а), а другая пара зарегистрирована как «супруги» по отношению друг к другу и в общей сложности имеются три супруги(а). При редактировании важно убедиться, что вторая пара — это действительно не отец и мать, не сын и невестка, не сестра и зять или какая-либо иная комбинация. Иногда эти отношения можно определить с некоторой долей определенности, а иногда они не поддаются определению. Когда вышеуказанным отношениям присваивается определенный код, группа редактирования должна исходить из необходимости произвести соответствующее вменение. Когда в полигамных домохозяйствах дополнитель-

ные супруги являются фактическими супругами, то при редактировании следует произвести проверку по полу и, возможно, по возрасту.

е) Наличие нескольких родителей

333. В данных о домохозяйстве не должно быть указано более двух «родителей», причем родители должны быть противоположного пола. Если появляется более двух родителей, то, вероятно, дополнительных родителей следует отнести к категории «прочие родственники». Иногда при переписях или обследованиях присваивается код для «родителя» или «кого-либо из родителей» мужа или жены, что позволяет получить до четырех «родителей» вместо двух, при этом не более двух родителей каждого пола.

ф) Когда при переписи собираются данные о родстве по признаку пола

334. Иногда при проведении переписей или обследований собираются данные о степени родства по признаку пола: «муж» и «жена» отдельно вместо категории «супруг(а)»; «сын» и «дочь» вместо категории «ребенок» и т. д. Если эти ответы не отредактировать, то в таблицах могут появиться данные, где дочери будут отнесены к мужскому полу, а мужа — к женскому. В этом случае группа редактирования должна установить приоритетность редактирования, то есть отдать предпочтение или родству или полу. В некоторых случаях, например в отношении мужа и жены, редактирование более важно, чем в других случаях, например для категории «маленький ребенок». Следует отметить, что не рекомендуется использовать родственные отношения по признаку пола, поскольку избыток данных, скорее, не проясняет, а, наоборот, маскирует характер родственных отношений и требует дополнительного редактирования.

г) Когда данные о родстве и семейном положении не согласуются между собой

335. Данные о степени родства и семейном положении должны согласовываться между собой, если происходит их наложение: лица, которые в графе «родство» указывают «супруг(а)», в графе «семейное положение» должны указывать «в браке». Группа редактирования выбирает, какую переменную изменить, если эти элементы данных не согласуются между собой. Иногда ясности в отношении родства нет, и поэтому нужно тщательно подойти к разработке спецификаций редактирования. Например, во многих странах зять может быть и братом супруга(и) (и поэтому не должен быть женат), а также супругом сестры (и должен быть женат).

336. В настоящее время появляются и другие, более современные проблемы в связи с сообщением данных о родстве. Когда два лица противоположного пола живут вместе, не состоя в брачных отношениях, то код отношений родства может быть или «не состоящий в браке партнер», или «супруг(а)». Если при переписи или обследовании указывается код для не состоящего в браке партнера, то тогда соответствующее семейное положение не должно указываться как «в браке», если только данное лицо не состоит в браке с кем-либо еще помимо того, с кем оно живет.

337. Аналогичная ситуация возникает, когда лица одного и того же пола проживают вместе, находясь или в любовных или не в любовных отношениях.

Лица, не находящиеся в любовных отношениях, могут быть закодированы как «сожители» или «не родственники». В отношении тех, кто состоит в любовных отношениях, для некоторых стран может быть приемлема категория «не состоящий в браке партнер». В этом случае группа редактирования также должна решить, как указывать соответствующее семейное положение. В рамках переписи невозможно провести различие между платоническими и любовными отношениями.

2. Пол

338. Пол является одной из самых простых характеристик с точки зрения сбора данных, но требует некоторого осмысления при редактировании. Это — одна из важнейших переменных, поскольку большинство характеристик, касающихся населения, анализируются на основе пола. Вменение данных о поле требует некоторого сопоставления с другими переменными. В некоторых случаях пол того или иного лица должен определяться на основе разницы между полами связанных с ним лиц, как правило, главы домохозяйства и супруги(а), а также между родителями и родни со стороны мужа или жены. Вероятно, не следует отмечать пол как «недостовверный» или «неизвестен», поскольку он имеет столь важное значение в качестве переменной. В связи с этим следует подумать, как лучше всего получить результаты, сопоставимые с реальной ситуацией в стране. В приложении IV содержится номер блок-схемы по редактированию данных, касающихся пола главы домохозяйства и супруга(и).

339. Если то или иное лицо не является главой домохозяйства или супругой(ом) главы, то других лиц, на которых можно ссылаться, нет; поэтому следует проверять другие элементы в записи данного лица. И если есть данные о рождаемости, то женщине следует присвоить код. Однако если, например, сведения о поле данного лица отсутствуют или недостоверны, но есть супруг или супруга, пол которого или которой указан, то при редактировании этому лицу можно вменить противоположный пол.

a) Когда код пола достоверен, но глава домохозяйства и супруг(а) одного пола

340. В тех случаях, когда встречаются серьезные противоречия, код пола необходимо менять, даже если существует достоверный код. Например, запись показывает, что есть вторая состоящая в браке пара и при этом домохозяйство уже имеет главу и супругу(а) или семейную пару в семье второго поколения. Если оба лица во второй паре регистрируют один и тот же пол, то можно использовать информацию о рождаемости и другие элементы данных, для того чтобы определить кто мужчина, а кто женщина. После этого можно исправить ошибочную запись о том или ином лице.

b) Когда в отношении мужчины имеется информация о рождаемости, а в отношении взрослой женщины нет

341. При редактировании можно обнаружить, что в отношении мужчины имеется информация о рождаемости и/или о детях, находящихся дома, то есть ошибка могла возникнуть из-за личного номера матери или другой

аналогичной переменной. Если супруг отсутствует, то пол можно поменять на женский, вместо того чтобы удалять информацию о рождаемости. Аналогичным образом, в отношении взрослой женщины, если отсутствует информация о рождаемости и не указаны дети, можно поменять данные на категорию «мужской» при определенных обстоятельствах, установленных группой редактирования.

с) Когда код пола является недостоверным и имеется супруг(а)

342. Если графа «пол» не заполнена или содержит недостоверные данные, то программа редактирования должна использовать записи о степени родства с главой домохозяйства и пол супруги(а), если он достоверен, — для того чтобы определить правильный код. Если отношения с главой домохозяйства обозначены как «глава домохозяйства», то программа проверяет, имеется ли супруг(а) (путем проверки данных о другом члене домохозяйства, в отношении которого указано «супруг(а)»). При определении кода пола супруга(и) код противоположного пола присваивается главе домохозяйства.

d) Когда код пола для супруга(и) недостоверен

343. Если степень родства с главой домохозяйства определена как «супруг(а)» и указывается пол главы домохозяйства, программа присваивает этому лицу пол, противоположный по отношению к главе домохозяйства.

e) Когда код пола недостоверен и имеется информация о женщине

344. По многочисленным признакам в переписном листе можно узнать, является ли респондент женщиной. Если программа еще не определила пол лица и имеются какие-либо показатели, касающиеся женщины, то в записи для этого лица следует указать женский пол. Например, если в редактируемых данных о конкретном лице указано достаточное число элементов, касающихся рождаемости, то в этом случае пол можно указать как женский. К данным о рождаемости относятся такие, как все дети, родившиеся живыми, дети, живущие в этом домохозяйстве, дети, живущие в других местах, умершие дети и дети, родившиеся живыми за последние 12 месяцев. Еще может быть так, что это лицо может быть матерью кого-то другого в данном домохозяйстве, причем номер строки для этого лица будет равняться номеру строки матери другого лица в данном домохозяйстве.

f) Когда код пола недостоверен и это лицо является мужем супруги

345. Если указанное лицо является мужем кого-либо еще в домохозяйстве исходя из элемента данных, показывающего номер строки мужа, то в графе «пол» следует указать «мужской».

g) Когда код пола недостоверен, а информации для определения пола недостаточно

346. Если группа редактирования совсем не использует метод динамического вменения, то значение «пол неизвестен» необходимо присваивать. К сожалению, это означает, что во всех таблицах должна быть дополнительная

колонка или дополнительный ряд или группа колонок или рядов для лиц, чей пол неизвестен. Поскольку пол — двоичная переменная, то величины можно присваивать попеременно, начиная с одной из двух и используя при этом противоположный пол для второй недоуверной записи, и затем продолжать в том же порядке.

h) Примечание об вмененных соотношениях полов

347. Если применяется вменение по методу «cold deck», то женский пол присваивается, как правило, чаще. Только у взрослых женщин есть данные о рождаемости, и их выбор искажается в какой-то степени случайно. В связи с этим, если имеющейся информации недостаточно, то лицо, о котором нет такой информации, чаще всего мужчина, а не женщина. Следовательно, необходимо предусматривать разработку матриц вменения, где учитывались бы общие соотношения между полами.

3. Дата рождения и возраст

348. Возраст является одной из наиболее трудных характеристик с точки зрения сбора и редактирования. Однако он, возможно, является важнейшей переменной, поскольку практически все связанные с населением характеристики анализируются исходя из возраста. Редактирование данных о возрасте требует подробного сопоставления с другими переменными величинами и другими лицами в доме. В большинстве случаев в основе вменения возраста должны лежать хранящиеся в памяти различия в возрасте родственников. Если исходя из этого возраст исчислить невозможно, то тогда следует использовать другие характеристики в записи об этом лице. Редактирование, возможно, потребует нескольких матриц вменения, включая возраст в разбивке по полу, семейному положению, степени родства и посещению школы; возрастные различия между матерью и ребенком; возрастные различия между мужем и женой; и возрастные различия между главой домохозяйства и супругой(ом).

a) Возраст и дата рождения

349. При структурном редактировании возраст вычисляется по дате рождения. При этом, однако, следовало бы рассмотреть разницу между возрастом и датой рождения. Как указывается во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда* (United Nations, 2008, para. 2.136), информацию о возрасте можно получить либо исходя из даты (года, месяца и дня) рождения, либо задав прямой вопрос о возрасте на момент последнего дня рождения.

350. Дата рождения дает более точную информацию и по возможности должна использоваться при всех обстоятельствах. Если точный день или даже месяц рождения неизвестны, то их можно заменить указанием на время года. Вопрос о дате рождения является уместным, когда людям известна дата их рождения, которая может быть установлена в соответствии с солнечным или лунным календарем, или выражена в годах, пронумерованных или определенных в традиционной народной культуре по названиям в рамках регулярного цикла.

351. Однако применительно к календарной системе, на которой основывается дата рождения, чрезвычайно важно, чтобы между счетчиком и рес-

пондентом было четкое взаимопонимание. Если возможны ответы некоторых респондентов со ссылкой на календарную систему, отличную от той, которую использовали другие респонденты, то в переписном листе необходимо предусмотреть отметку о применяемой календарной системе. Счетчикам не рекомендуется пытаться переводить даты из одной системы в другую. Необходимый перевод можно легче всего осуществить как часть компьютерного редактирования данных (United Nations, 2008, para. 2.137).

352. По ряду причин на прямой вопрос о возрасте, возможно, будет дан неточный ответ. Даже если все ответы основываются на одном и том же методе определения возраста, респондент может не понять, о каком возрасте требуются сведения: на момент последнего дня рождения, следующего дня рождения или ближайшего дня рождения. Кроме того, могут возникать и другие проблемы: возраст может быть округлен до ближайшей цифры, оканчивающейся на 0 или 5; может указываться приблизительный возраст без должного указания на это или умышленно даваться неправильный ответ (United Nations, 2008, para. 2.138).

Вставка 4

Редактирование и вменение возраста

При редактировании и вменении возраста нужно сделать следующее:

- ✎ приписать возраст, если вместо возраста стоит пробел;
- ✎ проверить минимальный возраст лиц, когда-либо состоявших в браке;
- ✎ проверить минимальный возраст главы домохозяйства;
- ✎ проверить минимальный возраст родителей; и
- ✎ произвести любые другие проверки, характерные для каждой страны.

353. Многие национальные бюро переписи/статистические управления регистрируют или дату рождения, или возраст, а не и то и другое вместе. Как указывается во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда* (United Nations, 2008), возраст в полных годах имеет очень важное значение: он используется при редактировании многих данных, а также в качестве размерности для многих матриц вменения. Не менее важно и то, что многие страны строят свою политику, исходя из данных о возрасте, поэтому необходимо прилагать все усилия, чтобы получать данные о возрасте самого высокого качества. Однако даже в идеальных ситуациях некоторые данные о возрасте указываться не будут. Следовательно, необходимо обеспечить, чтобы возраст рассчитывался должным образом и согласовывался с другими ответами отдельных членов домохозяйства.

b) Взаимосвязь между датой рождения и возрастом

354. Если данные о возрасте не собирались отдельно от даты рождения, то возраст следует вычислить во время редактирования структуры. Редактирование данных о возрасте наравне с другими данными обеспечит всестороннюю проверку на согласованность данных в записях и между записями. Но сначала нужно вычислить возраст исходя из даты рождения и даты проведения переписи. Необходимо проверить возраст после его вычисления по дате рождения, чтобы убедиться, что он вписывается в рамки даты проведения переписи.

355. Возраст детей, родившихся в год проведения переписи, но после даты ее проведения, будет вычислен как минус 1, и его необходимо исправить. Детей, зарегистрированных после даты проведения переписи, вероятно, следует из нее исключать. Однако, если после проверки обнаруживается, что дата рождения указана неправильно при регистрации счетчиками или при обработке, нужно использовать другие переменные для получения более точной оценки возраста.

с) Когда рассчитанный возраст выходит за верхний предел

356. Для переписей 2000 года и последующих большинство стран предпочитают записывать год рождения четырьмя цифрами. Для переписей, проводимых около 2010 года, приемлемым диапазоном будет примерно период с 1900-х или 2000-х годов и до года переписи. Хотя для работы компьютера достаточно и трех цифр, использование этих трех цифр для указания года может ввести в заблуждение и счетчиков, и работников статистических управлений. Иногда рассчитанный год будет выходить за верхнюю границу определенных переписью лет, и тогда потребуются вносить коррективы. Если перепись проводится в 2010 году, а человек сообщает, что родился в 1860 году, то расчетный возраст 150 лет очевидно выйдет за приемлемые пределы и его необходимо будет изменить.

d) Редактирование данных о возрасте

357. С помощью программы редактирования следует проверить согласованность указанного возраста данного лица с указанным возрастом его матери, отца или ребенка. Редактирование должно обеспечить минимальную разницу в годах между возрастом матери или отца и возрастом ребенка. В случае вменения возраста проверки на согласованность должны производиться по таким записям, как годы, прожитые в данном районе (длительность проживания) и самый старший законченный класс школы (уровень образования). Все эти проверки следует производить до изменения возраста или до указания вмененного возраста.

358. Редактирование следует начинать с проверки на достоверность. Если возраст указан правильно, специалисты, возможно, пожелают проверить, согласуется ли возраст данного лица с возрастом его/ее матери (если в домохозяйстве указана мать данного лица) и детей данного лица (если данное лицо является женщиной и имеет детей в домохозяйстве). Если данные о возрасте не согласуются между собой, то следует сделать заметку о возрасте этого лица и изменить его позднее.

e) Редактирование данных о возрасте при наличии данных о возрасте главы домохозяйства и супруги(а)

359. На следующем этапе редактирования необходимо определить наличие супруги(а). Если это так, то ее(его) возраст следует проверить на достоверность (по меньшей мере, X лет, в зависимости от установленного в стране минимального возраста вступления в брак). Если возраст не согласуется и если применяется динамическое вменение, то программа будет использовать специальное значение, полученное путем вменения, исходя из разницы в возрасте между мужем и женой. Диапазон возрастных различий меньше, чем сам возраст, поэтому в матрице вменения, используемой в программе, сохранится разница

в возрасте (из предыдущих записей) мужа и жены. Это значение прибавляется или вычитается из возраста супруга(и), чтобы получить рассчитанный возраст.

360. Для того чтобы этот рассчитанный возраст согласовывался с другими характеристиками, в матрицу вменения следует также включить семейное положение, длительность проживания и самый старший законченный класс школы. Исключение этих переменных может привести к тому, что вмененный возраст будет меньше количества лет, которые данное лицо прожило в конкретном месте, или меньше, чем это предполагает уровень обучения. Например, матрица вменения может показать возраст 8 лет, но данное лицо может указать, что они прожили в этом месте 10 лет. Без других переменных, когда программа редактирования производит редактирование данных о годах проживания, другая матрица вменения изменит годы проживания с правильного значения на неправильное.

f) Редактирование данных о возрасте главы домохозяйства при отсутствии данных о возрасте супруги(а) главы, но при наличии данных о возрасте ребенка

361. Когда нет возможности сопоставления с возрастом супруги(а) при определении возраста главы домохозяйства, то программа может проверить их взаимоотношения. Если эти отношения обозначены как «глава домохозяйства», программа редактирования может проверить другие записи о домашнем хозяйстве (если таковые имеются) в отношении сына или дочери, возраст которых, как известно, указан правильно. Программа проверяет возраст сына или дочери и рассчитывает возраст главы домохозяйства с использованием динамического вменения по «возрастной разнице» аналогично тому, как это делается в отношении мужа и жены. Как и ранее, при вычислении возраста принимаются во внимание длительность проживания и максимальный уровень образования. Итоговый возраст, таким образом, будет согласовываться с этими переменными и не будет очевидных ошибок, возникающих при включении числа лет, прожитых в данном районе, и самого старшего законченного класса школы в матрицу вменения.

g) Редактирование данных о возрасте главы домохозяйства при наличии данных о возрасте родителя главы домохозяйства

362. Если какое-либо лицо не подпадает ни под одну из описанных выше категорий, программа может включать поиск родителя этого лица в домохозяйстве. При обнаружении родителя такого лица можно подсчитать возраст с помощью матрицы вменения, используя разницу в возрасте. Разница в возрасте между ребенком и родителем обычно колеблется в более широких пределах, чем разница в возрасте между мужем и женой. По этой причине такое редактирование в программе применяется только тогда, когда методика, основанная на разнице в возрасте между мужем и женой, не дает результатов. В вычисляемых возрастных показателях следует учитывать показатели образования, самый старший законченный класс школы, длительность проживания в данном районе, семейное положение, рождаемость и степень экономической активности. В программе следует сделать предположение о том, что то или иное лицо находится по меньшей мере в минимально допустимом возрасте, если он/она когда-либо состоял(а) в браке, имеет детей или сообщает о любой степени экономической активности.

h) Редактирование данных о возрасте главы домохозяйства при наличии данных о возрасте внука/внучки

363. Когда то или иное лицо не подпадает ни под одну из вышеупомянутых категорий, программа может включать поиск внука/внучки данного лица в домохозяйстве. При обнаружении внука/внучки данного лица можно рассчитать возраст с помощью матрицы вменения, используя разницу в возрасте. Разница в возрасте между главой домохозяйства и его внуком/внучкой колеблется в значительно более широких пределах, чем разница в возрасте между мужем и женой или разница в возрасте между главой домохозяйства и его ребенком. По этой причине такое редактирование применяется в программе только тогда, когда методика, основанная на разнице в возрасте между мужем и женой и разнице в возрасте между главой домохозяйства и его ребенком, не дает результатов. В рассчитываемом возрасте следует учитывать показатели образования, включая самый старший законченный класс школы, продолжительность проживания в данном районе, семейное положение, рождаемость и экономическую активность. В программе следует сделать предположение о том, что то или иное лицо находится по меньшей мере в минимально допустимом возрасте, если он/она когда-либо состоял(а) в браке, имеет детей или участвует в любой экономической деятельности.

i) Редактирование данных о возрасте главы домохозяйства при отсутствии данных о возрасте других лиц

364. Если какое-либо лицо не подпадает ни под одну из вышеупомянутых категорий, программа может включать поиск другого родственника или лица, не состоящего в родственных связях с главой домохозяйства. При наличии такого лица и если такое лицо имеет зарегистрированный возраст, группа редактирования должна решить, стоит ли ей включать любую имеющуюся информацию в матрицу вменения, используя различия в возрасте. Однако эти различия в возрасте между главой домохозяйства и другими родственниками или лицами, не состоящими с ним в родственной связи, колеблются в таких широких пределах, что группа редактирования может вообще отказаться от этих усилий и просто использовать другие переменные для динамического вменения возраста главы домохозяйства. В любом случае в программе такое редактирование применяется только после того, как методика, основанная на разнице в возрасте между мужем и женой, главой домохозяйства и ребенком, главой домохозяйства и родителем, а также главой домохозяйства и внуком/внучкой, не дает каких-либо результатов. Однако если вычисляемый возраст определен, в нем должны учитываться характеристики образования, самый старший законченный класс школы, а также длительность проживания в округе, семейное положение, рождаемость и экономическая активность. В программе следует сделать предположение о том, что то или иное лицо находится по меньшей мере в минимально допустимом возрасте, если он/она состоял(а) когда-либо в браке, имеет детей или участвует в любой экономической деятельности.

j) Редактирование данных о возрасте супруги(а), когда уже определен возраст главы домохозяйства

365. Редактирование возрастных показателей супруги(а) обычно проводится одновременно с редактированием возрастных показателей главы домохо-

зайства, поскольку для совместного редактирования требуется информация от обоих лиц. Если, однако, проводится раздельное редактирование, когда возраст супруги(а) указан неверно или не согласуется с другими переменными, для определения наиболее приближенной оценки возраста супруги(а) следует применять матрицу динамического вменения, используя разницу в возрасте с главой домохозяйства и другие переменные. Как и ранее, при определении возраста следует учитывать характеристики образования, включая самый старший законченный класс школы, длительность проживания в данном районе, семейное положение, рождаемость и экономическую активность. В программе должно делаться предположение, что то или иное лицо находится по меньшей мере в минимально допустимом возрасте, если он/она когда-либо состоял(а) в браке, имеет детей или участвует в любой экономической деятельности.

к) Редактирование данных о возрасте других супружеских пар в домохозяйстве, когда известен возраст одного из супругов

366. При редактировании следует прежде всего установить, принадлежит ли данная запись лицу, состоявшему в браке. При положительном результате программа может включать поиск супруги(а) данного лица среди других записей о домохозяйстве. Если поиск супруга(и) оказывается безрезультатным, программа приступает к следующей части редактирования. Если с помощью поиска обнаруживается один из супругов, возраст супруга(и) следует проверить на достоверность (минимальный возраст X лет, в зависимости от установленного в стране минимального возраста вступления в брак). Если возраст не согласуется с другими данными и если применяется динамическое вменение, то тогда в программе используется специальное вмененное значение, полученное по разнице в возрасте между мужем и женой. Разница в возрасте колеблется не в столь широких пределах, как сами возрастные показатели, поэтому в матрицу вменения программы заносится (из предыдущих записей) разница в возрасте между мужем и женой. Это значение прибавляется или вычитается из возраста супруга(и) данного лица для определения вычисляемого возраста.

367. Для обеспечения согласования этого вычисляемого возраста с другими характеристиками, в матрицу вменения следует также включать семейное положение, длительность проживания и наивысший полученный уровень образования. Исключение этих переменных может привести к тому, что вычисляемый возраст будет меньше количества лет, которые данное лицо прожило в данном месте или меньше возрастного показателя, который соотносится с уровнем образования.

л) Редактирование данных о возрасте ребенка, когда уже определен возраст главы домохозяйства

368. Если речь идет о сыне или дочери главы домохозяйства, вычисляемый возраст может быть получен с помощью возрастного показателя главы домохозяйства, разницы в возрасте, длительности проживания и уровня полученного образования. И вновь при вычислении возраста следует учитывать характеристики образования, включая самый старший законченный класс школы, длительность проживания в данном районе, семейное положение, рождаемость и экономическую активность. В программе следует сделать предположение о том, что то или иное лицо находится по меньшей мере в минимально

допустимом возрасте, если он/она когда-либо состоял(а) в браке, имеет детей или участвует в любой экономической деятельности.

m) Редактирование данных о возрасте родителя, когда уже определен возраст главы домохозяйства

369. Если речь идет о родителе главы домохозяйства, вычисляемый возраст может быть получен с помощью возрастного показателя главы домохозяйства, разницы в возрасте, длительности проживания и уровня полученного образования. При вычислении возраста следует учитывать характеристики образования, включая самый старший законченный класс школы, длительность проживания в данном районе, семейное положение, рождаемость и экономическую активность. В программе следует сделать предположение о том, что то или иное лицо находится по меньшей мере в минимально допустимом возрасте, если он/она когда-либо состоял(а) в браке, имеет детей или участвует в любой экономической деятельности.

n) Редактирование данных о возрасте внука/внучки, когда уже определен возраст главы домохозяйства

370. Если речь идет о внуке/внучке главы домохозяйства, вычисляемый возраст можно получить с помощью возраста главы домохозяйства, разницы в возрасте, длительности проживания и уровня полученного образования. И вновь при вычислении возраста следует принимать во внимание характеристики образования, включая самый старший законченный класс школы, длительность проживания в данном районе, семейное положение, рождаемость и экономическую активность. В программе следует сделать предположение о том, что тому или иному лицу исполнилось по меньшей мере двенадцать лет, если он/она когда-либо состоял(а) в браке, имеет детей или участвует в любой экономической деятельности.

o) Редактирование данных по возрасту всех прочих лиц

371. Группе редактирования следует определить соответствующие матрицы вменения для других родственников и не родственников в домохозяйстве. Руководящие принципы будут зависеть от конкретной переписи или обследования и социально-экономических характеристик страны. Например, любое лицо, которое когда-либо состояло в браке, имело детей или участвовало в экономической деятельности наверняка должно быть по меньшей мере в возрасте, равном установленному той или иной страной минимальному возрасту. На основе этой информации, если применяется метод динамического вменения, значение, полученное по матрице вменения, должно быть не ниже минимального возраста. Аналогичным образом, если то или иное лицо посещает школу, имеет какую-либо школьную подготовку или умеет читать и писать, но не является главой домохозяйства, никогда не состояло в браке и не участвует в экономической деятельности, то тогда данное лицо следует отнести к группе, возраст которой меньше минимального возраста взрослых, но больше или равен минимальному школьному возрасту. После этого можно определить матричное вмененное значение для лиц младше минимального школьного возраста. При всем своем несовершенстве данный метод ограничивает разброс значений, которые могут приниматься в матрице вменения.

4. Семейное положение

372. Во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда* (пункты 2.144–2.151) семейное положение определяется как положение каждого человека в соответствии с законодательством о семье и браке или брачными обычаями данной страны. Категории семейного положения, подлежащие определению, включают, но не ограничиваются этим, следующие категории: *a)* одинокие лица, никогда не состоявшие в браке; *b)* лица, состоящие в браке; *c)* вдовы и вдовцы, не вступившие в новый брак; *d)* разведенные лица, не вступившие в новый брак; *e)* супруги, живущие раздельно. В некоторых странах категория *b* может потребовать выделения подкатегории лиц, которые по договору состоят в браке, но еще не живут в качестве мужа и жены. Во всех странах категория *e* должна включить супругов, живущих раздельно как по решению суда, так и де-факто, что можно отразить в виде отдельных подкатегорий, если это необходимо. Независимо от того, что живущие раздельно пары могут по-прежнему считаться супругами (поскольку они не имеют права вновь вступить в брак), ни одна из подкатегорий категории *e* не должна включаться в категорию *b*. В некоторых странах необходимо учитывать брачные союзы, основанные на обычном праве (которые, согласно обычному праву, признаются законными и имеющими юридическую силу), а также незаконные брачные союзы, часто называемые фактическим (гражданским) браком.

a) Редактирование данных о семейном положении

373. Группа редактирования должна принять решение относительно соответствующего минимального возраста вступления в первый брак для переписи или обследования. Минимальный возраст вступления в первый брак (примерно *X* лет) может отличаться в различных частях страны или в различных этнических группах. Если, например, сельское население вступает в брак раньше городских жителей, правила редактирования должны учитывать этот факт. Обычно национальное бюро переписи/статистическое бюро устанавливает этот возраст для самых ранних браков до регистрации, для того чтобы этот вопрос задавался только лицам старше установленного возраста. Лица более младшего возраста попадают в категорию лиц, «никогда не состоявших в браке», автоматически. Если, однако, всем задается вопрос о семейном положении, группа редактирования должна разработать соответствующую программу редактирования для всего населения.

b) Отнесение к категории «семейное положение» в случаях, когда не применяется метод динамического вменения

374. Хотя данные о семейном положении следует вносить в таблицы только для лиц в возрасте *X* лет и старше, где *X* равно минимальному возрасту вступления в первый брак, группы редактирования должны определить целесообразность и объем редактирования. Если в стране для недостоверных или несогласующихся ответов используются только графы «не указано» или «неизвестно», то при обнаружении недостоверных или несогласующихся первичных данных неправильные ответы должны заменяться кодом в графе «не указано». Если для лиц младше *X* лет ответ «никогда не состоял в браке» пропущен, то его следует приписать; поскольку статистические управления издадут наборы

данных для общего пользования, важно, чтобы в элементах данных, таких как семейное положение, всегда имелись соответствующие записи.

с) Отнесение к категории «семейное положение», когда применяется метод динамического вменения

375. При использовании метода динамического вменения при редактировании данных о семейном положении следует: *a)* произвести вменение значения, когда вводимые данные находятся за пределами интервала; и *b)* проводить проверку на соответствие между зарегистрированным семейным положением, степенью родства и возрастом.

d) Супруг(а) должен(должна) состоять в браке

376. Всем лицам, которым присвоен код «супруг(а)» в категории степени родства, должен присваиваться код лиц, состоящих в браке.

e) Супруг(а) супружеской пары

377. Если номер строки супруга(и) лица А (лицо В) является переменной величиной, тогда по отношению к лицу В лицо А должно квалифицироваться в качестве супруга(и); кроме того, оба лица А и В должны состоять в браке и принадлежать к противоположному полу.

f) Если имеется супруг(а), то глава домохозяйства должен(должна) состоять в браке

378. Если по семейному положению нет записи, но есть данные в графе «отношения к главе домохозяйства», согласно которым данное лицо определяется как глава, в программу следует включить проверочный поиск присутствия супруга(и) (путем проверки отношения к другим членам домашнего хозяйства). При наличии супруга(и) программа определяет семейное положение главы домохозяйства как «состоящего (состоящей) в браке».

g) Глава домохозяйства без супруги(а) и без детей

379. При отсутствии супруги(а) и если данное лицо относится к мужскому полу и имеет детей, программа производит вменение семейного положения по возрасту при наличии детей. Если нет детей, то программа может вменять семейное положение по возрасту и без наличия детей. Мужчина, который является главой домохозяйства, но жена которого не живет в домохозяйстве, с большой вероятностью относится к категории разведенных, живущих отдельно или вдовцов.

h) Если все остальные вычисления не дают результатов, следует производить вменение

380. Для лиц с кодами, выходящими за пределы диапазона, которым нельзя присвоить код на основе вышеупомянутых проверок, следующей проводится проверка возраста. Если возраст зарегистрирован достоверно, т. е. младше X лет, то при этом следует присваивать категорию «никогда не состоял в браке».

Во всех остальных случаях элемент данных следует классифицировать с помощью матрицы вменения. Необходимо составить матрицу вменения по полу и возрасту (двумерную), а также по полу, возрасту и родству (трехмерную), или по полу, возрасту, родству и числу детей, родившихся живыми (четырёхмерную). И вновь группам редактирования следует установить порядок редактирования, и, таким образом, при разработке матриц вменения важно помнить о том, какие элементы данных были отредактированы, а какие нет. Если до семейного положения были отредактированы только пол и степень родства, матрица вменения должна предусматривать графу «не сообщается» в других элементах данных.

l) **Взаимосвязь между возрастом и семейным положением молодых людей**

381. Для всех лиц, сообщивших достоверные данные о семейном положении, кроме категории лиц, «никогда не состоявших в браке», в отношении возраста следует провести проверку на согласованность. Все лица, когда-либо состоявшие в браке, должны быть в возрасте X или старше, где X равно минимальному возрасту вступления в брак для любого лица, установленному в конкретной стране. Если возраст составляет менее X лет или пропущен, следует провести новую проверку на согласованность по другим основным переменным (таким как число детей, родившихся живыми или экономическая активность). Если записи по этим элементам данных недостоверны, то в графе «семейное положение» следует указывать «никогда не состоял в браке»; во всех остальных случаях семейное положение не должно меняться.

5. Возраст вступления в первый брак

382. В соответствии со вторым пересмотренным изданием *Принципов и рекомендаций в отношении переписей населения и жилого фонда* (пункт 2.191) «датой первого брака» являются число, месяц и год заключения первого брака. Странам, в которых возникают трудности в установлении даты первого брака, рекомендуется собирать информацию о возрасте вступления в брак или о том, сколько лет назад был заключен брак (продолжительность брака). Важно включать не только первые браки, заключенные по договору, и фактические брачные союзы, но и браки, заключенные на основе обычая, а также церковные браки. В отношении женщин, которые на момент переписи овдовели, жили с мужем раздельно или состояли в разводе, следует получить информацию о «дате/возрасте/количестве лет, прошедших с момента расторжения первого брака». Информация о дате расторжения первого брака (если такое имело место) позволяет получить необходимые данные для исчисления «продолжительности первого брака» на этапе обработки данных в качестве производного показателя. В странах, где сведения о продолжительности брака имеют более высокую надежность, чем сведения о возрасте, составление таблиц данных о детях, родившихся живыми, по признаку продолжительности брака позволяет получить более точные оценки рождаемости, чем оценки на основе информации о детях, родившихся живыми, в их классификации по возрасту женщины. Данные о продолжительности брака можно получить путем вычитания возраста вступления в брак из возраста на момент переписи или непосредственно путем определения количества лет, прошедших с момента заключения брака.

383. Дату первого брака следует включать в перепись для всех лиц, когда-либо состоявших в браке. В программе должна проводиться проверка на соответствие: для лиц, никогда не состоявших в браке, информацию сообщать не следует, но лица, состоявшие в браке, должны иметь достоверные число, месяц и год. Группам редактирования необходимо принять решение о том, должны ли быть достоверными число и месяц: страны, не использующие метод динамического вменения, могут указать «неизвестно» для числа и месяца; страны, использующие динамическое вменение, могут произвести вменение числа и месяца, когда данные о них отсутствуют.

a) Графа возраста вступления в первый брак для лиц, никогда не состоявших в браке, не должна заполняться

384. Лица, которые никогда не состояли в браке, не должны сообщать возраст вступления в первый брак. Если для лиц, никогда не состоявших в браке, имеются достоверные данные, группа редактирования должна решить вопрос о том, изменить ли графу «семейное положение», или не заполнять графу «возраст» для этого лица. Если семейное положение должно быть изменено, страны, использующие только графу «не указано», будут применять этот код. Страны, использующие метод динамического вменения должны, вероятно, использовать возраст и пол для получения соответствующего ответа о семейном положении.

b) В графе для лиц, когда-либо состоявших в браке, должна быть соответствующая запись

385. Что касается года вступления в первый брак, то страны, не использующие метод динамического вменения, могут применять графы «не указано» или «неизвестно». Страны, применяющие метод динамического вменения, могут использовать другие переменные, например такие, как возраст супруга(и) или разницу в возрасте между супругами, число детей и детей, родившихся за последний год, для определения соответствующего года вступления в первый брак.

6. Рождаемость: дети, родившиеся живыми, и дети, живущие на момент переписи

386. Термин «дети, родившиеся живыми» означает общее число детей, родившихся живыми, из которого тем самым исключаются мертворожденные дети, выкидыши и аборт. Иногда демографы используют выражение «живорожденные дети», но в настоящем издании используются термины «дети, родившиеся живыми» или «родившиеся дети».

387. Совокупность, в отношении которой следует собирать данные по каждой из тем, содержащихся в этом разделе, включает женщин в возрасте 15 лет (или другом допустимом минимальном возрасте) и старше, независимо от того, состоят они в браке или нет, или женщин конкретных подкатегорий, таких как женщины, когда-либо состоявшие в браке. В странах, в которых не собираются и не сводятся в таблицы данные о женщинах в возрасте 50 лет и старше, все усилия следует сосредоточить на сборе данных только относительно женщин в возрасте от 15 до 50 лет, хотя при изучении последних данных о рождаемости в некоторых странах может оказаться целесообразным уменьшение нижнего возрастного предела на несколько лет (United Nations, 2007 para. 2.170).

а) Собираемые элементы данных о рождаемости

388. В издании *Принципы и рекомендации в отношении переписей населения и жилого фонда* (United Nations, 2008) Организация Объединенных Наций рекомендует получать информацию о трех элементах рождаемости: детях, родившихся живыми, дате рождения последнего ребенка, родившегося живым, и возрасте матери при родах первого, родившегося живым ребенка. Ответы на вопрос о возрасте, дате вступления в брак или продолжительности брака могут существенно улучшить оценки рождаемости, основанные на показателе родившихся живыми детей. Кроме того, во многих странах по-прежнему собирается информация о детях, живущих на момент переписи, что является подспорьем, особенно в ретроспективном анализе рождаемости.

389. В ходе переписей и обследований информация о рождаемости собирается от всех женщин с использованием установленного в каждой стране минимального возраста, а иногда также и максимального возраста.

б) Общие правила редактирования данных о рождаемости

390. Должны быть проверены лица женского пола младше установленного минимального возраста для предоставления данных о рождаемости и все лица мужского пола, и любая имеющаяся по ним информация о рождаемости должна быть удалена.

391. Цель редактирования данных о рождаемости состоит в том, чтобы добиться согласованности записей между собой и с данными о возрасте:

- а) общее число детей, родившихся живыми, не может превышать возраст данного лица плюс определенный минимальный возраст, установленный в каждой стране, помноженные на некий коэффициент. Этот коэффициент будет равен 1 в случаях, когда для женщин допускается рождение одного ребенка в год; коэффициент будет равен 1,5 в случаях перерыва в полтора года между родами и т. д. В целях редактирования см. следующий раздел «[Возраст при первом рождении](#)» для определения минимальной разницы в возрасте между матерью и старшим ребенком, родившимся живым;
- б) общее число детей, родившихся живыми, не может превышать суммарное число детей, живущих в домохозяйстве, проживающих в других местах и умерших. Если общее число детей превышает сумму этих слагаемых, то группы редактирования должны решить, какое из них имеет приоритет, и на этой основе могут внести соответствующие коррективы;
- в) если данные собираются как по живущим, так и по умершим детям, общее число этих детей не может превышать число детей, родившихся живыми;
- г) число детей, родившихся живыми, не может быть меньше данных по категории «дети, родившиеся за последние 12 месяцев»;
- д) в зависимости от страны и фактического общего числа детей, родившихся живыми, и детей, живущих на момент переписи, можно использовать матрицу вменения для данных о детях, родившихся за последние 12 месяцев, в целях распределения ответа по возрасту и числу детей, родившихся живыми. Однако необходимо проявлять боль-

шую осторожность при присвоении значения для категории «дети, родившиеся за последние 12 месяцев», когда имеет место пропуск. Для многих стран пропуск для этого элемента означает, что не родился ни один ребенок. Распределенные значения могут исказить данные;

- f) иногда страны собирают данные о детях родившихся живыми, детях, живущих на момент переписи, и другие элементы данных о рождаемости в распределении по признаку пола. В этих случаях отредактированные данные, представленные здесь, определяются в совокупности, но такие страны могут пожелать включать дополнительные проверки для учета дополнительной имеющейся информации. Такие дополнительные проверки включают точное определение того факта, что число мальчиков, родившихся живыми, представляет собой суммарное число мальчиков, живущих на момент переписи, и умерших мальчиков, а число девочек, родившихся живыми, является суммарным числом девочек, живущих на момент переписи, и умерших девочек. Что касается редактируемых данных о детях, в отношении которых не проводится дифференциации по полу, необходимо предпринять соответствующие действия, когда эти суммы не равняются их слагаемым.

c) Связь между данными о детях, родившихся живыми, и детях, живущих на момент переписи

392. Данные о детях, родившихся живыми, и детях, живущих на момент переписи, используются для опосредованных оценок как рождаемости, так и смертности. Результаты переписи или обследования организуются для групп женщин в разбивке по одногодичным или пятилетним возрастным интервалам. С помощью различных алгоритмов получают постоянные или изменяющиеся оценки смертности. Однако, чтобы получить наилучшие результаты, группам редактирования необходимо проявлять осторожность при определении соответствующего вида редактирования имеющихся данных.

393. Частью проблемы, связанной с разработкой общей программы редактирования, является то, что различным странам требуются разные виды информации. Например, в различных странах собираются следующие комплекты данных:

- a) только дети, родившиеся живыми;
- b) дети, родившиеся живыми, и дети, живущие на момент переписи (или без дифференциации полов, или в разбивке по признаку пола);
- c) дети, родившиеся живыми; дети, живущие на момент переписи; и дети, которые умерли (или без дифференциации полов, или в разбивке по признаку пола);
- d) дети, родившиеся живыми; дети, проживающие дома; дети, живущие отдельно; и дети, которые умерли (или без дифференциации полов, или в разбивке по признаку пола).

d) Редактирование, когда регистрируются только дети, родившиеся живыми

394. Если данная страна не применяет метод динамического вменения, недостоверное или недостающее значение для категории «дети, родившиеся

живыми» должно указываться как «неизвестно». В странах, где применяется метод динамического вменения, специалисты должны решить, желают ли они использовать динамическое вменение для всех элементов данных. Если специалисты применяют этот метод, данные для детей, родившихся живыми, можно получить на основе одногодичного возрастного интервала для женщины и по меньшей мере на основе еще одной характеристики. Можно также использовать одномерную матрицу для одногодичного интервала только для матери. Другими характеристиками могут быть такие элементы данных, как достигнутый уровень образования или религия, поскольку известно, что во многих странах рождаемость зависит от уровня образования или различий в религиозной принадлежности.

е) Редактирование, когда регистрируются дети, родившиеся живыми, и дети, живущие на момент переписи

395. Если имеются ответы как на вопрос о «детях, родившихся живыми», так и о «детях, живущих на момент переписи», в программе необходимо определить следующее:

- a) являются ли эти элементы данных внутренне сопоставимыми (равняется ли число детей, родившихся живыми, числу детей, живущих на момент переписи, или оно превышает это число);
- b) согласуется ли каждый элемент данных с возрастом женщины;
- c) согласуется ли показатель «дети, родившиеся живыми» с показателем «дети, родившиеся в последний год» (или последнее деторождение), если такие данные собираются.

396. Демографы используют элементы данных о детях, родившихся живыми, и детях, живущих на момент переписи, для получения косвенных оценок смертности. В связи с этим при редактировании данных необходимо сохранять зависимость между двумя этими элементами. Иногда регистрируется только один из этих двух элементов, а остальные остаются неизвестными. При простом редактировании можно было бы допустить, что в данных по детям, родившимся живыми, отсутствуют данные о смертности, и указывать оба элемента как одинаковые. Однако, если сделать эти два элемента одинаковыми, при косвенной оценке смертности не будут учитываться младенцы, которые могли бы умереть после родов, тем самым занижая показатель смертности и завышая показатель продолжительности жизни. Если таких случаев регистрируется в переписи или обследовании немного, это не приносит большого вреда. Однако, если это происходит с определенной частотой, которую можно было бы ожидать в странах, использующих косвенный метод, последствия могут быть значительными. Соответствующий пример приводится на рисунке 27, ниже.

397. В данном случае супруга сообщает о 5 детях, родившихся живыми, но по какой-то причине число детей, живущих на момент переписи, не было зарегистрировано. Респондент или счетчик не сообщили об этой величине или позже оператор, вводивший эти данные, ошибся с набором на клавиатуре. Многие страны разрабатывают программу редактирования, в которой значение «5» присваивается детям, живущим на момент переписи, исходя из числа детей, родившихся живыми. Однако при этом данные искажаются.

Рисунок 27

Иллюстрация домохозяйства с информацией о рождаемости

Лицо	Степень родства	Пол	Возраст	Дети, родившиеся живыми	Дети, живущие на момент переписи
1	Глава домохозяйства	1	60		
2	Супруг(а)	2	60	5	99
3	Дочь	2	40	3	3
4	Внучка	2	20	1	1
5	Внучка	2	18	0	0
6	Внучка	2	1		

Примечание: 99 = недостающие или недостоверные данные.

398. По сути дела, нет необходимости вообще менять это значение. Страны, которые не применяют метод динамического вменения, возможно, предпочтут не трогать «неизвестное» значение. Естественно, такое решение также приводит к искажениям, поскольку при таком редактировании принимается решение, что «неизвестные» и «известные» ответы имеют одинаковое распределение для таблиц. Если стране требуются данные о детях, родившихся живыми, и о детях, живущих на момент переписи, для определения косвенных оценок смертности, то, вероятно, такая страна сталкивается также с проблемами регистрации данных. В этом случае сохранение неизвестных значений в массиве данных также должно исказить результаты окончательного анализа. Женщин, для которых неизвестны ни число детей, родившихся живыми, ни число детей, живущих на момент переписи, нельзя учитывать при определении оценок смертности, поскольку невозможно определить разницу между детьми, родившимися живыми, и детьми, живущими на момент переписи.

399. В странах, в которых применяется метод динамического вменения, необходимо рассмотреть вопрос об определении недостающих данных, основываясь как минимум на других признаках рождаемости и возрасте женщин. Матрицы вменения могут обновляться, когда имеется достоверная информация о возрасте женщины, числе детей, родившихся живыми, и детей, живущих на момент переписи, и могут применяться, когда недостает такого элемента. Когда недостает информации о детях, родившихся живыми, в матрице вменения будут иметься возраст женщины и число детей, живущих на момент переписи. Когда недостает информации о детях, живущих на момент переписи, в матрице вменения будет приводиться возраст женщины и число детей, родившихся живыми.

400. Кроме того, при разработке матриц вменения важно помнить, что данные о числе детей, родившихся живыми, и числе детей, живущих на момент переписи, должны соответствовать разнице в возрасте между матерью и старшим ребенком (если имеется такая информация) и общему числу детей, родившихся живыми, при конкретном возрасте матери.

401. Например, разница между вмененным числом детей, родившихся живыми, и возрастом матери может быть как минимум 12 лет. В этом случае матрица вменения, использующая группы женщин с возрастным интервалом в 5 лет, в ряде случаев будет почти наверняка исчислять несовместимые данные.

402. На рисунке 28 в сопутствующей матрице вменения в верхней части по горизонтали приводятся значения возраста женщин, а сбоку (по вертикали) приводится число детей, родившихся живыми. Данные на рисунке представ-

Рисунок 28

Исходные значения для определения детей, живущих на момент переписи, когда возраст и число детей, родившихся живыми, достоверны

Дети, родившиеся живыми	Возраст												
	15	16	17	18	19	20	21	22	23	24	25–29	30–34	35+
0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	0	0	0
2		2	2	2	2	2	2	2	2	2	1	1	1
3			3	3	3	3	3	3	3	3	2	2	2
4					4	4	4	4	4	4	3	3	3
5					5	5	5	5	5	5	4	4	4

ляют собой вмененные значения для детей, живущих на момент переписи. В одних случаях ответы будут достоверными, а в других — нет. Если в программе встречаются данные для женщины 19 лет с 5 детьми, родившимися живыми, то показатель в 5 детей, живущих на момент переписи, вероятно, должен превышать критерий разницы в возрасте (разница в возрасте составляет 15 лет исходя из числа детей, живущих на момент переписи, и зарегистрированного возраста). Однако для 15-летней женщины данные ни о 5 детях, родившихся живыми (разница в возрасте составляет 10 лет), ни о 4 детях, живущих на момент переписи (разница в возрасте будет 11 лет), вряд ли будут приемлемыми.

403. Более точной матрицей вменения будет матрица, в которой к молодым женщинам применяется единый возрастной интервал. В этом случае в матрицу вменения будут вноситься лишь достоверные ответы о разнице в возрасте для конкретной возрастной группы, и только достоверные ответы можно будет получать на основе матрицы вменения.

f) Редактирование, когда имеются данные о детях, родившихся живыми, детях, живущих на момент переписи, и умерших детях

404. Данные по категории «дети, родившиеся живыми» представляют собой сумму показателей по категориям «дети, живущие на момент переписи» и «умершие дети». Ниже приводятся методы, иллюстрирующие, как можно ликвидировать любое несоответствие.

i) Когда имеются данные по всем трем элементам

405. Если имеется информация обо всех трех элементах, то в программе необходимо определить:

- a) являются ли все эти три элемента внутренне сопоставимыми, и является ли число детей, родившихся живыми, суммой числа детей, живущих на момент переписи, и числа умерших детей;
- b) соответствуют ли все три элемента возрасту данной женщины;
- c) согласуется ли число детей, родившихся живыми, с числом детей, родившихся в последний год (или последним деторождением), если такие данные собираются.

406. Если все эти элементы согласуются, то редактирование закончено. Однако, если они не согласуются, редактирование должно решить эту пробле-

му. Три элемента могут быть внутренне несопоставимыми: например, у женщины может быть 5 детей, родившихся живыми, но на момент переписи живут только два ребенка, а два — умерли. Группе редактирования следует решить, какая из переменных имеет приоритет по сравнению с другими. Во многих случаях женщина наверняка помнит всех своих детей, которых она родила, хотя она может забыть точное число своих умерших детей. Тогда группа редактирования, возможно, предпочтет согласиться с количеством детей, родившихся живыми, и числом детей, живущих на момент переписи, и провести вычитание, чтобы получить новые, согласующиеся данные о числе умерших детей.

ii) *Когда имеются данные по двум элементам*

407. Поскольку категория детей, родившихся живыми (CEB), представляет собой сумму числа детей, живущих на момент переписи (CS), и числа умерших детей (CD), то если имеются два элемента информации из трех, компьютерная программа может определить третью переменную:

- a) если известны CEB и CS , то $CD = CEB - CS$.
- b) если известны CS и CD , то $CEB = CS + CD$.
- c) если известны CEB и CD , то $CS = CEB - CD$.

Такие тесты обычно проводятся первыми. Как только программа определит, что все три элемента информации являются достоверными и согласующимися, то редактирование закончено.

iii) *Когда имеются данные по одному элементу*

408. Когда известен только один из трех элементов и если страна не применяет метод динамического вменения, два остальных элемента данных следует отнести к категории «неизвестно». Если же страна применяет метод динамического вменения, группе редактирования необходимо определить метод получения по меньшей мере еще одного элемента, а третий элемент следует затем определять путем вычитания или сложения. Двумерную матрицу можно использовать для определения второго значения рождаемости, основанного на первом элементе или одногодичном возрастном интервале для женщин. Если, например, известно число детей, родившихся живыми, то число детей, живущих на момент переписи, можно получить по матрице вменения, как описывалось выше, а затем путем вычитания получить число умерших детей. Аналогичным образом, если известно значение детей, живущих на момент переписи, то число детей, родившихся живыми, получают по матрице вменения исходя из одногодичного возрастного интервала женщины и числа детей, живущих на момент переписи, а число умерших детей получают путем вычитания.

iv) *Когда не имеется данных ни по одному из элементов*

409. Когда нет данных ни об одном из трех элементов, группа редактирования должна принять решение о том, как поступать дальше. Если в стране не применяется метод динамического вменения, то все элементы следует относить к категории «неизвестно» и их не следует применять в косвенных методах определения смертности или рождаемости. В странах, где применяется метод дина-

мического вменения, специалисты должны решить, хотят ли они использовать метод динамического вменения для всех элементов данных.

410. Если специалисты примут решение применять метод динамического вменения, то число детей, родившихся живыми, можно получить на основе одногодичного возрастного интервала для данной женщины и по меньшей мере на основе одной из других характеристик. Можно также применять одномерную матрицу для одногодичного возрастного интервала только матери. К другим характеристикам могут относиться такие элементы, как достигнутый образовательный уровень или вероисповедание.

411. Как только определяется первый элемент, можно предпринять упомянутые выше шаги по редактированию, когда имеются данные только по одному элементу, для получения второй характеристики рождаемости. Затем по первым двум элементам можно определить третий элемент. Все три элемента должны быть совместимыми, потому что матрицы вменения должны обновляться только тогда, когда все элементы сопоставимы. Полученный показатель рождаемости должен также быть сопоставимым с данными для других женщин в рассматриваемом географическом районе, поскольку информация, полученная от этих женщин, используется для обновления матрицы вменения.

g) Редактирование, когда имеются данные о детях, родившихся живыми, детях, живущих дома, детях, живущих вне дома, и умерших детях

i) Когда имеются данные по всем четырем элементам

412. Если имеются данные по всем четырем элементам, в программе необходимо определить:

- a) являются ли все четыре элемента внутренне сопоставимыми, для того чтобы число детей, родившихся живыми, представляло собой сумму числа детей, живущих дома, детей, живущих вне дома, и умерших детей;
- b) согласуется ли каждый из этих четырех элементов с возрастом женщины;
- c) согласуются ли данные о детях, родившихся живыми, с данными о детях, родившихся в последний год (или последним деторождением), если такая информация собирается.

413. Если все эти элементы согласуются, редактирование закончено. Однако, если они не согласуются, при редактировании необходимо устранить эти несоответствия. Как и в случае с тремя описанными выше элементами данных, все четыре элемента могут быть внутренне несопоставимыми. Опять-таки, группе редактирования следует принять решение, какая переменная имеет приоритет перед всеми остальными. Во многих случаях женщина, отвечающая на вопрос, наверняка помнит всех детей, которых она родила, хотя она может и забыть некоторых из детей, кто покинул дом, или точное число умерших детей. Тогда группа редактирования может предпочесть согласиться с числом детей, родившихся живыми, и детей, живущих на момент переписи (сумма числа детей, живущих вне дома, и числа детей, живущих дома), и произвести вычитание, чтобы получить новые, согласующиеся значения для

других переменных. Группе редактирования, возможно, придется разработать алгоритмы для различных комбинаций событий.

ii) *Когда имеются данные по трем из четырех элементов*

414. Число детей, родившихся живыми (*CEB*), представляет собой сумму числа детей, живущих дома (*CLH*), детей, живущих вне дома (*CLA*), и умерших детей (*CD*). Если имеются данные по любым трем из этих четырех элементов, то компьютерная программа может определить четвертую переменную:

если известны *CEB*, *CLH* и *CLA*, то $CD = CEB - CLH - CLA$;

если известны *CLH*, *CLA* и *CD*, то $CEB = CLH + CLA + CD$;

если известны *CEB*, *CLH* и *CD*, то $CLA = CEB - CLH - CD$;

если известны *CEB*, *CLA* и *CD*, то $CLH = CEB - CLA - CD$;

iii) *Когда имеются данные по двум из четырех элементов*

415. Если известны данные только по двум из этих элементов, то группе редактирования необходимо решить, что делать дальше. Как уже отмечалось выше, во многих странах женщины не сообщают о точном числе умерших детей. Другой элемент, который, как правило, опускается, представляет собой информацию о детях, живущих вне жилищной единицы, данные о которых также обычно невозможно получить напрямую. Поэтому следует проявлять особую осторожность при разработке переписного листа, при регистрации и при обработке данных с целью получения наиболее качественных данных по всем элементам рождаемости.

416. Данные о детях, живущих дома (*CLH*), можно получить путем суммирования числа детей в жилищной единице. Если только одна женщина в жилищной единице имеет соответствующий признак родства, простой подсчет должен дать число детей, живущих дома. Если этими признаками родства будут обладать больше женщин, то можно продолжать пользоваться программой редактирования при допущении, что данные о детях будут следовать немедленно за данными о матери в процессе сбора данных. Когда ничто другое не дает результатов, страны, в которых применяется метод динамического вменения, могли бы исчислить число детей, живущих дома, по возрасту матери и с учетом одной из других известных переменных. (См. ниже общие правила для вменения отдельных элементов рождаемости исходя из других элементов и возраста матери). Важно использовать одногодичный возрастной интервал, когда это возможно, а также единое число детей, родившихся живыми, живущих дома, живущих вне дома или умерших.

417. Например, данные о детях, родившихся живыми, и умерших детях могут быть достоверными, а данные о детях, живущих дома, и детях, живущих вне дома, могут быть недостоверными. В этом случае число детей, живущих дома, можно определить, суммируя число детей, имеющих соответствующее родство с матерью (при допущении, что мать является главой домашнего хозяйства). Тогда в наличии будут три из четырех элементов, а четвертый элемент — дети, живущие вне дома, — можно определить путем вычитания: $CLA = CEB - CLH - CD$.

418. Однако, когда известны только два элемента, скорее всего, придется регистрировать детей, родившихся живыми, и детей, живущих дома. Женщины, как правило, охотно сообщают о детях, родившихся живыми, а

информацию о детях, живущих дома, можно обычно получить путем наблюдения или работы с респондентами при регистрации, но такие решения не годятся для получения информации о детях, живущих вне дома, или умерших детях. Тогда при редактировании можно применить матрицу вменения, содержащую возраст женщины, и число детей, родившихся живыми (*CEB*), или, что еще лучше, возраст женщины, число детей, родившихся живыми (*CEB*), и число детей, живущих дома (*CLH*). С помощью этих переменных будет получена информация от той же женщины с теми же характеристиками детей, живущих вне дома (*CLA*).

419. В странах, в которых применяется только двумерная матрица для возраста женщины и числа детей, родившихся живыми (*CEB*), не включающая третий элемент, то есть детей, живущих дома (*CLH*), есть риск получить данные о числе детей, живущих вне дома (*CLA*), которые будут несовместимы с двумя другими. Например, если возраст женщины составляет 25 лет, а число *CEB* составляет 5, по матрице вменения можно получить значение 3 для детей, живущих вне дома. Если число детей, живущих дома, составляет 2, то редактирование данных проводится легко. Число умерших детей должно составлять 0, а элементы данных по рождаемости должны быть $CEB = 5, CLH = 2, CLA = 3, CD = 0$.

420. Однако число детей, живущих дома, в действительности может составлять 4, причем для определения числа детей, живущих вне дома, используется только возраст женщины и число детей, родившихся живыми. В этом случае число 3 для детей, живущих вне дома, приведет к несовместимости этих элементов. Число 5 для детей, родившихся живыми, будет меньше суммарного числа живущих детей (4 дома и 3 вне дома), а общий результат будет равняться 7. Поэтому следует применять трехмерную матрицу: для 5 *CEB* и 4 *CLH*, а значение в матрице вменения может составлять 1 для детей, живущих вне дома (нулевое значение следует определять путем вычитания числа умерших детей). Или в матрице вменения для детей, живущих вне дома, значение должно равняться 0 (а значение 1 должно определяться путем вычитания числа умерших детей). Для других пар известных элементов данных следует составлять аналогичные матрицы вменения, как это показано на рисунке 29.

421. В каждом случае имеются два из четырех элементов. Третий элемент определяется методом динамического вменения, а четвертый элемент — путем сложения или вычитания. Группы редактирования должны решить, какой наилучший путь избрать, исходя из культурных традиций.

Рисунок 29

Образец матрицы вменения, которую необходимо составить для пар известных элементов данных

Если известны эти элементы данных...		Применяйте метод динамического вменения для одного из следующих элементов (а затем производите вычитание или сложение)	
Дети, родившиеся живыми	Дети, живущие дома	Дети, живущие вне дома	Умершие дети
Дети, родившиеся живыми	Дети, живущие вне дома	Дети, живущие дома	Умершие дети
Дети, родившиеся живыми	Умершие дети	Дети, живущие дома	Дети, живущие вне дома
Дети, живущие дома	Дети, живущие вне дома	Дети, родившиеся живыми	Умершие дети
Дети, живущие дома	Умершие дети	Дети, родившиеся живыми	Дети, живущие вне дома
Дети, живущие вне дома	Умершие дети	Дети, родившиеся живыми	Дети, живущие дома

iv) Когда имеются данные только по одному элементу

422. Когда из четырех элементов известен только один, ситуация осложняется. Странам необходимо решить, как им действовать, когда недостает такого значительного объема информации. Если применяется метод динамического вменения, в первой матрице вменения, как отмечалось выше, используется такой элемент данных, как одногодичный возрастной интервал для женщины и еще один известный элемент для построения двумерной матрицы в целях вменения любого из остальных элементов. Как только определяются два элемента, два других остаются неизвестными по определению. Таким образом, дальнейшее применение динамического вменения для третьего элемента не должно приводить к несовместимости с остальными элементами, поскольку они неизвестны. Рассмотренная выше схема для двух известных и двух неизвестных элементов применяется для получения третьего элемента. Тогда четвертый элемент определяется путем вычитания. Все четыре элемента данных должны быть совместимыми.

v) Когда не имеется данных ни по одному из элементов

423. Когда нет данных ни по одному из этих четырех элементов, группа редактирования должна решить, как продолжать работу без них. Если в данной стране не применяется метод динамического вменения, все элементы должны быть отнесены к категории «неизвестно» и не должны использоваться в косвенных методах оценки уровня смертности или рождаемости. В странах, где применяется метод динамического вменения, специалисты должны решать, хотят ли они применять вменение для всех элементов данных.

424. Если специалисты принимают решение применять метод динамического вменения, то данные о числе детей, родившихся живыми, можно получить на основе одногодичного возрастного интервала для данной женщины и по меньшей мере на основе одной из других характеристик. Можно также применять одномерную матрицу для одногодичного возрастного интервала только матери. К другим характеристикам могут относиться такие элементы, как уровень образования или вероисповедание, поскольку известно, что во многих странах дифференциация показателей рождаемости существует в зависимости от различных уровней полученного образования или от религиозной принадлежности.

425. Как только определен первый элемент, использованный выше метод, когда известен только один элемент, можно использовать для получения второго элемента данных по рождаемости. Затем третий элемент можно получить по двум первым элементам, а четвертый элемент можно определить путем вычитания. Все эти четыре элемента должны быть совместимыми, потому что матрицы вменения необходимо обновлять только в случае совместимости всех элементов данных. Полученный показатель рождаемости также должен быть совместимым с показателями для других категорий женщин в данном географическом районе, поскольку информация, полученная от этих женщин, используется для обновления матрицы вменения.

h) Особый случай при наличии пяти или более элементов

426. В связи с тем что в некоторых небольших странах все более важную роль стала играть международная миграция, проводится сбор дополнительной

информации о детях, проживающих вне дома. Когда переменная для позиции «дети, живущие вне дома» подразделяется на категории «дети, живущие вне дома в своей стране» и «дети, живущие вне дома в других странах», процедуры при работе с четырьмя переменными — дети дома, дети вне дома, умершие дети и общее число детей — должны быть расширены, с тем чтобы учесть эту дополнительную информацию. Кроме того, как отмечалось ранее, хорошей идеей является введение единой строки матрицы для каждой возрастной группы женщин с полной информацией по рождаемости, вводимой в случае, когда все элементы являются достоверными, внутренне совместимыми и совместимыми с возрастом женщины; тогда в случае несовместимости информации по рождаемости (в том числе с возрастом) можно применять вменение, используя всю соответствующую строку целиком.

и) Значение единого источника данных для всех элементов по рождаемости

427. Таким образом, очень важно, если вообще возможно, исчислять все элементы исходя из данных по одной женщине, когда не известен ни один из показателей. Чтобы убедиться, что вся информация берется из одного источника (то есть от одной и той же женщины), возможно, потребуется разработать матрицы вменения, в которых применяются все данные о рождаемости. В этом случае матрицы вменения можно было бы обновлять лишь тогда, когда в программе редактирования установлено, что все элементы данных по рождаемости согласуются между собой. Как указано выше, в пункте [426](#), лучше не производить вменение элемента за элементом, однако, когда отсутствуют несколько элементов данных, следует использовать полную информацию, относящуюся к другой женщине.

й) Связь между данными о собственных детях, детях, проживающих в доме, и детях, живущих на момент переписи

428. Когда страны используют метод учета собственных детей для оказания помощи при проверке редактирования данных по рождаемости по мере его разработки и осуществления, информация по детям, живущим в доме, а также содержащаяся в матрице «мать–ребенок», может помочь в оценке надежности результатов редактирования. Поскольку лишь немногие страны используют этот метод для оказания помощи в процессе редактирования, он остается экспериментальным; тем не менее результаты его применения выглядят многообещающими.

7. Рождаемость: дата рождения последнего ребенка, родившегося живым, и рождение детей за 12 месяцев до переписи

429. Информация о последних рожденных детях помогает в получении оценок текущих показателей рождаемости непосредственно перед началом переписи или обследования. Одним из подходов является получение данных, касающихся даты рождения (день, месяц и год) последнего ребенка, родившегося живым, и пола этого ребенка, и выяснение того, жив ли этот ребенок. Второй подход состоит в сборе информации о детях, рожденных за последние

12 месяцев до переписи; использование второго подхода легче для счетчиков и респондентов, поскольку требуется лишь простой ответ «да» или «нет», а не представление конкретной даты.

430. В ходе обработки данных оценку числа детей, родившихся живыми, в течение 12 месяцев, непосредственно предшествующих дате переписи, можно получить из информации о дате рождения последнего живорожденного ребенка (а затем сохранить ее для перекодирования). Для оценки текущих по возрасту коэффициентов рождаемости и других показателей рождаемости данные, получаемые с помощью такого метода, являются более точными, чем информация о числе рождений, приходившихся на одну женщину в течение 12 месяцев, непосредственно предшествующих переписи (United Nations, 2008, paras. 2.188–2.191).

431. Следует отметить, что информация о дате рождения последнего ребенка, родившегося живым, не позволяет получить данные об общем числе детей, родившихся живыми в течение 12-месячного периода. Даже в случае отсутствия ошибок при сообщении данных о последнем живорожденном ребенке, эти сведения позволяют установить число женщин, которые родили по крайней мере одного живого ребенка в течение 12 месяцев, а не число рождений, поскольку лишь незначительная доля женщин родит более одного ребенка за один год (United Nations, 2008, para. 2.189).

432. Информация должна собираться только в отношении женщин в возрасте от 15 до 50 лет, которые сообщили о рождении хотя бы одного живого ребенка в течение всей своей жизни. Кроме того, информация должна собираться в отношении женщин всех категорий семейного положения, по которым собираются сведения о детях, рожденных живыми, в разбивке по признаку пола. Если сбор данных о родившихся живыми детях осуществляется лишь по выборке женщин, то информация о текущем уровне рождаемости также должна собираться по той же выборке (United Nations, 2008, para. 2.190).

433. В программу редактирования следует включать следующие виды редактирования. Дата рождения последнего ребенка должна указываться для всех женщин с возрастом между установленным в стране минимальным и максимальным возрастом. Программа должна обеспечивать проверку на соответствие. Например, следует исключать всю информацию по мужчинам и женщинам, не включенным в отобранную возрастную группу. Кроме того, женщины в отобранной возрастной группе, у которых количество родов в прошлом больше 0, должны указывать достоверное число, месяц и год последнего деторождения (или указывать на то, имело ли место рождение ребенка за последние 12 месяцев, если используется такая форма вопроса).

434. Группе редактирования необходимо решить, должны ли быть достоверными число и месяц рождения: группы редактирования, применяющие динамическое вменение, могут вменить число и месяц, когда они отсутствуют; группы, не применяющие метод динамического вменения, относят число и месяц к категории «неизвестно». Если специалисты по содержательной части переписи, в качестве которых обычно выступают демографы, хотят получить фактический возраст матерей на момент рождения их детей для перекодирования в целях проведения анализа рождаемости, тогда, вероятно, должен быть вменен как минимум месяц последних родов, если эти сведения отсутствуют. Затем можно получить новый код.

435. Аналогичным образом, некоторые демографы хотят проанализировать данные по месяцам, прошедшим с момента последних родов. Редактирование данных о годе и месяце рождения последнего ребенка дает информацию, необходимую для получения числа полных месяцев с даты рождения последнего ребенка. Когда таким образом получена дата последних родов, она может быть использована с целью определения новых кодов для месяцев с даты последних родов. (См. [приложение I](#) по информации получения новых кодов для месяцев с даты рождения последнего ребенка).

436. Если информация о годе рождения последнего ребенка отсутствует или является недостоверной, страны, не применяющие метод динамического вменения, могут записать в соответствующих графах «не указано» или «неизвестно». В странах, применяющих динамическое вменение, для получения даты рождения последнего ребенка можно применять другие переменные, такие как возраст и число детей, родившихся живыми.

437. В силу важности использования даты рождения последнего ребенка для измерения текущих национальных, региональных и местных показателей рождаемости, необходимо рассмотреть возможность проведения дополнительных проверок. Эффективный процесс редактирования предполагает проведение проверки в рамках домохозяйства для выявления ребенка или детей в возрасте от 0 лет и использование родственных отношений матери и этого ребенка или этих детей (или личного номера матери для данного ребенка, если эти данные собираются) для того, чтобы определить, регистрируется ли ребенок как последние роды для своей матери. Проверки должны проходить по двум направлениям: данные по новорожденным должны сверяться с данными по матерям, а последние рожденные дети должны сверяться по спискам домохозяйств.

438. Те страны, которые собирают также данные о числе смертей за год до проведения переписи или обследования, могут принять решение включить сравнительную проверку смертей младенцев (0 лет) за год до переписи и рождений последнего ребенка, когда последний рожденный ребенок записывается как «умерший» или «уже не живой». Хотя такая проверка нецелесообразна, если мать отсутствует по причине смерти или переезда или если по какой-либо причине о наличии ребенка не сообщается, определенную долю смертей младенцев можно тем не менее проверить таким способом.

8. Рождаемость: возраст при рождении первого ребенка

439. Возраст матери при рождении первого живорожденного ребенка используется для косвенной оценки рождаемости, основанной на данных о первых родах, и в целях получения информации о наступлении детородного возраста. Если этот вопрос включается в перепись, то информацию следует получить по каждой женщине, которая родила живым хотя бы одного ребенка (United Nations, 2008, para. 2.193).

440. Возраст при рождении первого ребенка определяется либо напрямую с помощью точно определенного элемента «возраст при рождении первого ребенка», либо по разнице в возрасте между текущим возрастом матери и возрастом старшего ребенка, если известен возраст старшего ребенка. Самый ранний возраст для рождения детей, установленный в странах, не является минимальным биологическим возрастом. Если, например, минимальный допус-

тимый возраст при рождении первого ребенка составляет 13 лет, то респондент может сообщить, а счетчик может зарегистрировать возраст какой-либо женщины при рождении, составляющий 11 или 12 лет. Тогда группам редактирования следует решить, менять ли минимальный допустимый возраст, исключить данные о деторождении или изменить либо возраст матери, либо возраст при первом рождении (применяя либо возраст ребенка, либо ее возраст в зависимости от переменных, используемых для определения разницы в возрасте). Аналогичным образом группы редактирования должны решить, какой «предельный возраст» является максимальным возрастом при рождении первого ребенка. Хотя женщины могут рожать детей в возрасте свыше 50 лет, это событие случается не очень часто, и чтобы исправить ошибки, при редактировании необходимо установить, какие возрастные отклонения являются реальными.

441. Важно помнить, что минимальный или максимальный возраст при рождении первого ребенка (и разница в возрасте между матерью и ее старшим ребенком, проживающим в домашнем хозяйстве) должны соответствовать обычаям и традициям страны. Специалисты должны решить, когда то или иное значение является случайным, а не законным возрастом при рождении первого ребенка. Когда устанавливаются соответствующие правила, то специалисты должны определить пути решения этой проблемы. Если метод динамического вменения не применяется, программа должна использовать категорию «неизвестно». Когда метод динамического вменения применяется, он может устанавливать возраст при рождении первого ребенка на основе данных о других женщинах того же возраста с аналогичным числом детей, родившихся живыми. Специалисты, разрабатывающие матрицу вменения, возможно, пожелают учитывать такие факторы, как проживание в городских/сельских районах (если рождаемость отличается в этих двух районах), участие женщины в рабочей силе (хотя текущий статус участия в рабочей силе не всегда соответствует ее статусу при рождении первого ребенка) и достигнутый уровень образования.

9. Смертность

442. Информация о числе смертей за последние 12 месяцев используется для оценки уровня и структуры смертности в разбивке по полу и возрасту в тех странах, где не представляется возможным непрерывно получать удовлетворительные статистические данные о смертности на основании записи актов гражданского состояния. Для обеспечения надежности оценочных данных по этому вопросу необходимо получить как можно более полную и достоверную информацию о числе смертей за последние 12 месяцев в разбивке по полу и возрасту. Широкая практика последних десятилетий по включению вопросов о смертности в опросные листы переписей привела к значительным улучшениям в плане применения косвенных методов для оценки показателей смертности взрослого населения (United Nations, 2008, para. 2.194).

443. В идеальном случае смертность по каждому домохозяйству следовало бы определять в виде общего числа смертей за последние 12 месяцев, предшествующих дате переписи. В случаях когда невозможно получить информацию о числе смертей за последние 12 месяцев, рекомендуется по крайней мере собрать данные о числе смертей среди детей в возрасте менее одного года. В отношении каждого зарегистрированного умершего лица должны также собираться данные о его имени, возрасте, поле, дате (число, месяц, год) смерти. Следует принять

меры к тому, чтобы четко объяснить респонденту, о каком именно отчетном периоде времени идет речь, во избежание ошибок, которые могут возникнуть вследствие недопонимания этого момента. Например, отчетный период можно точно определить, ориентируясь на какие-либо праздничные или исторические даты, отмечаемые в конкретной стране (United Nations, 2008, para. 2.195).

а) Возраст и пол умершего лица

444. Во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда* (United Nations, 2008) предлагается собирать данные об имени, возрасте и поле, а также числе, месяце и годе смерти лиц, которые умерли в течение года, предшествующего переписи. В странах, в которых не применяется метод динамического вменения, можно присваивать категорию «неизвестно» для каждой из этих переменных, если они недостоверны. Для стран, где применяется метод динамического вменения, можно использовать возраст (по возрастным группам), пол и год смерти в качестве параметров матриц вменения для других переменных. Фактические матрицы вменения, вероятнее всего, относятся к какой-либо конкретной стране, и членам группы редактирования придется работать совместно для разработки соответствующих матриц вменения. Для разработки наиболее приемлемой схемы редактирования может оказаться полезной структура населения страны или административно-территориальных единиц.

б) Причина смерти

445. В настоящее время некоторые страны собирают информацию о причине смерти для случаев летальных исходов в течение 12 месяцев до переписи. Из-за чувствительного характера этого вопроса, а подчас из-за сложности получения информации на местах в странах может применяться формулировка «Были ли причиной смерти несчастный случай или насилие?» для получения косвенной информации по ВИЧ/СПИДу в отдельных возрастных группах. Редактирование этого элемента данных обычно предполагает допущение о том, что если нужная информация не будет собрана или окажется недостоверной, то обычно по этой категории устанавливается значение «неизвестно». Если страна применяет метод вменения, то надлежит использовать данные «hot deck» с разбивкой по признаку пола и возрастным категориям 0, 1–4, а затем — по пятилетним возрастным группам.

с) Материнская смертность

446. В ходе текущего раунда переписей во все большем числе стран также задают вопросы о том, было ли умершее лицо женского пола и была ли она беременной на момент смерти. Этот элемент данных помогает определить уровень материнской смертности на национальном и региональном уровнях. Редактирование этого элемента может потребовать присвоения категории «неизвестно» недостоверным или пропущенным записям. Однако если страна выбирает метод вменения, то значения «hot deck» должны применяться очевидно только в отношении женщин и исключительно для вероятного возраста беременности (скорее всего, 12–54 года) и, вероятно, в возрастной разбивке по отдельным годам, а не по пятилетним возрастным группам.

d) Младенческая смертность

447. И наконец, во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда* (пункт 2.191) предлагается собирать информацию о детской смертности в отношении детей, «родившихся за последние 12 месяцев». Обычно такой вопрос должен задаваться только совместно с элементом данных о числе рождений за 12 месяцев до переписи. В случае использования другого текущего элемента данных по рождаемости — даты рождения последнего ребенка — этот элемент данных, вероятно, использовать не следует.

448. Данные о смертях и рождениях в течение года до проведения переписи помогают странам с налаженной системой учета естественного движения населения осуществлять проверку своих показателей младенческой смертности как на национальном, так и на региональном уровне. Страны, не имеющие налаженной системы учета естественного движения населения, могут использовать указанную информацию для получения оценок младенческой смертности. И вновь редакционная сравнительная проверка последних рождений для выявления числа лиц, умерших в течение года перед переписью, по сравнению с числом смертей младенцев младше года в этот же период представляет собой надлежащий метод и может предоставить полезную информацию об уровне младенческой смертности.

449. Редактирование по этому элементу данных требует некоторого осмысления и обычно применимо лишь к одной конкретной стране. В идеальном случае информация о детях, родившихся живыми, и детях, живущих на момент переписи, может быть использована для проверки сообщаемой информации; при наличии единственной взрослой женщины в домохозяйстве такая проверка сравнительно проста. При наличии нескольких женщин в жилищной единице необходимо внимательно отнестись к тому, чтобы дети были правильно соотнесены с соответствующими женщинами.

10. Сиротство в случае смерти матери/отца и номер строки матери

450. При сборе информации о сиротстве следует задавать два прямых вопроса: *a)* жива ли родная мать лица, зарегистрированного в домохозяйстве, на момент переписи; и *b)* жив ли родной отец лица, зарегистрированного в домохозяйстве, на момент переписи. Это исследование должно дать информацию о биологических родителях. Следовательно, необходимо принять меры к тому, чтобы исключить приемных и патронатных родителей. Поскольку в живых, как правило, остается более одного ребенка, который может сообщить о своем сиротстве, необходимо составить вопросы так, чтобы избежать двойного учета родителей при ответах, предоставленных их детьми (братьями/сестрами). В этих целях следует задать два дополнительных вопроса, а именно: *c)* является ли респондент самым старшим из выживших детей своей матери; и *d)* является ли респондент самым старшим из выживших детей своего отца. Таблицы должны строиться с учетом только самого старшего выжившего ребенка (United Nations, 2008, para. 2.199).

451. Редактирование таких элементов, как «мать, живущая на момент переписи» и «номер строки матери», взаимосвязано и должно проводиться сов-

местно. В отношении тех лиц, которые указывают в графе «мать, живущая на момент переписи» иные ответы кроме «да», необходимо проверить на достоверность номер строки матери; если запись сделана правильно, то матери, живущей на момент переписи, следует присвоить код для ответа «да». В отношении тех лиц, которые в графе «мать, живущая на момент переписи» указывают иное, чем «да», номер строки матери следует проверять, для того чтобы посмотреть, равен ли он 00 или номеру строки женщины в возрасте 12 лет или старше. И если один из этих случаев подходит, то программа исходит из того, что у данного лица есть мать, и демографическому статусу матери присваивает код «да». Если номер строки матери записан неправильно и матери, живущей на момент переписи, присвоен код «нет» или «не знает», то запись номера строки матери следует устранить. Во всех других случаях матери, живущей на момент переписи, следует присваивать код для ответа «не знает», а любую запись в графе «номер строки» следует устранять.

452. Конкретная страна может предпочесть не редактировать номер строки матери в отношении тех лиц, которые указали «нет» или «не знает», если мать жива на момент переписи. Во всех других случаях номер строки следует проверить на согласованность или присвоить, используя такие показатели, как степень родства того или иного лица и номер строки, пол, степень родства и возраст того лица, которое указано как мать. Когда имеются несоответствия или мать нельзя определить, то можно присвоить код для ответа «живет в другом месте». Следует отметить, что при редактировании структуры данных в случае, если глава домохозяйства не является первым лицом, и затем сдвигается на позицию первого лица, то номер строки матери может потребовать коррекции на одного или нескольких человек.

В. Характеристики миграции

453. В результате естественного прироста (рождаемость и смертность) и чистой миграции демографическая ситуация в стране со временем меняется. Миграция может быть долгосрочной (с рождения) или краткосрочной, которая определяется по предыдущему месту жительства и продолжительности проживания или по предыдущему конкретному моменту времени. Поскольку эти элементы данных часто взаимосвязаны, то некоторым странам, возможно, целесообразнее проводить совместное редактирование этих данных аналогично редактированию основных демографических переменных. Если используется нисходящий принцип, то порядок редактирования приобретает особое значение, поскольку некоторые элементы необходимо редактировать прежде других.

454. Элементы данных по миграции зачастую требуют более подробных кодов по сравнению с другими данными, поскольку в целях планирования и осуществления политики могут потребоваться более мелкие территориальные единицы. Подробная информация о небольших районах может потребоваться при планировании кадров для новой школы или больницы. Кроме того, могут потребоваться различные схемы кодирования и различные приемы редактирования для разных мест в стране и за ее пределами.

455. Традиционно в большинстве стран не наблюдаются значительные потоки международной миграции, поэтому основной упор делается на внутренней миграции, которая остается первоочередной задачей. Тем не менее в

мире растущей глобализации все больше внимания уделяется также международной миграции.

456. Для изучения миграции внутри страны (внутренней миграции) данные о месте рождения внутри страны и количестве лет, прожитых в данном районе, следует проверять на согласованность между собой, поскольку между этими двумя элементами существует очевидная взаимосвязь. Кроме того, существует некоторая обоснованная взаимосвязь между ответами разных членов домохозяйства. Например, если нет ответа на вопрос о количестве лет, прожитых ребенком в данном районе, его можно вычислить по ответу матери, а программа редактирования проверит, чтобы вмененное значение не превышало возраст ребенка.

457. Что касается международной миграции, то основное значение имеют страна рождения и год въезда в страну.

1. Место рождения

458. Под местом рождения в первую очередь понимается страна, в которой данное лицо родилось. Следует отметить, что страна рождения не обязательно должна быть связана с гражданством, что является отдельной темой (см. United Nations, 2008, para. 2.92 to 2.102 и раздел по стране рождения в настоящем издании). Для лиц, рожденных в стране проведения переписи (местных уроженцев), понятие о месте рождения включает также определенный вид территориальной единицы страны, в которой жила мать данного лица на момент его рождения. Однако в некоторых странах место рождения местных уроженцев определяется как территориальная единица, в которой фактически родилось данное лицо. Каждая страна должна уточнить, каким определением она пользовалась при проведении переписи (United Nations, 2008, para. 2.57).

a) Взаимосвязь между данными о стране рождения и количеством лет, прожитых в данном районе

459. Записи о месте рождения и продолжительности проживания можно проверить на согласованность, поскольку между этими двумя элементами существует тесная взаимосвязь. Кроме того, взаимосвязи существуют между разными членами домохозяйства, и можно делать допущения, исходя из данных о других членах семьи относительно того, мигрировало ли данное лицо или нет.

b) Присвоение категории «неизвестно» для недостоверных записей о месте рождения

460. Если страна не собирается применять метод динамического вменения, любые недостоверные ответы о месте рождения должны обозначаться категорией «неизвестно». Обычно любая страна не должна редактировать несогласующиеся ответы о членах семьи или географических районах, если только нет ошибок в кодировании.

c) Применение метода статического вменения для данных о месте рождения

461. Запись о стране рождения следует менять, только если она выходит за определенные пределы. Если для количества лет, прожитых в районе, при-

своен код «всегда», то следует присваивать код для страны. Если указана иная запись, чем «всегда», можно использовать информацию о предыдущем лице. Например, если предыдущее лицо — это мать, то количество лет, в течение которых мать жила в данном районе, можно сопоставить с возрастом этого лица. Если же количество лет, касающееся матери, больше или равно возрасту данного лица, то следует присвоить код для ответа «эта страна»; в противном случае необходимо присвоить категорию «страна рождения матери». Если страну рождения нельзя присвоить исходя из данных о матери, аналогичным образом можно использовать данные о других родственниках этого лица. Если после таких проверок запись классифицировать нельзя, то в графе «страна рождения» можно указать «неизвестно».

462. Поскольку в настоящее время страны предоставляют выборки своих данных для открытого использования, важно в ходе редактирования предоставлять конкретный код для пустых записей, поскольку они пропущены счетчиками, следующими порядку игнорирования тех или иных граф. Иными словами, часто в переписном листе счетчику дается указание пропустить вопрос о месте рождения, если лицо всегда проживало в данном месте. В ходе редактирования конкретным местам должен присваиваться код для оказания помощи пользователям, чтобы позднее им не приходилось сверяться с двумя источниками при составлении сводных таблиц.

d) Применение метода динамического вменения для данных о месте рождения

463. Как указано выше, запись о стране рождения следует менять, только если она выходит за определенные пределы. Если в графе о количестве лет, прожитых в данном районе, указано «всегда», то код для ответа «эта страна» следует присвоить в качестве страны рождения. Если запись иная, чем «всегда», то следует изучить информацию, полученную от других членов домохозяйства, чтобы найти сведения о стране рождения данного лица.

e) Присвоение места рождения, если у данного лица есть мать

464. Если в графе «страна рождения» стоит пробел или запись недостоверна, а в графе «продолжительность проживания» указано не «всегда», а что-то иное, то можно произвести поиск данных о матери этого лица. Если в домохозяйстве есть мать, то изучается запись о продолжительности проживания матери. Если в графе «количество лет, прожитых в данном месте» для матери указано «всегда», то в качестве страны рождения данного лица можно записать «эта страна». Если мать этого лица не всегда жила в данном районе, а возраст данного лица меньше или равен количеству лет, прожитых матерью в этом месте, программа может также записать «эта страна» в графу о стране рождения. Если возраст данного лица превышает количество лет, которые мать прожила в данном районе, а сведения о стране рождения матери достоверны, то страна рождения этого лица классифицируется так же, как и страна рождения матери.

f) Присвоение места рождения для ребенка главы домохозяйства

465. Если мать данного лица в домохозяйстве не указана, а рассматриваемое лицо является сыном или дочерью главы домохозяйства, то для получения данных о месте рождения можно произвести несколько проверок, используя

информацию из записи о главе домохозяйства. Если в записи о главе домохозяйства в графе «количество лет, прожитых в данном районе» указано «всегда», то программа должна указать «эта страна» в качестве страны рождения в записи об этом лице. Если в графе «количество лет, прожитых в данном районе» для главы домохозяйства указано не «всегда», а возраст рассматриваемого лица меньше или равен количеству лет, которые глава домохозяйства провел в этом районе, то программа также должна указать «эта страна» в качестве страны рождения данного лица. Однако, если возраст данного лица превышает количество лет, которые глава домохозяйства провел в этом районе, программа должна записать название страны рождения главы домохозяйства, если у него указан правильный код страны рождения.

г) Присвоение места рождения для ребенка, не являющегося ребенком главы домохозяйства

466. В зависимости от того, старше или младше то или иное лицо определенного возраста (возраста X), установленного группой редактирования, могут производиться совершенно разные вменения. Если возраст меньше значения X, страну рождения необходимо исчислять исходя из первой предшествующей записи для ребенка в возрасте до X лет по возрасту и полу.

h) Присвоение места рождения для взрослых замужних женщин

467. Если возраст этого лица составляет X или более лет и этим лицом является женщина, то программа должна проверить, есть ли у нее в домохозяйстве муж. Если у этой женщины муж есть и у него правильный код страны рождения, то программа должна указать код его страны рождения в ее записи. Если муж не имеет достоверного кода страны рождения, нужно проверить его запись о количестве лет, прожитых в этом районе. Если для количества лет, прожитых мужем в данном районе, подходит запись «всегда», то код страны рождения женщины должен быть «эта страна». Если количество лет, прожитых мужем в данном районе, не соответствует записи «всегда», то страну рождения женщины следует исчислить исходя из возраста и пола.

и) Присвоение места рождения для взрослых незамужних женщин

468. Хотя женщина, чей возраст превышает какой-то минимальный возраст, установленный группой редактирования, не имеет мужа в домохозяйстве, она может быть матерью ребенка в этом домохозяйстве. В таком случае программа должна отыскать ее старшего ребенка. Если ребенка найти нельзя, программа может исчислить страну рождения по возрасту и полу. Если код ребенка для страны рождения указан правильно и количество сообщенных лет, прожитых матерью в данном районе, больше возраста ребенка, то программа должна исчислить страну рождения по возрасту и полу. Если количество лет, прожитых матерью в данном районе, меньше или равно возрасту ребенка, то программа должна присвоить ей страну рождения ребенка.

j) Присвоение места рождения для лиц мужского пола

469. Для определения места рождения того или иного лица мужского пола программа редактирования должна попытаться найти данные о его жене или, если он является главой домохозяйства, программа должна попытаться

найти данные о его детях. Сначала программа пытается найти данные о жене этого человека. Если она их находит, а количество лет, прожитых им в данном районе, меньше или равняется ее показателям, то в записи об этом человеке указывается страна рождения жены. Если количество лет, прожитых этим человеком в данном районе больше, чем у его жены, то страну рождения следует исчислить по возрасту и полу, используя матрицу вменения. Если мужчина является главой домохозяйства, имеет сына или дочь в этом домохозяйстве и количество лет, прожитых им в этом районе, меньше или равняется возрасту ребенка, то программа должна указать ту же страну рождения, что и его ребенка. Если длительность проживания данного лица в этом районе больше, чем возраст его ребенка, то программа должна произвести вменение по возрасту.

2. Гражданство

470. Сведения о гражданстве должны собираться таким образом, чтобы население можно было разбить на три категории: *a)* граждане по рождению; *b)* граждане, натурализованные путем подачи заявления о предоставлении гражданства, путем оптации, брака или иными способами; и *c)* иностранцы. Помимо этого следует собирать сведения о стране гражданства иностранцев. Важно записать страну гражданства как таковую, а не использовать прилагательное, указывающее на гражданство, так как некоторые из этих прилагательных одновременно обозначают этнические группы.

471. Кодирование информации о стране гражданства должно проводиться достаточно подробно, чтобы обеспечить возможность идентифицировать каждую из всех стран гражданства, представленных среди иностранного населения в стране. В целях кодирования странам рекомендуется использовать числовую систему кодирования, представленную в документе «Стандартные коды стран и районов для использования в статистике» (Standard Country or Area Codes for Statistical Use) (United Nations, 1999). Применение стандартных кодов для классификации иностранного населения по стране гражданства повысит актуальность таких данных и позволит организовать обмен информацией между странами об их иностранном населении. Если та или иная страна решит объединить страны гражданства в большие группы, рекомендуется прибегнуть к использованию стандартных региональных и субрегиональных классификаций, описанных в вышеупомянутой публикации (United Nations, 2008, para. 2.97).

a) Редактирование данных о гражданстве

472. Гражданство зависит от принятых в каждой стране определений. В большинстве стран, но не во всех, лица, родившиеся в стране, автоматически становятся гражданами этой страны по рождению. Следовательно, при редактировании необходимо рассмотреть взаимосвязь между местом рождения и гражданством и, возможно, присвоить категорию «граждане по рождению» тем лицам, которые родились в этой стране.

b) Взаимосвязь между этнической/расовой принадлежностью и гражданством

473. Некоторые страны осуществляют также сбор информации об «этнической принадлежности» или «расовой принадлежности», которая может дать дополнительные сведения, с тем чтобы использовать их при определении

гражданства, особенно когда полученный ответ является недостоверным. Во многих странах данные об этническом происхождении и гражданстве мигрантов в первом и втором поколениях должны почти полностью согласовываться между собой. Для стран с давней историей международной миграции эта характеристика, возможно, не имеет такой ценности, но ее все же можно рассматривать наряду с другими переменными.

с) Взаимосвязь между натурализацией и гражданством

474. В странах, где существует натурализация, требования в отношении натурализации могут включаться, а могут и не включаться в пункты переписи. Если, например, требуется период постоянного проживания, то для проверки соблюдения срока, необходимого для натурализации, можно использовать элемент данных «продолжительность проживания». Затем, если то или иное лицо родилось за границей и указало неправильный или не согласующийся ответ в отношении гражданства, группы редактирования могут указать категорию «натурализованный» для обозначения гражданства. Другим лицам, которые не подпадают под требование о продолжительности проживания, необходимой для натурализации, может присваиваться запись «иностранец» методом вменения «cold deck».

d) Взаимосвязь между продолжительностью проживания и гражданством

475. Элемент «продолжительность проживания» может не включаться в переписной лист или может оказаться неоднозначным при определении гражданства, или же группа редактирования может отказаться от его использования. Затем, если данные по гражданству указаны неправильно или не согласуются с местом рождения, то, если не применяется метод динамического вменения, указывается ответ «неизвестно». Страны, применяющие метод динамического вменения для определения недостоверных величин, должны пользоваться по меньшей мере двумя характеристиками (и одной из них, вероятно, должен быть элемент «место рождения») для получения «известной» информации, исходя из сведений об аналогичных лицах в данном географическом районе.

3. Продолжительность проживания

476. Под продолжительностью проживания имеется в виду промежутки времени вплоть до даты проведения переписи, выраженной в полных годах, в течение которых это лицо проживало: *a)* в населенном пункте, который является его или ее постоянным местом жительства в момент переписи; и *b)* в крупной или более мелкой административно-территориальной единице, в которой расположен этот населенный пункт (United Nations, 2008, para. 2.64).

a) Редактирование данных о продолжительности проживания

477. Как и страна рождения, продолжительность проживания имеет важное значение при составлении статистических данных о перемещении населения. В некоторых случаях какая-то подгруппа населения может быть

намного более мобильной, чем вся нация в целом. При редактировании этого элемента данных учитывается место рождения данного лица и ответы других членов домохозяйства. Данные о «продолжительности проживания» должны редактироваться вместе со сведениями о «прежнем местожительстве» или «местожительстве в определенный момент в прошлом».

b) Проживание и продолжительность де-факто/де-юре

478. Результаты редактирования могут зависеть от того, как проводится перепись: де-факто или де-юре. Поскольку во время переписи де-юре собираются сведения о постоянном месте жительства, данные о продолжительности проживания могут не выявить ту же информацию, что и при переписи де-факто, когда регистрация проводится по месту проживания в ночь переписи. Кроме того, при кодировании и редактировании необходимо учитывать тех, кто или «всегда» жил в данном месте или «никогда не покидал» его. В отношении этих лиц программа редактирования должна пропустить проверку на согласованность и другие приемы редактирования.

c) Взаимосвязь между возрастом и продолжительностью проживания

479. На начальном этапе процесса редактирования следует произвести проверку на согласованность между возрастом и местом рождения и на достоверность данных о количестве лет, прожитых в данном населенном пункте или административно-территориальной единице. Количество лет, прожитых тем или иным лицом в одном населенном пункте или административно-территориальной единице, не может превышать возраст этого лица. Кроме того, тот, кто родился за пределами страны, не мог всегда проживать в данном населенном пункте или административно-территориальной единице. Программа должна приписывать «всегда» при указании количества лет, прожитых в населенном пункте или административно-территориальной единице, если это количество лет превышает возраст, а страной рождения является эта страна. Если количество лет, прожитых в населенном пункте или административно-территориальной единице, превышает возраст, а страной рождения является не эта страна, то возраст данного лица следует соизмерять с количеством лет, прожитых в населенном пункте или административно-территориальной единице. В этом случае исходят из того, что рассматриваемое лицо, хотя и родилось за пределами этой страны, переехало в данный населенный пункт или административно-территориальную единицу в возрасте менее одного года.

d) Взаимосвязь между местом рождения и продолжительностью проживания

480. В том случае, если вводимые данные выходят за определенные пределы, следует произвести такую же проверку, как и в отношении места рождения. Нужно найти предыдущих родственников (мать, главу домохозяйства, мужа, ребенка). Вменение должно производиться на основе найденной информации. Однако прежде чем приписать то или иное значение, нужно убедиться, что оно согласуется с возрастом и местом рождения того лица, запись о котором редактируется.

е) Для тех лиц, которые всегда жили в данном месте

481. Если на вопрос о количестве лет, прожитых в данном населенном пункте или административно-территориальной единице, то или иное лицо отвечает «всегда», но страна его рождения не «эта страна», то группа редактирования, возможно, соотнесет возраст данного лица с продолжительностью проживания в данном населенном пункте или административно-территориальной единице. Специалисты будут исходить из того, что, хотя человек родился за пределами этой страны, он переехал в данный населенный пункт или административно-территориальную единицу в возрасте менее одного года. На следующем этапе редактирования будет произведена проверка на достоверность данных о количестве лет проживания в населенном пункте или административно-территориальной единице. Поскольку промежуток времени, в течение которого человек жил в данном населенном пункте или административно-территориальной единице, не может превышать возраст этого лица, его возраст будет соотнесен в этой ситуации с количеством лет, прожитых в данном населенном пункте или административно-территориальной единице.

ф) Определение продолжительности проживания данного лица по продолжительности проживания матери

482. Если данная категория не имеет достоверного кода, то программа может произвести общую проверку по записям в поисках данных о матери этого лица в данных о домохозяйстве. При положительном результате запись, касающаяся матери, может дать полезную информацию для указания недостающих величин. Если мать данного лица всегда проживала в этом населенном пункте или административно-территориальной единице и ее страной рождения является «эта страна» (как и должно быть), то программа отнесет количество лет, прожитых этим лицом в населенном пункте или административно-территориальной единице, к категории «всегда». Если страной рождения матери является не «эта страна», даже если в категории «количество лет, прожитых ею в данном населенном пункте или административно-территориальной единице» указано «всегда», это свидетельствует о том, что в категориях матери что-то указано неправильно. Тогда программа проигнорирует страну рождения матери и припишет возраст по продолжительности проживания в населенном пункте или административно-территориальной единице. Если в отношении количества лет, прожитых матерью в данном населенном пункте или административно-территориальной единице, указано не «всегда», но код присвоен правильно, а возраст данного лица меньше количества лет, которые мать прожила в данном населенном пункте или административно-территориальной единице, то программа редактирования вернется назад и проверит страну рождения матери. Если страна рождения матери указана как «эта страна», программа соотнесет возраст этого лица с количеством лет, прожитых в населенном пункте или административно-территориальной единице. Однако, если возраст этого лица равен или превышает количество лет, прожитых матерью в населенном пункте или административно-территориальной единице, программа соотнесет «количество лет, прожитых матерью в населенном пункте или административно-территориальной единице», с количеством лет, прожитых этим лицом в данном населенном пункте или административно-территориальной единице.

g) Определение продолжительности проживания того или иного лица по продолжительности проживания ребенка

483. Если рассматриваемое лицо является ребенком (сыном или дочерью), программа редактирования должна проверить запись о главе домохозяйства для возможного получения информации, которая может помочь присвоению значений для недостающих сведений о продолжительности проживания. Если глава домохозяйства родился в «этой стране» и всегда проживал в данном населенном пункте или административно-территориальной единице, программа укажет «всегда» для количества лет, прожитых ребенком в данном населенном пункте или административно-территориальной единице. Если глава домохозяйства всегда проживал в данном населенном пункте или административно-территориальной единице, но родился не в «этой стране», возраст ребенка будет соотнесен с количеством лет, прожитых им в населенном пункте или административно-территориальной единице. Если в графе «количество лет, проведенных в населенном пункте или административно-территориальной единице» для главы домохозяйства указано не «всегда», но при этом код является правильным, эту информацию можно использовать, если она согласуется с редактируемыми данными о возрасте ребенка. Если возраст ребенка равен или превышает количество лет, прожитых главой домохозяйства в данном населенном пункте или административно-территориальной единице, то программа использует данные о количестве лет, прожитых главой домохозяйства в населенном пункте или административно-территориальной единице, для указания количества лет, прожитых там сыном или дочерью. Если возраст ребенка меньше того периода времени, который глава домохозяйства проживал в населенном пункте или административно-территориальной единице, то программа присвоит значение в зависимости от страны рождения главы домохозяйства. Это значение будет «всегда», если глава домохозяйства родился в «этой стране»; в противном случае программа соотнесет возраст сына или дочери с количеством лет, прожитых в данном населенном пункте или административно-территориальной единице.

h) Продолжительность проживания лица, когда нет никакой другой информации

484. Когда все, о чем говорилось выше, не дает достоверного значения, программа может указать в графе «количество лет, прожитых в населенном пункте или административно-территориальной единице» в отношении этого лица «не сообщается» или «неизвестно». Если значение по-прежнему недостоверно, то значение «неизвестно» следует указывать, если не используется метод динамического вменения. Страны, применяющие метод динамического вменения при определении недостоверных величин, должны пользоваться по меньшей мере двумя характеристиками для получения «известной» информации, исходя из сведений об аналогичных лицах в данном географическом районе.

4. Прежнее место жительства

485. Под прежним местом жительства понимается крупная или более мелкая административно-территориальная единица либо другая страна, в которой данное лицо проживало непосредственно до приезда в нынешнюю тер-

риториально-административную единицу на постоянное жительство (United Nations, 2008, para 2.67).

a) Редактирование данных о прежнем месте жительства

486. Элемент данных «прежнее место жительства» следует редактировать вместе с элементом «продолжительность проживания». Если человек родился в данном месте (страна, населенный пункт или административно-территориальная единица, в зависимости от пункта переписи) и никогда оттуда не переезжал, в этом вопросе необходимо ставить или пробел или конкретный код «никогда не покидал». Однако пропуски могут вызвать проблемы при составлении таблиц, поэтому группе редактирования необходимо решить, как лучше справиться с такой ситуацией.

b) Прежнее место жительства при изменении границ

487. Границы стран со временем меняются, поэтому нужно следить за тем, чтобы это соответствующим образом отражалось в схемах кодирования. Кроме того, коды следует формировать так, чтобы можно было составлять логические группы. Например, как говорилось выше, первая цифра в трехзначном цифровом коде может представлять континент проживания, вторая цифра — регион на этом континенте, и третья цифра — страну в этом регионе.

c) Если человек никуда не переезжал с рождения

488. Обработчики данных составляют таблицы по определенным отдельным элементам. Таким образом, специалисты должны быть уверены в том, что кроме других кодов, присваиваемых месту жительства, используется специальный код «родился здесь». В этом случае программа должна различать тех, кто родился в данном месте, и тех, кто родился в каком-то одном месте, а затем переехал в другое место в том же географическом районе.

d) Использование данных о других лицах в единице

489. Если «прежнее место жительства» указано неправильно или не согласуется с другими элементами, то, как правило, применяется прием, аналогичный тому, который используется при редактировании элемента «продолжительность проживания». Программа редактирования может проверить прежнее место жительства матери, если она указана в жилищной единице. Затем она может проверить прежнее место жительства главы домашнего хозяйства, используя данные о детях и взрослых в тех странах, куда взрослые перемещаются нечасто.

e) Когда нет данных о других приемлемых лицах для редактирования прежнего места жительства

490. Если вышеуказанные процедуры не дают достоверного значения, программа может для этого лица указать «не сообщается», или «неизвестно», когда речь идет о количестве лет, прожитых на прежнем месте. Если значение остается неправильным, то следует приписать «неизвестно», когда не применяется метод динамического вменения. Страны, применяющие метод динамическо-

го вменения для определения недостоверных величин, должны пользоваться по меньшей мере двумя характеристиками для получения «известной» информации, исходя из сведений об аналогичных лицах в данном географическом районе.

5. Место жительства в определенный момент в прошлом

491. Под местом жительства в определенный момент в прошлом понимается крупная или более мелкая административно-территориальная единица, либо другая страна, в которой данное лицо проживало в определенный момент, предшествующий переписи. Выбранная базисная дата должна быть наиболее подходящей для национальных целей. В большинстве случаев предполагается, что это срок в один год или пять лет, предшествующий переписи. Базисная дата первого вида обеспечивает текущую статистику миграции в течение одного года, а базисная дата второго вида может быть более подходящей для сбора данных в целях проведения анализа международной миграции, хотя, вероятно, может быть менее пригодной для анализа текущей внутренней миграции. При выборе базисной даты необходимо также принимать во внимание вероятность наличия у отдельных лиц способности точно вспомнить свое постоянное место жительства, где они проживали за один год или за пять лет до даты переписи. В странах, проводящих перепись каждые пять лет, пятилетний срок для большинства лиц может быть легко увязан со временем предыдущей переписи. В других случаях человек может более точно вспомнить, где он проживал, за один год, чем за пять лет до проведения переписи.

492. Однако некоторые страны могут использовать другую базисную дату, отличную от периода в один или пять лет, предшествующих переписи, поскольку оба эти промежутка времени могут вызвать трудности, связанные с восстановлением в памяти необходимых данных. При определенных национальных условиях может возникнуть необходимость устанавливать базисную дату, которая будет ассоциироваться с важным событием, запомнившимся большинству населения. Кроме того, для изучения международной миграции может быть полезной информация относительно года прибытия в страну (United Nations, 2008, para. 2.69).

493. Элемент «место жительства в определенный момент в прошлом» редактируется аналогично данным о прежнем месте жительства. Как правило, страны спрашивают и «продолжительность проживания», и «прежнее место жительства» или просто «место жительства в определенный момент». Если человек родился в месте регистрации (страна, населенный пункт или административно-территориальная единица, в зависимости от пункта переписи) и никогда оттуда не переезжал, здесь нужно или оставить пробел, или поставить специальный код «никогда не покидал». Как указывалось ранее, пропуски могут вызвать проблемы при составлении таблиц. Затем применяются те же процедуры для прежнего места жительства, которые описаны в трех предыдущих пунктах.

6. Год прибытия

494. Второе пересмотренное издание *Принципов и рекомендаций в отношении переписей населения и жилого фонда* подразделяет переменные по миграции на внутреннюю миграцию и международную миграцию. Год прибытия обычно обозначает год прибытия в страну из какого-то места за пределами этой

страны. Таким образом, год прибытия — это элемент данных, по которому вопрос, как правило, задается вместе с дополняющим его вопросом, то есть место жительства до прибытия в данную страну (United Nations, 2008, para. 2.103).

a) Взаимосвязь между возрастом и годом прибытия

495. На первом этапе процесса редактирования следует произвести проверку на согласованность между возрастом и местом рождения и на достоверность сведений относительно года прибытия в данный населенный пункт или административно-территориальную единицу. Количество лет, прожитых тем или иным лицом в каком-либо населенном пункте или административно-территориальной единице, не может превышать возраст этого лица. Кроме того, человек, родившийся за пределами страны, не мог всегда проживать в данном населенном пункте или административно-территориальной единице. Программа должна приписывать ответ «всегда» году прибытия в населенный пункт или административно-территориальную единицу, если количество лет, прожитых в населенном пункте или административно-территориальной единице, превышает возраст, а страной рождения является эта страна. Если количество лет со времени прибытия в данный населенный пункт или административно-территориальную единицу, превышает возраст, а страной рождения является не эта страна, то одним из методов решения этой проблемы является соотнесение возраста данного лица с количеством лет, прожитых в населенном пункте или административно-территориальной единице. В этом случае исходят из того, что рассматриваемое лицо, хотя и родилось за пределами этой страны, переехало в данный населенный пункт или административно-территориальную единицу в возрасте менее одного года.

496. Для помощи пользователям открытых выборочных статистических данных статистические управления должны предоставить коды для ответа «менее одного» и «всегда» для этого элемента данных. Код «всегда» должен обычно присваиваться фактическому текущему постоянному месту жительства для содействия в прямом составлении таблиц. Код «менее одного» позволит пользователям быть уверенными в том, что они учли каждого человека в данной совокупности при составлении сводных таблиц.

b) Взаимосвязь между местом рождения и годом прибытия

497. В том случае, если вводимые данные выходят за установленные пределы, следует произвести такие же проверки, как и в отношении места рождения. Нужно найти предыдущих родственников (мать, главу домохозяйства, мужа, ребенка). Вменение должно производиться на основе найденной информации. Однако прежде чем приписать то или иное значение, нужно убедиться, что оно согласуется с возрастом и местом рождения лица, запись о котором редактируется.

c) Для лиц, которые всегда жили в данном месте

498. Если на вопрос о количестве лет со времени прибытия, прожитых тем или иным лицом в данном населенном пункте или административно-территориальной единице, указывается ответ «всегда жил здесь», но страна его рождения не «эта страна», то группа редактирования, возможно, захочет ис-

пользовать возраст данного лица для определения года его прибытия в данный населенный пункт или административно-территориальную единицу. Специалисты будут исходить из того, что, хотя рассматриваемое лицо и родилось за пределами этой страны, оно прибыло в данный населенный пункт или административно-территориальную единицу в возрасте менее одного года. На следующем этапе редактирования будет произведена проверка на достоверность данных относительно года прибытия в населенный пункт или административно-территориальную единицу. Поскольку промежуток времени, в течение которого то или иное лицо проживало в данном населенном пункте или административно-территориальной единице, не может превышать возраст этого лица, его возраст будет соотнесен с количеством лет, проведенных в данном населенном пункте или административно-территориальной единице.

d) Определение прибытия данного лица исходя из года прибытия матери

499. Если данная категория не имеет достоверного кода, то программа может произвести общую проверку по записям в поисках данных о матери этого лица в данных о домохозяйстве. При положительном результате запись, касающаяся матери, может дать полезную информацию для определения недостающих величин. Если мать данного лица всегда проживала в этом населенном пункте или административно-территориальной единице, и ее страной рождения является «эта страна» (как и должно быть), то программа в категории количества лет, прожитых этим лицом в населенном пункте или административно-территориальной единице, запишет «всегда». Если страной рождения матери является не «эта страна», даже если в категории количества лет, прожитых ею в данном населенном пункте или административно-территориальной единице, указано «всегда», это свидетельствует о наличии ошибки в категориях, касающихся сведений о матери. Тогда программа проигнорирует страну рождения матери и определит возраст по году прибытия в данный населенный пункт или административно-территориальную единицу. Если в отношении года прибытия матери в данный населенный пункт или административно-территориальную единицу указано не «всегда», но код присвоен правильно, а возраст данного лица меньше количества лет, прошедших с момента прибытия матери в данный населенный пункт или административно-территориальную единицу, то программа редактирования вернется назад и проверит страну рождения матери. Если страна рождения матери указана как «эта страна», программа соотнесет возраст этого лица с количеством лет, прожитых в населенном пункте или административно-территориальной единице. Однако, если возраст того или иного лица равен или превышает количество лет, прошедших с момента прибытия матери в населенный пункт или административно-территориальную единицу, программа соотнесет категорию «год прибытия матери в населенный пункт или административно-территориальную единицу» с годом прибытия данного лица в населенный пункт или административно-территориальную единицу.

e) Определение года прибытия ребенка исходя из года прибытия главы домохозяйства

500. Если рассматриваемое лицо является ребенком (сыном или дочерью), программа редактирования должна проверить запись о главе домохозяйства для возможного получения информации, которая может помочь

присвоению значений для недостающих сведений относительно года прибытия. Если глава домохозяйства родился в «этой стране» и всегда проживал в данном населенном пункте или административно-территориальной единице, программа укажет ответ «всегда» для количества лет, прожитых ребенком в данном населенном пункте или административно-территориальной единице. Если глава домохозяйства всегда проживал в данном населенном пункте или административно-территориальной единице, но родился не в «этой стране», возраст ребенка будет соотнесен с количеством лет, прожитых им в населенном пункте или административно-территориальной единице. Если в графе года прибытия в данный населенный пункт или административно-территориальную единицу для главы домохозяйства указано не «всегда», но при этом код остается правильным, эту информацию можно использовать, если она согласуется с возрастом в редактируемых данных ребенка. Если возраст ребенка равен или превышает возраст, определяемый годом прибытия главы домохозяйства в данный населенный пункт или административно-территориальную единицу, то программа использует год прибытия главы домохозяйства в населенный пункт или административно-территориальную единицу в качестве года прибытия этого ребенка в данный населенный пункт или административно-территориальную единицу. Если возраст ребенка меньше возраста, определяемого годом прибытия главы домохозяйства в населенный пункт или административно-территориальную единицу, то программа определит значение в зависимости от страны рождения главы домохозяйства. Это значение будет «всегда», если глава домохозяйства родился в «этой стране»; в противном случае программа соотнесет возраст сына или дочери с количеством лет, прожитых в данном населенном пункте или административно-территориальной единице.

f) Определение года прибытия данного лица при отсутствии какой-либо другой информации

501. Когда все, о чем говорилось выше, не дает достоверного значения, программа может указать «не сообщается» или «неизвестно» в отношении года прибытия рассматриваемого лица в данный населенный пункт или административно-территориальную единицу». Если значение по-прежнему недостоверно, то значение «неизвестно» следует указывать, если не используется метод динамического вменения. Страны, применяющие метод динамического вменения при определении недостоверных величин, должны пользоваться надлежащим числом характеристик для получения «известной» информации, исходя из сведений об аналогичных лицах в данном географическом районе.

7. Взаимосвязь между продолжительностью проживания и годом прибытия

502. Важно отметить, что некоторые страны делают основной упор на внутреннюю миграцию и включают графу о продолжительности проживания (часто с указанием предыдущего места жительства). Другие страны, делающие акцент на международную миграцию, включают пункт, касающийся года прибытия (часто с указанием места жительства до переезда). В большинстве стран наблюдается либо значительная внутренняя миграция и небольшая международная миграция или значительная иммиграция и небольшая внутренняя миграция. В некоторых странах, однако, может наблюдаться как внутренняя,

так и международная миграция, и, таким образом, им необходимо включать в переписные листы оба элемента данных.

503. При включении в переписной лист обоих элементов сотрудники статистического управления должны проявить особую внимательность при разработке схем редактирования, чтобы не получить в итоге внутренне несогласованные данные. Иными словами, переменные величины для возраста, продолжительности проживания и года прибытия должны рассматриваться совместно для обеспечения того, чтобы суммарное количество лет, соответствующее продолжительности проживания и времени, прошедшему с момента прибытия, не превышало показателей возраста. Следовательно, программистам необходимо рассматривать все три переменные одновременно, когда это имеет место.

504. При использовании динамического вменения сотрудники статистических учреждений должны использовать данные «hot deck», включающие многомерные массивы данных для учета различных возрастных групп и сроков в годах. Кроме того, если продолжительность проживания и год прибытия представляют собой отдельные годы, в данных «hot deck» также должны использоваться сроки в разбивке по одному году, так как обновление данных, например по какому-либо пятилетнему периоду, может привести к конфликту в процессе вменения.

505. Наряду с этим следует проявлять особую осторожность, когда в процессе такой проверки собираются сгруппированные данные по продолжительности проживания или году прибытия или по обоим этим элементам, а также в ходе разработки и вменения данных «hot decks». Сгруппированные данные вызывают проблемы взаимного наложения данных. Страны могут принять решение о том, что оптимальным подходом в такой ситуации может быть запись «неизвестно».

8. Постоянное место жительства

506. Как правило, «постоянное место жительства» определяется в целях переписи как место, где данное лицо проживает на момент проведения переписи и где он или она находились в течение некоторого времени или намерены остаться на некоторое время (United Nations, 2008, paras. 1.461–1.463). Странам при определении постоянного места жительства рекомендуется применять пороговый период в 12 месяцев в соответствии с одним из двух следующих определений:

- a) место, в котором данное лицо, проживало постоянно в течение большей части последних 12 месяцев (то есть по меньшей мере шесть месяцев и один день), не считая временного отсутствия по причине отпусков и командировок, или в котором это лицо намерено прожить не менее шести месяцев;
- b) место, в котором данное лицо, проживало постоянно в течение по меньшей мере последних 12 месяцев, не считая временного отсутствия по причине отпусков и командировок, или в котором это лицо намерено прожить не менее 12 месяцев.

507. При этом страны, проводящие переписи с регистрацией де-факто, могут включать дополнительный элемент данных в графу «постоянное место жительства» для получения информации как де-юре, так и де-факто. Редактирование этого элемента варьируется в зависимости от конкретных особенностей страны. Для никогда не переезжавших лиц постоянное место жительства

совпадает с местом текущего проживания; следовательно, отсутствующую информацию можно заполнять напрямую.

508. Тем не менее, когда данные показывают движение населения, ситуация усложняется. Обычно, если для этого элемента ставится пропуск, страны исходят из того, что постоянное место жительства и текущее место проживания совпадают друг с другом или что счетчик и/или респондент просто пропустили эту информацию.

509. Однако, когда данные о продолжительности проживания или о годе прибытия дают те или иные доказательства изменения места жительства, тогда сотрудники статистических учреждений возможно захотят сделать попытку разработать методы получения максимально достоверных оценок по конкретным географическим районам или по всей стране в целом. Хотя конкретные схемы редактирования будут зависеть от особенностей отдельной страны, в качестве последнего средства необходимо, по всей видимости, использовать категорию «неизвестно».

510. Если счетчик получает указание оставить пропуск в случае совпадения постоянного места жительства и места регистрации, в ходе редактирования в графе «постоянное место жительства» необходимо указывать код места регистрации. Другая переменная должна указывать на то, что редактор внес такое изменение. Наличие полного набора кодов поможет пользователям данных открытой выборки в составлении собственных полных таблиц данных.

С. Социальные характеристики

511. Социальные характеристики в разных странах отличаются друг от друга, но, как правило, они включают в себя элементы данных, характеризующие различные аспекты социокультурных условий в стране. Элементы данных, касающиеся образования, включая грамотность, посещение учебных заведений и образовательную подготовку, а также область знаний и образовательный ценз можно классифицировать в соответствии с категориями, приведенными в пересмотренном издании Международной стандартной классификации образования (МСКО) 1997 года, выпущенном Организацией Объединенных Наций по вопросам образования, науки и культуры (ЮНЕСКО) (United Nations, 2008, paras. 2.202–2.230).

1. Способность читать и писать (грамотность)

512. Сбор данных о грамотности должен осуществляться в отношении всех лиц в возрасте 10 лет и старше. Однако в ряде стран некоторые лица в возрасте от 10 до 14 лет могут вскоре стать грамотными благодаря учебе в школе, и процент грамотного населения для данной возрастной группы может не соответствовать действительности и вводить в заблуждение. Поэтому при проведении международных сравнительных исследований грамотности данные о ней следует собирать в отношении всех лиц в возрасте от 15 лет и старше. В тех случаях, когда страны собирают сведения в отношении лиц более молодого возраста, при составлении таблиц с данными о грамотности следует по крайней мере проводить различие между лицами младше 15 лет и лицами в возрасте 15 лет и старше (United Nations, 2008, para. 2.202).

513. Каждая страна должна установить минимальный возраст для составления таблиц с данными о грамотности; при этом группы редактирования должны решить, какой минимальный возраст нужно выбрать для редактирования данных о грамотности, поскольку для внутреннего пользования могут понадобиться дополнительные таблицы. По мере составления переписного листа группы редактирования должны принять решение относительно минимального возраста с точки зрения сбора данных и с какого образовательного уровня этот вопрос уже не стоит задавать. В связи с этим, если респондент уже достиг определенного уровня образования, счетчику, видимо, уже нет необходимости задавать ему вопрос о грамотности. Тем не менее эта графа должна заполняться во время редактирования с целью помощи исследователям и другим лицам, использующим открытые данные.

514. При редактировании данных о грамотности прежде всего проверяется последняя законченная ступень обучения; если в соответствующей графе стоит ответ «грамотный» исходя из спецификаций, то следует присвоить код для ответа «да». Лица на определенном уровне обучения должны считаться грамотными. В тех случаях, когда обнаруживается неправильный код для грамотности, следует указать то или иное значение. При этом должна стоять запись или «не указано», или значение следует определить, используя матрицу вменения, исходя из конкретных переменных, таких как последняя ступень обучения и пол. Элемент «последняя законченная ступень обучения» зависит от принятого в каждой конкретной стране определения понятия «грамотный».

2. Посещение учебных заведений

515. В принципе информация о посещении учебных заведений должна собираться для лиц всех возрастов. Это относится, в частности, к группам населения официально установленного возраста школьного обучения, который составляет, как правило, от 5 до 29 лет, но может варьироваться в зависимости от страны и особенностей ее национальной системы образования. В случае более широкого диапазона собираемых данных, охватывающего сведения о посещении дошкольных учреждений и/или участия в других систематизированных образовательных и учебных программах, организуемых для взрослых на производственных предприятиях и предприятиях сферы услуг, в общинных организациях и других неучебных учреждениях, возрастной диапазон может устанавливаться в зависимости от конкретных обстоятельств (United Nations, 2008, para. 2.209).

а) Редактирование данных о посещении учебных заведений

516. В каждой стране группа редактирования должна решить, по каким возрастным группам собирать данные о посещении учебных заведений. Поскольку в большинстве стран обучение подразделяется еще и на несколько уровней, то для определения этих уровней по возрасту специалисты должны также решить, какие возрастные группы соответствуют различным уровням обучения. При этом записи по всем остальным лицам необходимо изменить. Если программа редактирования дает несогласующиеся ответы для данной категории, необходимо изменить или возраст, или сведения о посещении учебных заведений. Как правило, возраст устанавливается ко времени редактирования, поэтому изменения касаются именно данных о посещении учебных заведений. Счетчикам следует дать указание опускать данные о посещении учебных за-

ведений для лиц старше заранее установленного возраста, если это подходит для данной страны. В тех случаях, когда лица среднего возраста продолжают получать среднее или высшее образование, то может быть и нецелесообразно устанавливать верхний возрастной предел для данных о посещении учебных заведений. Предполагается, что ответы и комбинации ответов проверяются еще до переписи на предварительном этапе, поэтому такие решения можно принимать до фактического проведения переписи.

b) Прием на очное или заочное обучение

517. Некоторые страны могут по своему усмотрению собирать информацию об очном или заочном обучении. Если такой элемент включается в перепись, его можно редактировать вместе с данными о посещении учебных заведений или в качестве отдельного пункта.

c) Согласованность между сведениями о посещении учебных заведений и экономической деятельностью

518. При редактировании в первую очередь следует проверять данные о согласованности с другими важными элементами данных, например с основной экономической деятельностью. Если в данных об основной экономической деятельности указано посещение учебного заведения и респондент ответил, что его или ее основной деятельностью является посещение школы, то в графе «посещение учебного заведения» нужно присвоить код для ответа «да», а основную экономическую деятельность указать как «учащийся», то есть ответы должны быть согласованы между собой. Во всех других случаях должны приниматься любые достоверные ответы.

d) Присвоение той или иной категории недостоверным или несогласующимся записям о посещении учебных заведений

519. Если указанные данные выходят за установленные пределы, а запись, касающаяся последней законченной ступени обучения, достоверна, то значение следует присваивать, используя матрицу вменения по возрасту, полу и последней ступени обучения. Если последняя ступень имеет неправильный код, то для указания сведений о посещении учебных заведений следует использовать запись в графе «грамотность». Если для указания грамотности дан неправильный код, то запись о посещении учебных заведений следует делать только на основе возраста и пола.

520. Матрицы вменения могут понадобиться для того, чтобы показать различные характеристики посещения учебных заведений в разбивке по полу и возрасту (иногда по одногодичному возрастному интервалу или по небольшим возрастным группам).

3. Образовательная подготовка (последняя завершенная ступень или уровень)

a) Редактирование данных об образовательной подготовке

521. При редактировании данных по образовательной подготовке (последняя ступень или уровень) следует: а) проверить на согласованность досто-

верную запись и возраст и *b*) произвести вменение записи, если первоначальная запись выходит за установленные пределы. Как указывалось ранее, в странах, которые не используют метод динамического вменения, должна быть запись «не указано». В странах, где используется метод динамического вменения, указание пола и годового возрастного интервала потребуется для молодых людей, а указание пола и небольших возрастных групп потребуется для детей старшего возраста. Для стран, где указываются сведения о последней ступени или последнем уровне образования, могут понадобиться многомерные матрицы вменения (United Nations, 2008, para. 2.215). Следует обратиться к приложению I по предложениям касательно перекодирования графы «текущая ступень образования», исходя из данных посещения учебных заведений и самой высокой завершенной ступени образования.

b) Минимальный возраст для образовательной подготовки

522. В каждой стране группы редактирования должны определить минимальный возраст поступления в школу. При установлении минимального возраста высший завершенный уровень, как правило, не должен превышать возраст данного лица плюс некоторая постоянная величина (представляющая этот минимальный возраст поступления в школу). Кроме того, необходимо использовать годовые возрастные интервалы для детей, поскольку обновление матриц вменения может привести к ошибкам в том случае, если возрастные группы весьма широкие по своему диапазону.

c) Взаимосвязь между возрастом и образовательной подготовкой

523. Группа редактирования должна также определиться с тем, сколько несоответствий допустимо в наборе данных. Как правило, лучше внести изменения в нескольких исключительных случаях, когда не согласуются возраст и образовательная подготовка, чем принимать большое количество ответов, которые действительно несовместимы. Исходя из этого, в тех случаях, когда исходная запись выходит за установленные пределы или не согласуется с возрастом, ту или иную запись можно приписать. Для стран, где не используется метод динамического вменения, можно ввести категорию «не указано». Там, где этот метод применяется, запись можно получить исходя из возраста (включая годовые возрастные интервалы для лиц школьного возраста), пола и процента посещаемости учебных заведений. ЮНЕСКО признает грамотность как категорию, отдельную от образовательной подготовки, и поэтому «способность читать и писать», вероятно, не следует использовать в качестве оценочного показателя в матрице вменения.

4. Область знаний и образовательный ценз

524. Информация о лицах, собираемая по признаку уровня образования и области знаний, важна для анализа соответствия между предложением и спросом на квалифицированную рабочую силу с учетом конкретной специализации на рынке труда. Это одинаково важно для планирования и регулирования объема выпуска учащихся учебных заведений и программ профессиональной подготовки различных уровней, типов и профиля (United Nations, 2008, para. 2.223).

525. Лица моложе 15 лет (или другого заранее определенного возраста) не должны указывать информацию об области знаний и/или образовательных цен-

зах. Для лиц в возрасте 15 лет и старше должна быть взаимосвязь между уровнем образовательной подготовки и областью знаний и/или образовательными цензами. Каждый раз, когда возникают недостоверные записи, страны, не использующие метод динамического вменения, могут ввести запись «неизвестно». Страны, применяющие метод динамического вменения, могут рассмотреть вопрос об использовании возраста, пола, образовательной подготовки и по возможности рода занятий для определения области знаний и/или образовательного ценза.

5. Вероисповедание

526. В целях переписи понятие вероисповедания может быть определено как *a)* религиозная или духовная вера, которой отдается предпочтение независимо от того, представлена или нет эта вера организованной группой, или *b)* принадлежность к организованной группе, придерживающейся определенной религиозной или духовной доктрины. Каждая страна, которая при проведении переписи учитывает вероисповедание, должна пользоваться тем определением, которое лучше всего соответствует ее собственным нуждам, и должна указывать в опубликованных результатах переписи применяемое определение (United Nations, 2008, para. 2.152).

a) Редактирование данных о вероисповедании

527. Вероисповедание — одна из переменных, соответствующих примерам стандартного редактирования, приведенным в главе II. В отношении элементов данных по вероисповеданию, в отличие от других элементов, наблюдается высокая доля «отсутствия ответа», и этот фактор необходимо учитывать, поскольку некоторые люди могут быть не склонны сообщать о своей религиозной принадлежности. Достоверное значение (включая «отсутствие ответа») получается для каждого лица или непосредственно от другого члена домохозяйства, если такое значение имеется, или от другого главы домохозяйства с аналогичными характеристиками. Группа редактирования должна составить план логического редактирования для использования в отношении других социальных переменных. Сначала следует определить главу домохозяйства и отредактировать данные о нем независимо от того, является ли он или она первым лицом в жилищной единице. Если лицо с неправильно указанным или неизвестным вероисповеданием является главой домохозяйства, то нужно предпринять следующие шаги:

b) Отсутствие данных о вероисповедании главы домохозяйства, но наличие данных о вероисповедании какого-либо другого лица в жилищной единице

528. В качестве первого шага нужно определить, есть ли достоверные данные о вероисповедании кого-либо другого в жилищной единице, и приписать первый достоверный ответ о вероисповедании.

c) Отсутствие данных о вероисповедании главы домохозяйства или любого другого лица в жилищной единице

529. Если вероисповедание кого-либо в домохозяйстве не указывается, следует или приписать категорию «неизвестно» (если данная страна не использует метод динамического вменения), или вменить вероисповедание по данным

последнего главы домохозяйства с аналогичными характеристиками, включая возраст и пол, а также язык, место рождения и другие переменные с учетом соответствующих обстоятельств.

d) Отсутствие данных о вероисповедании лица, не являющегося главой домохозяйства

530. Если данное лицо не является главой домохозяйства и не указывает свое вероисповедание, группа редактирования может принять решение приписать вероисповедание главы домохозяйства.

6. Язык

531. Имеются три категории данных о языке, которые могут быть собраны при проведении переписи (United Nations, 2008, para. 2.156), а именно следующие:

- родной язык, определяемый как язык, на котором обычно говорили в доме опрашиваемого лица в его раннем детстве;
- обычно употребляемый язык, определяемый как язык, на котором в настоящее время говорит или на котором наиболее часто говорит опрашиваемое лицо в его нынешнем доме;
- умение говорить на одном или нескольких обозначенных языках.

a) Редактирование данных о языке

532. Из трех разных категорий языка, которые могут быть указаны в переписном листе, первые две — родной язык и обычно употребляемый язык — взаимосвязаны. Если в переписном листе указаны обе категории, группы редактирования должны рассмотреть возможность их совместного редактирования. Если одна из них указана неправильно, другую можно использовать для восполнения записи.

b) Редактирование данных о языке: глава домохозяйства

533. Язык — это еще одна переменная, соответствующая примерам, представленным в главе II. Группы редактирования должны составить план логического редактирования для использования в отношении других социальных переменных, редактируя в первую очередь данные о главе домохозяйства. Если лицо, чей язык указан неправильно или неизвестен (родной язык или обычно употребляемый язык), является главой домохозяйства, в первую очередь необходимо определить, есть ли кто-либо еще в данной жилищной единице, чей язык указан правильно, и приписать первому достоверный язык. Если таких лиц нет, нужно приписать категорию «неизвестно», если не применяется метод динамического вменения, или вменить язык по данным последнего главы домохозяйства с аналогичными характеристиками, включая возраст и пол, а также иные языковые переменные, место рождения и другие переменные в зависимости от обстоятельств.

c) Редактирование данных о языке: лица, отличные от главы домохозяйства

534. Если данное лицо не является главой домохозяйства, а язык указан неправильно, то ему следует приписать язык главы домохозяйства.

d) Редактирование данных о языке: использование данных об этническом происхождении или месте рождения

535. Язык и этническое происхождение, а иногда и место рождения тесно взаимосвязаны, и в некоторых странах эти данные можно редактировать совместно. Кроме того, группы редактирования должны рассмотреть возможность присвоения кодов, чтобы отразить соотношение между этими переменными. В зависимости от количества знаков в коде и распределения языков и этнических групп в стране можно установить соответствия, с тем чтобы легче было восполнить неизвестные или несогласующиеся ответы.

e) Редактирование данных о языке: родной язык

536. Если родной язык неизвестен, но данное лицо является филиппинцем или родилось на Филиппинах, то можно приписать соответствующий эквивалентный язык: тагальский, илоканский или какой-либо другой язык, распространенный на Филиппинах. Язык присваивается таким способом только главе домохозяйства, а код этого языка присваивается другим членам домохозяйства, но в каждой стране группа редактирования должна рассматривать конкретные обстоятельства, в том числе географические данные (например, проживание в городских или сельских районах), возраст или другие элементы.

f) Редактирование данных о языке: умение говорить на обозначенном языке

537. Умение говорить на обозначенном языке — это третья переменная, соответствующая примерам, представленным в главе II. Кроме того, в первую очередь следует редактировать данные о главе домохозяйства. Если значение для языка главы домохозяйства недостоверно или неизвестно, сначала надо посмотреть, есть ли кто-либо еще в данной жилищной единице с достоверной способностью говорить на том или ином языке, и приписать первый достоверный язык. После этого, если таких лиц в жилищной единице нет, следует или приписать «неизвестно», если в данной стране не используется метод динамического вменения, или исчислить язык по данным последнего главы домохозяйства с аналогичными характеристиками (например, возраст и пол, а также место рождения и другие переменные в зависимости от обстоятельств). Если данное лицо не является главой домохозяйства, а способность говорить на обозначенном языке недостоверна, то следует приписать эту способность главе домохозяйства.

7. Этническая принадлежность и коренные народы

538. Решение о сборе и распространении информации об этнических или национальных группах населения в ходе переписи зависит от целого ряда соображений и национальных особенностей, включая, например, национальные потребности в таких данных, приемлемость и деликатность постановки вопросов об этнической принадлежности в рамках проводимой в той или иной стране переписи. Выявление этнокультурных характеристик населения страны имеет все большее значение в контексте миграции, интеграции и политики в отношении групп меньшинств (United Nations, 2008, para. 2.160).

539. Во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда* (пункт 2.163) рекомендуется

проявлять особую внимательность при выявлении коренных народов, обычно в качестве подгруппы элемента данных об этнической принадлежности. Следует также позаботиться о разработке перечней кодов для обеспечения однозначной идентификации принадлежности к коренной народности в целях разработки надлежащих схем редактирования и таблиц для помощи в планировании и формировании политики в отношении коренных народов. Например, могут потребоваться отдельные коды для одной и той же группы населения, если она ведет кочевой образ жизни по сравнению с той же группой, которая поселилась для проживания в жилых районах. Можно разработать специальные схемы редактирования, особенно с помощью «поисковых» файлов для конкретных групп коренных народов в целях обеспечения их надлежащей и полной идентификации для последующего включения в таблицы. Для таких групп можно также разработать особые процедуры вменения или использовать дополнительные категории в рамках имеющихся данных «hot decks».

a) Редактирование данных об этнической принадлежности

540. Некоторые другие переменные, если они указываются в ходе переписи, могут быть полезными при «определении» этнической принадлежности, когда она указана неправильно или неизвестна. Во многих странах существует взаимосвязь между местом рождения — как в данной стране, так и в других странах — и этнической принадлежностью. Графа «родной язык» также зачастую является хорошим показателем этнической принадлежности во многих странах, поскольку категории, а значит и коды, будут аналогичными или вообще одинаковыми.

b) Редактирование данных об этнической принадлежности: глава домохозяйства

541. Этническое происхождение также соотносится с примерами, приведенными в главе II. Группы редактирования должны следовать той схеме, которая уже приводилась для других социальных переменных. Данные о главе домохозяйства должны редактироваться в первую очередь. Если лицо, данные об этническом происхождении которого недостоверны или неизвестны, является главой домохозяйства, следует сначала найти достоверные данные об этническом происхождении какого-либо другого лица в жилищной единице и приписать первый достоверный ответ об этнической принадлежности. Если такого лица нет, то тогда следует приписать «неизвестно» или, если данная страна не применяет метод динамического вменения, исчислить этническую принадлежность по данным последнего главы домохозяйства с аналогичными характеристиками (возраст и пол, а также язык, место рождения и другие переменные, которые могут быть подходящими в зависимости от обстоятельств).

c) Редактирование данных об этнической принадлежности: лица, не являющиеся главой домохозяйства

542. Если данное лицо не является главой домохозяйства, а этническое происхождение указано неверно, то следует приписать этническое происхождение главы домохозяйства.

d) Редактирование данных об этнической принадлежности: использование сведений о языке и месте рождения

543. Этническое происхождение и язык, а иногда и место рождения тесно взаимосвязаны, и в отношении некоторых стран эти сведения можно редактировать совместно. Кроме того, группы редактирования должны продумать вопрос о присвоении кодов, с тем чтобы отразить взаимосвязь этих переменных. В зависимости от количества знаков в коде и распределения этнических групп и языков в стране можно построить соответствия, которые помогут восполнить неизвестные или несогласующиеся ответы.

544. Например, если этническое происхождение неизвестно, но данное лицо говорит на одном из языков, распространенных на Филиппинах, и родилось на Филиппинах, можно присвоить соответствующее эквивалентное этническое происхождение, то есть филиппинское. Как правило, таким способом присваивается этническое происхождение только главы домохозяйства (а другим членам присваивается этот код), но в каждой стране группа редактирования должна учитывать конкретные обстоятельства, в том числе географические данные (например, проживание в городских и сельских районах), возраст или другие элементы.

8. Нетрудоспособность

545. Статус нетрудоспособности позволяет разделить население на нетрудоспособных и трудоспособных лиц. То или иное лицо с нетрудоспособностью определяется как лицо, которое подвергается повышенному риску, по сравнению с остальным населением, столкнуться с ограничениями при выполнении конкретных задач или участия в ролевой деятельности. Организация Объединенных Наций рекомендует для определения статуса нетрудоспособности использовать четыре области функционирования: 1) ходьба; 2) зрение; 3) слух; и 4) когнитивная способность (United Nations, 2008, paras. 2.351–2.352 and 2.367–2.371).

546. Вопрос, задаваемый с целью выявления лиц с нетрудоспособностью, должен включать широкие категории нетрудоспособности, с тем чтобы каждое лицо могло определить для себя наличие и отсутствие каждого вида нетрудоспособности. Для контроля данных о нетрудоспособности можно использовать следующие категории на основе международной классификации расстройств здоровья, нетрудоспособности и инвалидности (МКРНИ): 1) функционирование и нетрудоспособность, включая функции организма и структуру организма (расстройства), деятельность (ограничения) и участие (ограничения); и 2) ситуативные факторы, включая факторы окружающей среды и личные факторы (United Nations, 2008, para. 2.354).

⁷ При комбинировании различных областей функционирования, например постановки вопроса о зрении или слухе, респонденты зачастую приходят в замешательство, полагая, что для предоставления утвердительного ответа они должны испытывать трудности в обеих этих сферах. Кроме того, получение данных о числе лиц с конкретными ограничениями является полезным для целей как внутреннего планирования, так и международной сопоставимости.

a) Вопросы переписи, касающиеся нетрудоспособности

547. Рекомендуется уделять особое внимание разработке вопросов переписи для измерения показателей нетрудоспособности. Формулировки и структура вопросов оказывают значительное влияние на точность выявления лиц с нетрудоспособностью. Каждая область функционирования требует отдельного вопроса⁷. Используемые формулировки должны быть простыми, ясными и недвусмысленными. В любом случае необходимо избегать негативных терми-

нов. Важно адресовать вопросы об инвалидности каждому отдельному члену домохозяйства и избегать общих вопросов о наличии лиц с нетрудоспособностью в домохозяйстве. В случае необходимости ответы за того или иного нетрудоспособного члена семьи может представить доверенный респондент. Весьма важно то, чтобы каждый член домохозяйства учитывался в индивидуальном порядке, а не с помощью сводного вопроса. Предоставлению данных о нетрудоспособности может содействовать использование нескольких дифференцированных категорий ответов (United Nations, 2008, para. 2.373).

b) Редактирование данных о нетрудоспособности

548. Если кто-то не отвечает на задаваемые вопросы, связанные с нетрудоспособностью, то в этом случае трудно определить, обусловлен ли пропуск данного пункта отсутствием нетрудоспособности или нежеланием данного респондента отвечать на этот вопрос по какой-то другой причине. Группа редактирования в стране должна решить, следует ли редактировать этот элемент данных обычным порядком, приписывая «неизвестно», если не используется метод динамического вменения, или используя ответы других лиц, если применяется динамическое вменение. В качестве альтернативы специалисты могут решить, что следует принимать только те ответы, где указывается наличие нетрудоспособности, и что любой недостоверный ответ следует записывать как «нетрудоспособность отсутствует». В последнем случае динамическое вменение не применяется.

c) Многие виды нетрудоспособности

549. Страны, собирающие информацию о разных видах нетрудоспособности, должны видоизменить приемы редактирования. Программа редактирования должна следить за тем, сколько всего может быть видов нетрудоспособности и за дублированием и распределением этих видов нетрудоспособности. Как и прежде, большинство стран сочтут нецелесообразным использовать данные, построенные на сведениях о нетрудоспособности других лиц, поэтому в случаях недостоверности, возможно, придется использовать такие записи, как «неизвестно» или даже «неизвестно, имеется ли нетрудоспособность».

d) Редактирование данных о причинах нетрудоспособности

550. Группа редактирования в той или иной стране должна решить, следует ли редактировать этот элемент данных обычным порядком, присваивая ответ «неизвестно», если не применяется метод динамического вменения, или используя ответы других лиц, если применяется динамическое вменение. В качестве альтернативы специалисты могут решить, что будут приниматься только те ответы, где указывается причина нетрудоспособности, и в этом случае матрица вменения использоваться не будет.

D. Экономические характеристики

551. Информацию о статусе экономической активности следует в принципе собирать в отношении всего населения, но на практике такая ин-

формация собирается по каждому лицу, достигшему или превышающему минимальный возраст, установленный в каждой стране в соответствии с ее условиями. Минимальный возраст окончания школы не должен автоматически приниматься за нижний возрастной предел для сбора информации о статусе экономической активности. В странах, где обычно многие дети участвуют в сельскохозяйственных работах или других видах экономической деятельности (например, разработка полезных ископаемых, ткачество и мелкая торговля), потребуется выбрать более низкий минимальный возраст, чем в странах, для которых подобная занятость детей нехарактерна.

552. В таблицах данных об экономических характеристиках необходимо как минимум проводить различие между лицами в возрасте до 15 лет и лицами в возрасте 15 лет и старше; страны, где минимальный возраст окончания школы превышает 15 лет и где есть экономически активные дети младше этого возраста, должны стремиться получить данные об экономических характеристиках таких детей в целях достижения международной сопоставимости данных, по крайней мере для лиц в возрасте 15 лет и старше. Участие в экономической деятельности пожилых мужчин и женщин после достижения ими обычного пенсионного возраста также зачастую недооценивается. Это требует пристального внимания при определении экономически активного населения. Максимальный возрастной предел для измерения экономически активного населения, как правило, не используется, поскольку значительное число пожилых людей старше пенсионного возраста могут регулярно или время от времени принимать участие в экономической деятельности (United Nations, 2008, para. 2.241).

553. Каждая страна должна определить минимальный возраст для участия в экономической деятельности. Странам, которые заинтересованы в сборе данных о детском труде, возможно, потребуется установить низкий минимальный возраст, но при этом необходимо помнить о том, что могут возникнуть некоторые искажения в статистике, если дети, которые не входят в число работающих, будут ошибочно зарегистрированы как участвующие в составе рабочей силы. После установления минимального возраста производится редактирование элементов экономической активности, с тем чтобы включить в таблицы данные о лицах в возрасте X лет или старше; поэтому редактирование данных о детях в возрасте до X лет понадобится только для того, чтобы убедиться, что все соответствующие графы не заполнены. Чтобы упростить составление всех таблиц, следует исключить любые ответы, которые могли быть внесены для детей в возрасте до X лет.

1. Статус экономической активности

554. Статус экономической активности составляется из нескольких экономических переменных, часть из которых описывается ниже. Этих переменных может быть достаточно для сбора данных, но в целях обработки и анализа данных возможно потребуется изменить их классификацию.

555. «Статус текущей экономической активности» — это отношение того или иного лица к экономической деятельности, основанное на коротком отчетном периоде, таком как одна неделя. Использование показателя текущей экономической активности считается наиболее подходящим для стран, где на экономическую активность людей не оказывают большого влияния сезонные

или другие факторы, вызывающие колебания в течение года. Этим отчетным периодом в одну неделю может быть любая установленная фиксированная неделя в недавнем прошлом, полная последняя календарная неделя или последние семь дней до переписи (United Nations, 2008, para. 2.248).

556. Согласно документам Организации Объединенных Наций (United Nations, 2008, para. 2.253), занятое население включает всех лиц старше определенного возраста, которые на протяжении короткого отчетного периода либо в одну неделю, либо в один день *a*) выполняли определенную работу за плату или с целью получения прибыли или пополнения семейного дохода деньгами или натурой; или *b*) временно отсутствовали на работе, на которой они уже работали и с которой они сохраняли официальную связь, или на работе на основе самозанятости, такой как коммерческое предприятие, ферма или предприятие сферы услуг.

557. *Экономически неактивное на данный момент население*, или, что то же самое, лица, не входящие в состав *рабочей силы*, включает всех лиц, которые и не были заняты, и не были безработными в течение короткого отчетного периода, используемого для оценки текущей экономической активности, в том числе лиц, не достигших минимального возраста, установленного для измерения численности экономически активного населения (United Nations, 2008, paras. 2.278–2.279).

a) Категории, относящиеся к статусу экономической активности

i) Незанятое население

558. Согласно документам Организации Объединенных Наций (United Nations, 2008, para. 2.271), незанятое население включает всех лиц старше определенного возраста, которые на протяжении отчетного периода отвечали следующим критериям:

- a*) не имели работы: не были заняты в сфере оплачиваемого труда или не были самостоятельно заняты;
- b*) на данный момент готовы к работе: были готовы к оплачиваемому наемному труду или к самостоятельной занятости в течение отчетного периода;
- c*) искали работу: предпринимали конкретные шаги в определенный недавний период в поисках оплачиваемой работы по найму или самостоятельной работы. Конкретные шаги могли включать регистрацию на государственной или частной бирже труда; заявления к работодателям; регистрацию на рабочих площадках, фермах, у заводских ворот, на рынках или других людных местах; помещение объявления в газету или ответы на объявления; просьбу о содействии друзей и родственников; поиски земельного участка, здания или оборудования для создания собственного предприятия; поиски финансовых средств; и обращение за разрешением или лицензией. При классификации безработного было бы полезно проводить различие между лицами, ищущими работу впервые, и прочими ищущими работу лицами.

559. Как правило, для классификации в качестве безработного то или иное лицо должно удовлетворять всем трем вышеназванным критери-

ям. Однако в ситуациях, когда традиционные методы поиска работы имеют ограниченное значение, когда рынок рабочей силы в значительной мере не организован или ограничен, когда включение в экономическую жизнь иммигрантов на данный момент недостаточно или когда рабочую силу в большинстве своем составляют самостоятельно занятые лица, стандартное определение незанятости может применяться при менее строгом соблюдении критерия «поиск работы». Такое смягчение критерия поиска работы предназначено в первую очередь для тех развивающихся стран, где данный критерий не применим ко всем случаям незанятости. Несмотря на это смягчение критерия поиска работы, которое в крайних случаях позволяет полностью отказаться от него, остаются применимыми два основных критерия, а именно: «не имеющие работы» и «готовые к работе в данный момент» (United Nations, 2008, para. 2.272).

560. Редактирование данных о незанятости по таким позициям, как «увольнение из-за отсутствия работы», «поиск работы», могло ли данное лицо получить работу, и «последний год работы» (если присутствует), должно производиться совместно. Кроме того, они должны быть сопоставимы с ответом об экономической активности и в большинстве случаев не должны указываться, если заполнены такие графы, как проработанное время, отрасль, род занятий, категория работника и место работы. И если специалисты по содержательной части переписи решат, что в графе «увольнение из-за отсутствия работы» необходимо сделать запись, когда там имеется пропуск или ответ недостоверен, то тогда можно использовать матрицу вменения по возрасту и полу и, возможно, по образовательной подготовке данного лица.

ii) Поиск работы

561. Редактирование показателя «поиск работы» должно производиться одновременно с редактированием показателя «увольнение из-за отсутствия работы» и «почему не ведется поиск работы». Сотрудники, занимающиеся содержанием переписи, должны разработать такие приемы редактирования, где использовались бы записи по этим элементам данных, с тем чтобы можно было исчислить и другие элементы. При редактировании следует учитывать местные и региональные условия, а также переменные величины переписи или обследования.

iii) Экономически неактивное на данный момент население

562. К категории «экономически неактивного на данный момент» населения или к категории лиц, «не входящих в состав рабочей силы», относятся все лица, которые не были ни «заняты», ни «безработными» в течение короткого отчетного периода, используемого для оценки текущей экономической активности (United Nations, 2008, para. 2.278). В зависимости от причины, по которой они не являются «на данный момент экономически активными», их можно классифицировать по одной из следующих групп:

- a) посещающие то или иное учебное заведение;
- b) выполняющие домашние обязанности;
- c) живущие на пенсию или доход от капитала;
- d) не работающие по иным причинам.

О редактировании показателя «экономически неактивное на данный момент население» говорится выше в разделе о редактировании данных об экономической активности.

iv) Почему не ведется поиск работы

563. Данный элемент следует редактировать только в отношении тех, для кого сделана запись «не ведется поиск работы»; у всех других в этом месте должен быть пробел. В качестве альтернативы, если в графах «род занятий», «отрасль» и «статус занятости» есть достоверные записи, то следует ввести код для лиц, «имеющих работу, но не работающих». Этот код обозначает экономически активных лиц, которые имели работу, но не работали в течение отчетного периода. Во всех остальных случаях, если не используется динамическое вменение, можно приписать ответ «неизвестно». Для стран, где используется метод динамического вменения, можно присвоить запись, применяя данные с указанием возраста, пола и основной деятельности.

b) Редактирование данных о статусе экономической активности

564. Экономическая активность, как правило, подразделяется на следующие категории:

экономически активное население

- 1) работающие лица;
- 2) неработающие лица;

экономически неактивное население

- 3) учащиеся;
- 4) домохозяйки;
- 5) пенсионеры и лица, получающие доход от капитала;
- 6) прочие.

i) Работающие лица

565. Если выбирается категория 1 («работающие лица»), следует указать такие переменные величины, как отработанное время, род занятий, отрасль, статус экономической активности и место работы. Если они не указаны, их следует отредактировать и заполнить, либо указав «неизвестно», либо с помощью значений «cold deck» или «hot deck». Если выбирается категория 1, то вместо переменных «увольнение из-за отсутствия работы», «поиск работы» и «последний отработанный год» следует оставить пропуски. Если они указаны, надо изменить на ПРОПУСК.

ii) Экономическая активность неработающих лиц

566. Если выбирается категория 2 («неработающие лица»), следует указать переменные «увольнение из-за отсутствия работы», «поиск работы» и «последний отработанный год». Если они достоверно не указаны, их следует отредактировать и указать либо как «неизвестно», либо с помощью значений «cold-deck» или «hot-deck». Если выбираются категории 2–6, то вместо таких переменных, как отработанное время, род занятий, статус экономической ак-

тивности и место работы, должны быть пропуски. Если они заполнены, следует оставить ПРОПУСК.

iii) *Экономическая активность учащихся и пенсионеров*

567. Если выбирается категория 3, «учащийся», специалисты по содержательной части переписи должны решить, нужно ли вставить в графе «посещение учебных заведений» запись «да, в учебном заведении». Если выбирается категория 5, «пенсионер», специалисты по содержательной части переписи должны решить, нужно ли устанавливать определенный возраст для выхода на пенсию.

iv) *Когда запись об экономической активности недостоверна и указаны переменные, касающиеся работающих лиц*

568. Если запись об экономической активности недостоверна и указаны некоторые переменные для отработанного времени, рода занятий, отрасли и места работы, то экономической активности респондента следует присвоить код 1 «работающее лицо». Для выбора приемлемого ответа, возможно, понадобится матрица вменения.

v) *Когда запись об экономической активности недостоверна и указаны переменные, касающиеся неработающих лиц*

569. Если указаны какие-либо переменные для граф «увольнение из-за отсутствия работы», «поиск работы» и «последний отработанный год», то записи для экономической активности следует присвоить значение от 2 до 6. Если данное лицо посещает школу, то код, вероятно, должен быть 3. Для лиц пожилого возраста значение кода должно быть, вероятно, 5. В других случаях специалисты по содержательной части переписи могут принять решение использовать матрицу вменения для присвоения приемлемого ответа.

vi) *Когда запись об экономической активности недостоверна и не указаны никакие экономические переменные*

570. Если по каким-либо элементам экономической активности ответа нет, то специалисты по содержательной части переписи, возможно, захотят использовать матрицы вменения для определения наиболее приемлемого ответа, а затем исчислить и другие экономические элементы данных.

2. Отработанное время

571. Отработанным временем является все время, фактически затраченное на производство товаров и услуг в регулярные рабочие часы и в сверхурочное время, в течение отчетного периода, принятого для определения экономической активности при проведении переписи. В случае короткого отчетного периода (например, недели, предшествующей проведению переписи), отработанное время рекомендуется измерять в часах. При этом отработанное время можно измерить путем запроса отдельной информации по каждому дню недели. Если отчетный период продолжительный (например, 12 месяцев, предшествующих проведению переписи), отработанное время следует измерять в неделях или днях, где это осуществимо, или исходя из более продолжительных интервалов

времени. Отработанное время должно также включить время, затраченное на деятельность, которая хотя и не ведет непосредственно к производству товаров и услуг, все же определяется как часть производственных и служебных задач и обязанностей, например подготовка, ремонт или содержание в порядке рабочего места или рабочих инструментов. На практике оно будет также включать время простоя в ходе этих видов деятельности, например время, затраченное на ожидание, обеспечение готовности к ожидаемому заданию и другие короткие перерывы. Следует исключить более длительные перерывы для приема пищи и время, когда работа не выполнялась по причине отпуска, праздничных дней, болезни или производственных конфликтов (United Nations, 2008, para. 2.323).

572. Этот элемент данных следует редактировать только в отношении тех лиц, чей ответ об экономической активности был «имеет работу, работающий» или «самостоятельно занятый, работающий». Для некоторых стран отработанное время должно также включаться для лиц, ведущих домашнее хозяйство. Следует принимать категории, которые заранее определены группой редактирования. Если не применяется метод динамического вменения, то пропуски, нулевые значения или нецифровые коды следует заменить на ответ «не сообщается», и специалисты по содержательной части переписи могут по своему усмотрению заменить переменную экономической активности на ответ «не работает», если указанные часы равны нулю.

573. В случае использования динамического вменения минимальные переменные для матрицы вменения включают данные по возрастным группам и полу, но можно использовать и другие переменные, например основные категории образовательной подготовки, рода занятий или отрасли.

3. Род занятий

574. Род занятий означает вид работы, выполнявшейся в течение отчетного периода занятым лицом (или вид работы, выполнявшейся ранее, если речь идет о безработном), независимо от отрасли хозяйства или статуса занятости, к которым должно быть отнесено данное лицо (United Nations, 2008, para. 2.301).

575. Данный элемент следует редактировать только в отношении тех лиц, чья экономическая активность указана как «работающий». Если не применяется метод динамического вменения, то пропуски, нулевые или недостоверные ответы следует заменить на запись «не сообщается».

576. Коды для рода занятий, как правило, разрабатываются таким образом, чтобы различные цифры представляли коды для основного и неосновного рода занятий. Дополнительные записи, которые почти неизбежны для категории «род занятий», сделают кодирование еще более трудным.

577. При использовании метода динамического вменения минимальные переменные для матрицы вменения включают данные по возрастным группам и полу, но могут использоваться и другие показатели, такие как основные категории образовательной подготовки или отрасли.

4. Отрасль

578. В соответствии со вторым пересмотренным изданием *Принципов и рекомендаций в отношении переписей населения и жилого фонда* (пункт 2.306) термин «отрасль» относится к деятельности заведения, в котором занятое лицо

работало в течение отчетного периода, установленного для сбора данных об экономических характеристиках (или работало в последнее время, если речь идет о неработающем). Для ориентации относительно выбора работы/деятельности, подлежащей классификации, см. пункт 2.307 (там же).

579. Этот элемент данных следует редактировать только в отношении лиц, чья экономическая активность указана как «работающий». Если динамическое вменение не используется, то пропуски, нулевые или недостоверные ответы следует заменить на запись «не сообщается».

580. Коды для отрасли, как правило, разрабатываются таким образом, чтобы различные цифры представляли коды основных и неосновных отраслей. Дополнительные записи, которые почти неизбежны для этого элемента, делают кодирование еще более трудным.

581. В случае применения метода динамического вменения минимальные переменные для матрицы вменения включают указание возрастных групп и пола, но могут использоваться и другие переменные, такие как основные категории образовательной подготовки или отрасли.

5. Статус занятости

582. Термин «статус занятости» относится к положению экономически активного лица в связи с его или ее занятостью, то есть к виду прямого или косвенного договора о найме, заключенного данным лицом с другими лицами или организациями в отношении своей работы. Основные критерии, применяемые для определения таких групп классификации, — это тип экономического риска, элементом которого является степень связи лица со своим местом работы, и степень его полномочий в отношении предприятия и других работников, которую лицо имеет или будет иметь на этой работе. Следует побеспокоиться об обеспечении классификации экономически активного лица по статусу занятости на основе той же работы (работ), которая (которые) использовалась (использовались) для классификации этого лица по роду занятий, отрасли или сектору (United Nations, 2008, para. 2.310).

583. Экономически активное население следует классифицировать по статусу занятости следующим образом (United Nations, 2008, para. 2.311):

- a) наемные работники, среди которых можно различать работников со стабильными контрактами (включая штатных работников) и других работников;
- b) работодатели;
- c) самостоятельные работники;
- d) работающие безвозмездно на семейном предприятии члены семьи;
- e) члены производственных кооперативов;
- f) лица, не поддающиеся классификации по статусу занятости.

584. Владельцев-управляющих акционерных предприятий, которые, как правило, должны быть отнесены к числу работающих по найму, но которых в определенных описательных и аналитических целях может оказаться более предпочтительным группировать вместе с работодателями, следует выделять отдельно.

585. Данный элемент следует редактировать только в отношении лиц, чья экономическая активность указана как «работающий». Если не используется динамическое вменение, то пропуски, нулевые или недостоверные ответы можно заменить на запись «не сообщается». В случае применения метода динамического вменения переменные для матрицы вменения включают указание возрастных групп и пола, но могут использоваться и другие показатели, такие как основные категории образовательной подготовки или отрасли.

6. Доход

586. Темы переписи, связанные с экономическими характеристиками населения, представленными во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда*, ставят в центр внимания экономически активное население, как оно определено в рекомендациях Международной организации труда (МОТ), где понятие экономического производства соотнесено с системой национальных счетов (СНС) (United Nations, 2008, para. 2.331). Экономически активное население включает всех мужчин и женщин, которые заняты или могут быть заняты в сфере производства товаров и услуг в соответствии с определением СНС в течение отчетного периода времени (United Nations, 2008, para. 2.237).

587. В рамках этой структуры доход может быть определен как *a*) ежемесячный доход в деньгах и/или натурой от работы, выполняемой каждым экономически активным лицом, или *b*) совокупный годовой доход домохозяйств в деньгах и/или натурой независимо от источника. Сбор достоверных данных о доходах, особенно доходах от самостоятельной деятельности и от собственности, чрезвычайно труден при проведении общих опросов на местах, в особенности при переписи населения. Включение дохода в неденежной форме еще более усложняет проблему. Сбор данных о доходах при переписи населения, даже если он ограничивается данными о денежных доходах, вызывает особые проблемы, связанные с увеличением объема работы и, в частности, с ошибками в ответах. Исходя из этого данная тематика, включая более широкое определение дохода, обычно считается более подходящей для использования при выборочном обследовании. Тем не менее в зависимости от национальных потребностей страны могут пожелать получать ограниченную информацию о денежных доходах. Собранная на основе такого определения информация может оказаться полезной для статистики распределения доходов, потребления и накопления в домохозяйствах в дополнение к непосредственным целям переписи.

588. Во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда* выделяются два вида дохода: индивидуальный доход и доход домохозяйства. Оба элемента требуют аналогичного редактирования. Для индивидуального дохода, если не применяется метод динамического вменения, недостоверные ответы о доходе следует отмечать как «не указано» или «неизвестно». Если используется метод динамического вменения, то для создания матрицы вменения данных о доходе можно использовать такие показатели, как возраст, пол, образовательная подготовка, отрасль, род занятий, и др.

589. Для этой переменной доход домохозяйства представляет собой сумму всех доходов домохозяйства, которая вносится в запись о жилом фонде. Однако редактирование с применением метода динамического вменения про-

исходит примерно так же, когда используются не данные о каждом отдельном лице, а показатели возраста, пола, уровня образования главы домохозяйства. По вопросам дополнительного рассмотрения проблем перекодирования данных о домохозяйстве и семейном доходе см. [приложение I](#).

7. Институциональный сектор

590. Институциональный сектор занятости определяет организационно-правовую форму и основные функции, поведение и цели конкретного заведения, с которым связана работа того или иного лица (United Nations, 2008, para. 2.335).

591. Существует взаимосвязь между некоторыми возможными отраслями и родом занятий и институциональным сектором занятости (корпорация, государственный сектор, некоммерческое учреждение, домохозяйство и другие). Некоторые страны могут по своему усмотрению проверять эту взаимосвязь между переменными, с тем чтобы убедиться в том, что в таблицах не будет несоответствий, если представить эти показатели в сводной таблице.

592. Для целей редактирования страны, не использующие метод динамического вменения, должны обозначить институциональный сектор, если он неизвестен, как «неизвестно». Страны, использующие метод динамического вменения, должны рассмотреть возможность использования показателей возраста и пола и, возможно, основной отрасли или рода занятий лиц с аналогичными характеристиками в данном географическом районе.

8. Занятость в неформальном секторе

593. В случаях когда деятельность неформального сектора играет важную роль в создании рабочих мест и получении дохода, некоторые страны используют элементы данных о статусе экономической активности и другие экономические элементы в целях выявления неформального сектора. Другие страны задают конкретные вопросы по поводу участия в неформальном секторе (United Nations, 2008, para. 2.337).

594. Схема редактирования данных по неформальному сектору должна быть прямой и простой. Если занятость в неформальном секторе не зависит от занятости в формальном секторе, то для пропущенных и недостоверных записей могут присваиваться ответы «неизвестно»; также можно использовать данные «hot deck», базирующиеся на информации по возрасту и полу. Если занятость в неформальном секторе находится в зависимости от занятости в формальном секторе, тогда можно включить дополнительную переменную из матрицы «hot deck», чтобы указать, занято ли данное лицо также и в формальном секторе (United Nations, 2008, para. 2.343).

9. Место работы

595. «Место работы» — это место, где занятое лицо в данный момент выполняет свою работу и где обычно занятое по найму лицо выполняет основную работу, используемую для определения его/ее других экономических характеристик, таких как род занятий, отрасль и статус занятости. Хотя информацию о месте работы можно использовать для разработки территориальных профилей

занятой рабочей силы (в отличие от демографических профилей по месту жительства), однако главная цель состоит в увязке информации о месте работы с данными о месте жительства (United Nations, 2008, para. 2.346).

596. Поскольку элемент «место работы» используется в статистике о миграции, необходимо, чтобы любые изменения в сообщаемой информации отражали конкретные географические районы. В связи с этим группы редактирования в конкретной стране могут по своему усмотрению ставить запись «неизвестно» для недостоверных ответов и анализировать только «известные» случаи.

597. Процедуры кодирования этого элемента потребуют больше времени и будут более сложными, если принимаются дополнительные записи, которые должны кодироваться. Если для цифр будет установлена определенная иерархия, например первая цифра будет обозначать провинцию или область, вторая — район и т.д., то в этом случае кодирование, возможно, будет более эффективным и более точным.

598. При составлении матриц вменения обработчики данных должны убедиться в том, что в эти матрицы будут вноситься только наиболее вероятные территориальные единицы. Возможно, было бы целесообразно начинать каждый раз новый набор данных «cold deck» по каждой административно-территориальной единице или другому географическому району, чтобы убедиться, что нельзя для этого использовать предыдущие значения. В сами матрицы вменения можно включить показатели возраста и пола и, возможно, измененные основные категории главного рода занятий или отрасли. Кроме того, для показателя работы внутри страны и за ее пределами, возможно, понадобятся разные матрицы вменения.

599. В настоящей главе были рассмотрены переменные, относящиеся к населению и рекомендуемые в издании *Принципы и рекомендации*. Странам необходимо тщательно протестировать отдельные переменные величины и их пространственную взаимосвязь с другими переменными как в «кабинетных» условиях, так и в условиях проводимых перед переписью обследований для обеспечения надежных и полных ответов. Поскольку элементы, связанные с населением, в отличие от переменных по жилому фонду, обычно включаются в сводные таблицы по многочисленным различным комбинациям, требуется их тщательная проверка.

Глава V

Редактирование данных о жилом фонде

600. В спецификациях для редактирования данных о жилом фонде учитывается достоверность отдельных элементов данных, а также согласованность между ними. Знание конкретных соотношений между элементами для данной страны позволяет планировать редактирование с точки зрения проверки на согласованность таким образом, чтобы обеспечить более высокое качество данных для сведения их в таблицы. Например, крыша жилищной единицы не должна быть из цемента, если стены построены из бамбука. Аналогичным образом, для того чтобы в строении был туалет со смывом, ванна или душ, в нем должен быть трубопровод.

601. Как и в случае с данными о населении, группа редактирования должна решить, что делать: вписывать «не указано», вычислять значение для показателя «неизвестно» или другое значение методом статического вменения («cold deck») или производить динамическое вменение («hot deck») по характеристикам других жилищных единиц. Здесь, как и во многих других случаях, предпочтительнее производить динамическое вменение, поскольку оно устраняет те виды исчислений, которые необходимы на стадии табулирования, когда есть информация только в самих таблицах, позволяющая принять решение о неизвестных величинах. Созданные таким образом матрицы вменения дают данные вместо пропусков, недостоверных величин или устраненных несоответствий, когда есть другие связанные с этим элементы с достоверными ответами. В некоторых странах могут быть различия в жилищных характеристиках по стране, но их почти нет в большинстве населенных пунктов, если их брать отдельно. В других странах могут быть существенные различия по конкретным элементам в разных населенных пунктах, особенно в городских и сельских районах. Эти различия необходимо учитывать при разработке матриц вменения и, в частности, в отношении исходных статических величин. Группа редактирования может по своему усмотрению уточнить обстоятельства, при которых следует заполнить тот или иной пропуск, взяв за основу предыдущую жилищную единицу с другими аналогичными характеристиками.

602. За исключением случая, когда в стране нет информации о помещениях коллективного (группового) проживания, каждому порядковому номеру должна соответствовать одна (и только одна) запись о жилище (см. главу III, в которой изложен ряд процедур обеспечения качества). В зависимости от решений группы редактирования, программа редактирования может создать запись о жилище, если она отсутствует. Аналогичным образом, эта программа может также устранить одну или более записей, если они дублируются или повторяются несколько раз.

603. В идеальном случае каждая запись о жилищах должна редактироваться избирательно только для подходящих элементов данных. Редактируемые

элементы могут отличаться в зависимости от городских/сельских, климатических и иных условий. Однако на практике лишь у нескольких стран есть время и опыт разработки и применения многомерных матриц для изменения недостающих или несогласующихся данных. Еще меньше стран фактически применяют выборочное редактирование.

604. Тем не менее, скорее по эстетическим, чем по техническим причинам, и в особенности применительно к элементам данных по жилому фонду, поскольку редактирование стало более сложным и подробным, в настоящее время все большее значение придается обеспечению того, чтобы выбранные географические районы содержали только «соответствующие» ответы. Например, если определенные географические районы страны не имеют электроснабжения, в них не должно быть кондиционеров воздуха, электрических холодильников или электроплит. Редактирование может быть применено для решения таких проблем, которые возникают в определенных географических районах для обеспечения отсутствия каких-либо отклонений от нормы в конечном наборе данных. Лучше всего, вероятно, использовать метод «ускоренного рассмотрения» полученных ответов для исключения действительно не относящихся к делу ситуаций. Например, хотя могут быть случаи приобретения отдельными обеспеченными жителями того или иного района газовых генераторов при отсутствии электроснабжения, группа редактирования может принять решение не включать такие случаи в набор данных.

605. Информация, собираемая с помощью переписного листа, будет также зависеть от типа жилых помещений (жилищная единица или помещения для группового проживания) и от того, свободна она или занята. Редактирование данных о жилых помещениях для коллективного или группового проживания может ограничиваться только теми элементами данных, которые собраны или в помещениях группового проживания или в других жилищных единицах.

606. По определению бездомные лица не могут иметь записей о жилище. Если тем не менее такие записи есть, поскольку страна предпочитает иметь для них идентификаторы, то их можно обрабатывать так же, как и записи для помещений коллективного проживания, или надо будет применять совсем другие приемы редактирования или не применять их вообще.

607. Иногда для конкретного элемента данных можно допустить запись «не сообщается». Это может потребоваться в тех случаях, когда у группы редактирования в данной стране нет хорошей основы для вменения ответов по той или иной характеристике. Решение оставить ответы «не сообщается» должно соотноситься с требованием выдавать соответствующие сведенные в таблицы характеристики в целях планирования и осуществления политики. Если планирующим органам необходима выборочная информация, т. к. ответы «не сообщается» распределяются так же, как и регистрируемые данные, то распределение случаев «не сообщается» не должно вызывать проблем. Однако если ответы «не сообщается» вносят какие-то искажения, то процедура вменения после сбора данных может быть проблематичной, особенно для небольших районов или конкретных условий. Например, если респонденты, занимающие по принятому в стране определению «не соответствующее стандарту» жилье, отказываются раскрыть некоторые из характеристик этого жилья и счетчик их не регистрирует, планирующие органы могут не иметь возможности ввести программы по исправлению положения, чтобы облегчить такие нестандартные условия.

608. Редактировать данные о жилье, как правило, проще, чем данные о населении, поскольку соответствующие сводные таблицы имеют обычно значительно меньшую сложность. В большинстве стран отдельные жилищные характеристики собираются только по различным территориальным уровням. Как говорилось выше, страны, предпочитающие не использовать метод динамического вменения, должны определить идентификатор для категорий «неизвестно», чтобы использовать его, когда возникают недостоверные или несогласующиеся ответы.

609. В странах, где используется метод динамического вменения, группа редактирования должна разработать простые матрицы вменения с параметрами, позволяющими проводить различия между жилищными характеристиками. Для большинства стран переменная «тип жилых помещений», будь то жилищная единица или помещения для коллективного проживания, включая тип единицы в рамках этих категорий, является наилучшей первичной переменной для динамического вменения.

610. Для некоторых стран географические районы можно использовать в качестве одного из измерений этих матриц вменения. Можно использовать и показатель условий владения. Например, если в данной стране около половины единиц арендуется, а другие находятся в собственности, условия владения являются удобным показателем для включения в качестве одного из измерений матрицы вменения. Однако если арендуется только около 5 процентов единиц, то более приемлемыми будут какие-то другие характеристики. Условия владения зачастую являются полезной переменной, которую можно использовать в матрицах вменения, особенно в странах с большой долей основных условий владения. К другим заслуживающим внимания характеристикам относятся тип стен и наличие электричества.

611. Для каждой страны конкретные переменные, включенные в качестве измерений матриц вменения, должны соответствовать переменным в наборе данных, поэтому в связи с элементами данных о жилье необходимо следить за тем, чтобы отдельные элементы, а также их комбинации различались по характеристикам.

А. Основные и дополнительные темы

612. Единицами учета при проведении переписи жилого фонда являются *a)* строения; *b)* жилые помещения и *c)* проживающие в них лица. В Организации Объединенных Наций был разработан список основных тем для редактирования, представляющих общий интерес и ценность, а также имеющих значение с точки зрения проведения всесторонних статистических сопоставлений на международном уровне. Для удобства пользователей ниже приводятся предлагаемые коды для этих и ряда других дополнительных тем. Эти темы сгруппированы по типу единиц учета.

1. Жилые помещения: тип жилых помещений (основная тема)

613. В приводимой ниже классификации дается описание системы трехзначных кодов, разработанной Организацией Объединенных Наций (United Nations, 2008, paras.2.412–2.454) для группирования в широкие категории

жилищных единиц и жилых помещений для коллективного проживания с аналогичными структурными характеристиками. Распределение жильцов (населения) по различным группам дает ценную информацию о жилых помещениях, имеющих на момент переписи. Эта классификация также дает полезную основу для стратификации при проведении выборочных обследований. Жилые помещения можно подразделить на следующие категории:

1. Жилищные единицы
 - 1.1. Обычные жилища:
 - 1.1.1. имеющие все основные удобства;
 - 1.1.2. не имеющие всех основных удобств
 - 1.2. Прочие жилищные единицы:
 - 1.2.1. временные жилищные единицы;
 - 1.2.2. передвижные жилищные единицы;
 - 1.2.3. импровизированные жилищные единицы;
 - 1.2.4. жилищные единицы в капитальных строениях, не предназначенных для жилья;
 - 1.2.5. другие помещения, не предназначенные для жилья
2. Жилые помещения для коллективного проживания
 - 2.1. Гостиницы, пансионаты и другие сдаваемые внаем меблированные помещения
 - 2.2. Специализированные учреждения:
 - 2.2.1. больницы;
 - 2.2.2. исправительные учреждения (тюрьмы, колонии);
 - 2.2.3. военные учреждения;
 - 2.2.4. религиозные учреждения (монастыри, конвенты и т. д.);
 - 2.2.5. дома престарелых и аналогичные учреждения;
 - 2.2.6. студенческие и аналогичные им общежития;
 - 2.2.7. жилые помещения для персонала (например, общежития для сотрудников или младшего медперсонала);
 - 2.2.8. сиротские приюты;
 - 2.2.9. прочее
 - 2.3. Лагеря и общежития для рабочих:
 - 2.3.1. военные лагеря;
 - 2.3.2. рабочие лагеря;
 - 2.3.3. лагеря беженцев;
 - 2.3.4. лагеря для внутренне перемещенных лиц;
 - 2.3.5. прочее
 - 2.4. Прочее

614. Группам редактирования необходимо разработать такие методики редактирования, которые обеспечат внутреннюю согласованность информации, касающейся всех жилых помещений для коллективного проживания и жилищных единиц. Если конкретный показатель типа жилья неизвестен или недостоверен, то группам редактирования следует применить такую методику, которая позволяет использовать другие показатели, чтобы определить тип жилых помещений. В противном случае, когда такой показатель недостоверен,

следует присвоить значение «неизвестно», если не применяется динамическое вменение. Национальным статистическим органам/бюро переписей, применяющим динамическое вменение в отношении недостоверных показателей, необходимо использовать по меньшей мере две характеристики, например тип строения, условия владения, количество комнат, размер площади или сведения о занятости помещений, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

2. Местонахождение жилых помещений (основная тема)

615. Местонахождение жилых помещений — это географическая переменная, о которой говорится в главе III, посвященной редактированию структуры данных.

3. Сведения о занятости (основная тема)

616. Решение регистрировать жилые помещения, жильцы которых временно отсутствуют или временно присутствуют, в качестве «занятых» или «незанятых» придется рассматривать в зависимости от того, проводится ли перепись населения де-юре или де-факто. И в том и в другом случае было бы полезно отличать по мере возможности жилые помещения, которые используются в качестве основного места жительства, от жилых помещений, используемых в качестве второго места жительства. Это особенно важно, если второе жилище значительно отличается по своим характеристикам от основного жилья, что, например, относится к тем случаям, когда люди, проживающие в сельских районах, переезжают в определенные периоды года из своих постоянных сельских жилых помещений в простейшие строения, размещенные на сельскохозяйственных угодьях (United Nations, 2008, para. 2.466). Ниже приводится рекомендуемая классификация обычного жилья по статусу занятости:

1. Занятое
2. Незанятое
 - 2.1. Незанятое на сезонной основе:
 - 2.1.1. загородные дома;
 - 2.1.2. сезонные рабочие поселки;
 - 2.1.3. прочее
 - 2.2. Незанятое не на сезонной основе:
 - 2.2.1. дополнительное жилье;
 - 2.2.2. для сдачи;
 - 2.2.3. для продажи;
 - 2.2.4. предназначено под снос;
 - 2.2.5. прочее

617. Если жилищная единица занята, то число жильцов и подсчет записей о заселении не должны равняться нулю. Если никаких проживающих лиц не зарегистрировано, то данная единица или свободна, или жильцы в ней отсутствуют. Как отмечалось ранее, в разделе о структурном редактировании специалисты должны разработать соответствующие процедуры, для того чтобы определить, свободна ли данная единица. Если она зарегистрирована как заня-

тая, но фактически свободна, то необходимо разработать какой-то метод для определения типа незанятости, указав «неизвестно» или используя метод динамического вменения. Если данная единица зарегистрирована как свободная, но можно определить, что она фактически занята (по имеющейся информации о числе жильцов или подсчете записей о заселении), то показатель занятости необходимо поменять на «занято».

618. Если этот показатель недостоверен, число жильцов равно нулю и не представлено никаких записей о заселении данной единицы, то, если не используется динамическое вменение, следует указать «свободно по неизвестной причине». Если этот показатель недостоверен, но число жильцов не равно нулю или имеются записи о заселении данной единицы, следует указать «занято». Страны, применяющие метод динамического вменения для недостоверных величин (для исчисления данных о занятости помещений), должны использовать по меньшей мере две характеристики для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе, или в качестве альтернативы указать «свободно по неизвестной причине».

4. Форма собственности (основная тема)

619. Данная тема относится к форме собственности самой жилищной единицей, а не к земельному участку, на котором она расположена (United Nations, 2008, para. 2.467). Форму собственности не следует смешивать с условиями владения. Необходимо получить информацию, показывающую, принадлежат ли данные жилищные единицы государственному сектору (центральное правительство, местные органы власти, государственные корпорации), или частному сектору (домохозяйства, частные корпорации, кооперативы, жилищные объединения или иные субъекты). Вопрос нередко ставится более широко, чтобы указать, полностью ли оплачено жилое помещение, производится ли его покупка в рассрочку или не взят ли под него залог. Ниже приводится следующая классификация жилых помещений по форме собственности:

1. Занимаемые владельцем
2. Занимаемые другими лицами:
 - 2.1. в собственности государства;
 - 2.2. в частной собственности;
 - 2.3. в муниципальной собственности;
 - 2.4. в кооперативной собственности;
 - 2.5. прочее

620. Если форма собственности связана с условиями владения, то это следует учитывать при разработке программы редактирования; если связи между ними нет, то тогда, вероятно, форма собственности не будет зависеть от других жилищных переменных. В случае недостоверности показателя «форма собственности», если не применяется динамическое вменение, следует указать значение «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, например, строительный материал для стен, условия владения, тип жилищной единицы и количество комнат, для получения

«известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

5. Число комнат (основная тема)

621. Комната определяется как площадь в пределах жилищной единицы или других жилых помещений, окруженная стенами от пола до потолка или до крыши или высотой не менее двух метров, достаточная для размещения одной кровати для взрослого человека, то есть не менее четырех квадратных метров. Таким образом, комнатами можно считать спальни, столовые, гостиные, кабинеты, пригодные для жилья мансарды, помещения для прислуги, кухни, помещения, используемые для профессиональных или деловых целей, а также другого рода отдельные помещения, используемые или предназначенные для жилья и отвечающие критериям в отношении стен и площади. Коридоры, веранды, вестибюли, ванны и туалетные комнаты не должны считаться комнатами, даже если они и отвечают этим критериям. Для целей национальной статистики можно отдельно собирать сведения по помещениям менее четырех квадратных метров, которые в других отношениях подходят под определение «комнаты», если считается, что их число оправдывает сбор таких данных (United Nations, 2008, para. 2.472).

622. Поскольку число комнат может не зависеть от других жилищных переменных, в случае недостоверности показателя, следует указать значение «неизвестно», если не используется динамическое вменение. Страны, применяющие метод динамического вменения в отношении недостоверных величин, должны использовать по меньшей мере две характеристики, например тип жилищной единицы, строительный материал для стен, условия владения жильем и сведения о занятости помещений, для получения «известной» информации, исходя из аналогичных жилищных единиц в данном географическом районе.

6. Число спальных комнат (дополнительная тема)

623. Помимо регистрации числа комнат, в ряде национальных переписей собирается информация о числе спален в жилищной единице, которая является единицей учета по данной теме. Спальня определяется как комната с кроватью, используемая для ночного отдыха (United Nations, 2008, para. 2.475).

624. Иногда счетчики указывают число спален, которое больше числа комнат¹⁰. Если такие факты отмечаются и если страна использует значение «не указано» только для недостоверных или несогласующихся ответов, то число спален следует обозначать как «не указано». Если используется динамическое вменение, спальни следует «оценивать» с помощью матрицы вменения, где в качестве одного из элементов стоит число комнат. Таким образом, число спален будет не больше числа комнат, потому что показатель для спален будет обновляться лишь в том случае, если показатели для комнат и спален согласуются между собой. Проще всего использовать линейный массив с числом комнат в качестве элементов, а в этих элементах будет указано число спален. Более сложная матрица вменения может включать число лиц в данной жилищной единице и тип структуры.

¹⁰ Если указаны и комнаты и спальни, то эти данные следует редактировать вместе, и число спален не должно превышать число комнат. Поскольку число спален — это «дополнительная» тема, то редактирование производится, если есть оба показателя.

625. В противном случае при недостоверности показателя числа спален следует указывать «неизвестно», если не используется динамическое вменение. Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики (одна из них — число комнат) для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

7. Полезная площадь помещения (дополнительная тема)

626. Жилая площадь — это полезная площадь в жилищных единицах, то есть площадь, измеряемая в пределах наружных стен жилищных единиц, исключая нежилые подвалы и чердаки. В многоквартирных зданиях должны исключаться все площади общего пользования. Подходы к жилищным единицам должны отличаться от подходов к жилым помещениям для коллективного проживания (United Nations, 2008, para. 2.476).

627. Жилая площадь может быть связана с числом комнат и/или числом спальных комнат, поэтому группа редактирования в конкретной стране, возможно, захочет принять это во внимание при разработке программы редактирования. К другим полезным элементам данных для динамического вменения можно отнести число жильцов и число жильцов на одну комнату. В большинстве случаев жилая площадь не зависит от других переменных редактирования, связанных с жильем. Возможно, потребуется указать единицу измерения, например квадратные метры. В случае недостоверности показателя, если не используется динамическое вменение, следует указать значение «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных величин, должны использовать по меньшей мере две характеристики, например тип жилищной единицы, строительный материал для стен, условия владения и сведения о занятости помещений для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

¹¹ Жилищная единица фактически является единицей учета в отношении следующих переменных: система водоснабжения, туалетное и канализационное оборудование, ванное оборудование, кухонное оборудование, освещение и удаление мусора.

8. Система водоснабжения (основная тема)¹¹

628. Согласно документам Организации Объединенных Наций (United Nations, 2008, para. 2.479), основные сведения о системе водоснабжения, которые следует собирать во время переписи, касаются того, обеспечиваются ли жилищные единицы водопроводной водой. Из этих сведений будет видно, обеспечиваются ли жилые помещения водой из коммунальной системы водоснабжения или из индивидуальной установки, например водонапорного бака или насоса. Единицей учета для этой темы является жилищная единица. Необходимо также указать, имеется ли водонапорный кран внутри жилищной единицы, а если нет, то находится ли он не далее определенного расстояния от входной двери. Рекомендуемое расстояние составляет 200 метров на основании того, что в этих пределах жильцы смогут брать воду для бытовых нужд без чрезмерных усилий. Помимо месторасположения водопроводного крана особого внимания заслуживает также источник получаемой воды. В связи с этим с точки зрения водоснабжения рекомендуется следующая классификация жилищных единиц:

1. Водопровод внутри жилищной единицы:
 - 1.1. от коммунальной системы водоснабжения;
 - 1.2. из частного источника
2. Водопроводный кран вне жилищной единицы, но не далее 200 метров от нее:
 - 2.1. от коммунальной системы водоснабжения:
 - 2.1.1. для исключительного пользования (жильцов данной жилищной единицы);
 - 2.1.2. для совместного пользования;
 - 2.2. из частного источника:
 - 2.2.1. для исключительного пользования;
 - 2.2.2. для совместного пользования
3. Прочее

629. Коммунальная система водоснабжения представляет собой техническую службу, которая управляется и инспектируется государственными органами. Такие системы обслуживаются, как правило, государственным или муниципальным органом, но в некоторых случаях они создаются кооперативным или частным заведением.

630. Элементы данных, касающиеся водохозяйственных установок — системы водоснабжения, снабжения питьевой водой, туалетное и канализационное оборудование, ванное оборудование и горячее водоснабжение — вероятно, следует редактировать вместе. Поскольку эти элементы тесно связаны между собой, если одного из них недостает или он указан неправильно, то другие можно использовать для определения недостающего показателя. В отношении географических районов, где нет проточной воды, специалисты, возможно, должны использовать специальные приемы редактирования данных о жилищных единицах. Если же в данном районе есть другие единицы с аналогичными характеристиками, то эти элементы данных рекомендуется использовать для динамического вменения, если таковое производится.

631. В случае недостоверности показателя системы водоснабжения, если не применяется метод динамического вменения, следует указывать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики. Они, как правило, включают тип жилищной единицы, а также туалетное, канализационное и ванное оборудование. Эти сведения можно использовать для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

9. Основной источник питьевой воды (основная тема)

632. Данные по снабжению питьевой водой необходимо редактировать совместно с данными о системе водоснабжения. Здесь может также применяться большинство из описанных выше критериев. Поскольку в переписной лист обычно включается бутилированная питьевая вода и другие нетрадиционные источники питьевой воды, их также необходимо включать в редактирование (United Nations, 2008, para. 2.483).

633. В случае недостоверности показателя по питьевой воде, если не применяется метод динамического вменения, следует указать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики. Они, как правило, включают тип жилищной единицы, а также систему водоснабжения, туалетное, канализационное и ванное оборудование. Эти сведения можно использовать для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

10. Тип туалета (основная тема) и

11. Канализационная система (основная тема)

634. Для получения наиболее согласованных результатов данные по типу туалета и канализационной системе должны редактироваться совместно с другими переменными по санитарно-техническим системам. Хотя во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда* эти две переменных объединены, в 2010 году они были разъединены. Тем не менее эти элементы следует редактировать совместно и, если возможно, использовать для этого одни и те же матрицы динамического вменения.

635. Некоторые страны сочли целесообразным расширить данную классификацию, включив в нее туалеты без слива, чтобы выделить некоторые широко используемые типы туалетов и указать на определенный уровень санитарных условий. Ниже приводится следующая рекомендуемая Организацией Объединенных Наций классификация жилищных единиц по типу туалета (United Nations, 2008, para. 2.487):

1. С туалетом внутри жилищной единицы:
 - 1.1. смывной туалет/туалет с ручным смывом;
 - 1.2. прочее
2. С туалетом за пределами жилищной единицы:
 - 2.1. для исключительного пользования (жильцов данной жилищной единицы):
 - 2.1.1. смывной туалет/туалет с ручным смывом;
 - 2.1.2. туалет с вентилируемой выгребной ямой благоустроенного типа;
 - 2.1.3. туалет с невентилируемой и закрытой выгребной ямой;
 - 2.1.4. отверстия или вырытые в земле ямы с временным покрытием или без него;
 - 2.1.5. прочее
 - 2.2. для совместного пользования:
 - 2.2.1. смывной туалет/туалет с ручным смывом;
 - 2.2.2. туалет с вентилируемой выгребной ямой благоустроенного типа;
 - 2.2.3. туалет с невентилируемой и закрытой выгребной ямой;
 - 2.2.4. отверстия или вырытые в земле ямы с временным покрытием или без него;
 - 2.2.5. прочее

3. Отсутствие какого-либо туалета:

- 3.1. подсобные средства или помещения с ведрами (удаление экскрементов вручную);
- 3.2. использование природного окружения, например кустов, рек, ручьев и т.д.

636. Данные о типе туалетного и канализационного оборудования являются еще одними элементами, касающимися водоснабжения в жилищных единицах, и их следует редактировать совместно с другими элементами, связанными с водоснабжением. Такие показатели, как «для личного пользования», «для совместного пользования», «только для жильцов данного жилого помещения» и т. д., можно использовать при определении согласованности данных. Если они не согласуются, то их можно использовать для того, чтобы установить, какой путь редактирования избрать для выявления проблемы. Когда есть одна или более других связанных с водоснабжением переменных, то оценку неизвестной или несогласующейся информации можно произвести, не прибегая к показателю «неизвестно» или к динамическому вменению. Однако, если это не дает достоверного значения и если не используется динамическое вменение, следует указывать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, в том числе, как правило, тип жилищной единицы, а также водоснабжение, строительный материал для стен, условия владения и сведения о занятости помещений, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

12. Ванные помещения и оборудование (основная тема)

637. Согласно документам Организации Объединенных Наций (United Nations, 2008, para. 2.490), необходимо получить сведения о том, установлена ли в пределах каждого комплекса жилых помещений стационарная ванна или душевая установка. Единицей учета для этой темы является жилищная единица. Могут быть собраны дополнительные сведения о том, находятся ли эти виды удобств в исключительном пользовании жильцов данной жилищной единицы или нет и имеется ли для ванны горячая или только холодная вода. Однако в некоторых районах мира предложенная выше классификация может не отвечать национальным потребностям. Вместо этого может быть важным, например, провести разграничение с точки зрения доступности между жилыми помещениями с отдельной ванной комнатой, отдельной ванной комнатой в здании, открытыми кабинами для мытья в здании и использованием общественными банями. Ниже приводится следующая рекомендуемая классификация жилищных единиц по наличию и типу ванных помещений и оборудования:

1. Жилищная единица со стационарной ванной или душевой установкой
2. Жилищная единица без стационарной ванны или душевой установки:
 - 2.1. Наличие стационарной ванны или душевой установки вне жилищной единицы:
 - 2.1.1. для исключительного пользования (жильцов данной жилищной единицы);
 - 2.1.2. для совместного пользования
 - 2.2. Отсутствие стационарной ванны или душевой установки

638. Данные о типах ванн помещений и оборудования следует редактировать вместе с другими элементами, касающимися водоснабжения. Такие показатели, как «для личного пользования», «для совместного пользования» или «исключительного пользования», можно использовать для определения согласованности показателей, и, если они не согласуются, для того чтобы установить, какой путь редактирования избрать для выявления проблемы. Если есть одна или более переменных, касающихся водоснабжения, оценку неизвестной или несогласующейся информации можно производить, не прибегая к указанию «неизвестно» или к динамическому вменению. Однако, если никакие приемы не помогают и если не используется динамическое вменение, в случае недостоверности показателя следует указывать «неизвестно». Страны, применяющие метод динамического вменения для исчисления недостоверных величин, должны использовать, по меньшей мере, две характеристики. Они включают, как правило, тип жилищной единицы, а также водоснабжение, строительный материал стен, условия владения или данные о занятости помещения. Эти сведения используются для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

13. Наличие кухни (основная тема)

639. Согласно второму пересмотренному изданию *Принципов и рекомендаций в отношении переписей населения и жилого фонда* (United Nations, 2008, параг. 2.494), сбор данных о наличии кухни может предоставить удобную возможность и для сбора сведений о типе оборудования, используемого для приготовления пищи, например кухонной плиты, электрической плиты или очага с открытым огнем, а также о наличии кухонной раковины и места хранения продуктов, обеспечивающего их сохранность. Ниже приводится рекомендуемая классификация жилищных единиц по наличию кухни или других мест, отведенных для приготовления пищи:

1. Жилищная единица, включающая кухню:
 - 1.1. для исключительного пользования (жильцов данной жилищной единицы);
 - 1.2. для совместного пользования
2. Жилищная единица, включающая другое место для приготовления пищи, например мини-кухню:
 - 2.1. для исключительного пользования;
 - 2.2. для совместного пользования
3. Жилищная единица без кухни или другого места для приготовления пищи:
 - 3.1. Наличие кухни или другого места для приготовления пищи вне жилищной единицы:
 - 3.1.1. для исключительного пользования;
 - 3.1.2. для совместного пользования
 - 3.2. Отсутствие кухни или другого места для приготовления пищи

640. При редактировании данных о кухонных помещениях и оборудовании такие показатели, как «для личного пользования», «для совместного пользования», «для исключительного пользования» и т. д., можно использовать при определении согласованности показателей, и, если они не согласуются

ся, для установления того, какой путь редактирования избрать для выявления проблемы. Когда есть одна или две переменных, касающихся наличия кухонных помещений и оборудования, оценку неизвестной или несогласующейся информации можно производить, не прибегая к значению «неизвестно» или к динамическому вменению. Однако в случае недостоверности данного показателя, если динамическое вменение не используется, следует указывать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, которые, как правило, включают тип жилищной единицы, а также водоснабжение, строительный материал для стен, условия владения и сведения о занятости помещений, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

14. Топливо, используемое для приготовления пищи (основная тема)

641. В связи с необходимостью строгого контроля за использованием природных ресурсов в ряд национальных переписей жилого фонда включается тема о топливе, используемом для приготовления пищи. Единицей учета является жилищная единица; под «топливом, используемым для приготовления пищи», понимается топливо преимущественно для приготовления горячей пищи. Если используются два вида топлива (например, электричество и газ), при переписи указывается наиболее часто используемое. Классификация видов топлива, используемого для приготовления пищи, зависит от национальных условий и может включать электричество, газ, мазут, уголь, дрова и навоз. Полезно также собирать эту информацию в отношении жилых помещений для коллективного проживания, особенно если число таких комплексов в стране является значительным (United Nations, 2008, para. 2.496).

642. Ответ на вопрос о типе топлива, используемого для приготовления пищи, следует редактировать вместе с ответами о наличии кухонных помещений и оборудования. Группа редактирования определяет соотношение обеих переменных и разрабатывает приемы редактирования для их проверки на согласованность. Такие показатели, как «для частного пользования», «для совместного пользования», «для исключительного пользования» и т. д., возможно, будут использоваться для определения согласованности величин, а если они не согласуются, то для того чтобы установить, какой путь редактирования избрать для выявления проблемы. Если есть одна или две переменные, касающиеся приготовления пищи, то оценку неизвестной или несогласующейся информации можно произвести, не прибегая к значению «неизвестно» или к динамическому вменению. Однако в случае недостоверности показателя, если не используется динамическое вменение, следует указывать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, включая тип кухонного помещения и оборудования, а также тип строения, строительный материал для стен, условия владения и сведения о занятости помещений, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

15. Освещение и/или электроснабжение (основная тема)

643. Следует собирать сведения о типе освещения жилых помещений — электричество, газ, керосиновые лампы или какие-либо другие источники. Если имеется электрическое освещение, то некоторые страны возможно пожелают собрать сведения о том, поступает ли электроэнергия из коммунальной сети, от электростанции или из какого-либо иного источника, например от промышленного предприятия. Помимо типа освещения странам следует оценивать информацию о наличии электричества для иных целей (таких как приготовление пищи, нагрев воды и отопление помещений). Если жилищные условия в стране дают возможность получать эту информацию из данных о типе освещения, то необходимости в дополнительных обследованиях не возникает (United Nations, 2008, para. 2.497).

644. В случае недостоверности данного показателя, если не используется динамическое вменение, следует указывать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных величин, должны использовать по меньшей мере две характеристики, включающие, как правило, тип жилищной единицы, а также строительный материал для стен, условия владения и сведения о занятости помещений, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

16. Основной метод удаления твердых отходов (основная тема)

645. Согласно второму пересмотренному изданию *Принципов и рекомендаций в отношении переписей населения и жилого фонда* (пункт 2.500), эта тема касается сбора и удаления твердых бытовых отходов, образуемых в результате проживания жильцов. Единицей учета здесь является жилищная единица. Ниже приводятся основные принципы классификации жилищных единиц по типу метода удаления твердых отходов:

1. твердые отходы, собираемые на регулярной основе специально назначенными организациями или лицами;
2. твердые отходы, собираемые на нерегулярной основе специально назначенными организациями или лицами;
3. твердые отходы, собираемые организациями и лицами, взявшими на себя эту задачу;
4. жильцы выбрасывают отходы на местную свалку, за состоянием которой следят власти;
5. жильцы выбрасывают отходы на местную свалку, за состоянием которой власти не следят;
6. жильцы сжигают твердые отходы;
7. жильцы закапывают твердые отходы;
8. жильцы сбрасывают твердые отходы в реку/море/ручей/пруд;
9. жильцы компостируют твердые отходы;
10. другие способы.

646. Данные о твердых отходах не зависят от других жилищных переменных. В случае недостоверности данного показателя, если не используется динамическое вменение, следует указывать «неизвестно». Страны, применяющие метод динамического вменения в отношении достоверных показателей,

должны использовать по меньшей мере две характеристики. Как правило, они включают тип жилищной единицы, а также строительный материал для стен, условия владения, сведения о занятости помещений или тип кухонного оборудования, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

17. Тип отопления и используемая энергия (дополнительная тема)

647. Эта тема относится к типу отопления жилых помещений и энергии, используемой для данной цели. Единицей учета являются все жилые помещения. Эта тема неактуальна в ряде стран, где в силу их географического положения и климата нет необходимости в отоплении жилых помещений. «Тип отопления» означает систему, используемую для обогрева большей части площади. Это может быть либо центральное отопление для обогрева всех комплексов жилых помещений или отдельного их комплекса, либо нецентральное с отоплением отдельных жилых помещений печкой, камином или каким-либо другим нагревателем. Показатель «энергия, используемая для отопления» тесно связан с типом отопления и касается преобладающего источника энергии, например твердого топлива (уголь, лигнит и их продукты, дрова), мазут, газ (природный или сжиженный) и электричество (United Nations, 2008, para. 2.501).

648. Тип отопления и энергия, используемая для отопления, связаны между собой, а также с наличием горячей воды и другими видами теплоснабжения, используемыми в жилищной единице, например с электричеством и магистральным газом. Группы редактирования должны учитывать наличие этих элементов при разработке спецификаций для программ редактирования по виду отопления и энергии, используемой для обогрева. Тип отопления может не зависеть от других жилищных показателей, поэтому эти данные можно редактировать отдельно. Однако если показатель «энергия, используемая для отопления» неизвестен или не согласуется с другими данными, то программа может проверить вид энергии, используемой для освещения. И наконец, в случае недостоверности данного показателя, если не используется динамическое вменение, следует указывать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, например тип жилищной единицы, а также строительный материал для стен, условия владения и сведения о занятости помещений, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в географическом районе.

18. Наличие горячей воды (дополнительная тема)

649. Данная тема относится к обеспеченности жилых помещений горячей водой. Горячая вода — это вода, нагретая до определенной температуры и подаваемая жильцам в помещение по трубам через кран. Сведения могут собираться в отношении наличия горячей воды в жилых помещениях, либо вне их только для исключительного или совместного пользования, либо ее отсутствия вообще (United Nations, 2008, para. 2.502).

650. Наличие горячей воды может быть связано со средствами подогрева воды, хотя использование солнечной энергии для подогрева воды может

быть и не связано с другими жилищными элементами. Группа редактирования должна определиться в отношении соответствующих приемов редактирования в зависимости от других жилищных элементов и географического положения. В конечном счете в случае недостоверности данного показателя, если не используется динамическое вменение, следует указывать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, такие как характеристики для воды, подаваемой по трубопроводу, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

19. Наличие магистрального газа (дополнительная тема)

651. Данная тема относится к наличию или отсутствию подвода магистрального газа в жилые помещения. Магистральный газ обычно определяется как природный или промышленный газ, который распределяется по трубам и потребление которого регистрируется. Эта тема может быть неактуальна для ряда стран, где либо отсутствуют источники природного газа, либо нет развитой системы газовых магистралей (United Nations, 2008, para. 2.503).

652. Магистральный газ не связан с другими жилищными элементами, за исключением типа освещения и топлива, используемого для приготовления пищи. Группа редактирования должна определить приемлемый путь редактирования, а также произвести проверку на согласованность. Если данный показатель остается недостоверным или не согласуется с другими данными, то, когда не используется динамическое вменение, следует указывать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, такие как энергия, используемая для отопления, тип строения, тип жилищной единицы, строительный материал для стен, условия владения и сведения о занятости помещений, для получения «известной» информации на базе сведений об аналогичных единицах в данном географическом районе.

20. Использование жилищной единицы (дополнительная тема)

653. «Использование жилищной единицы» означает полное или неполное использование жилищной единицы для целей проживания (пребывания). Жилищная единица может использоваться для проживания или для коммерческих, а также производственных или иных целей (United Nations, 2008, para. 2.504).

654. Показатель «использование жилищной единицы» не зависит от других жилищных элементов. В случае недостоверности данного показателя, если не используется динамическое вменение, следует указывать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, такие как тип жилищной единицы, строительный материал для стен, условия владения и форма собственности, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

21. Занятость более чем одним домохозяйством (основная тема)

655. Занятость более чем одним домохозяйством не зависит от других жилищных данных. Если значение недостоверно, страна должна посчитать число глав домохозяйств и использовать полученное число. Важно отметить, что такое редактирование должно осуществляться после структурного редактирования, определяющего главу домохозяйства.

22. Число жильцов (основная тема)

656. Каждое лицо, обычно проживающее в какой-либо жилищной единице или в комплексе жилых помещений для коллективного проживания, следует регистрировать как жильца. Исходя из этого учетной единицей здесь является жилое помещение. Однако, поскольку переписи жилого фонда обычно проводятся одновременно с переписями населения, применимость этого определения зависит от того, содержит ли собранная и зарегистрированная информация в ходе переписи населения указания на то, где именно каждое лицо находилось в день переписи, или указание на постоянное место проживания этого лица. В отношении тех, кто занимает передвижные жилищные единицы, такие как лодки, фургоны, автоприцепы, следует четко разграничивать лиц, занимающих их в качестве жилого помещения, и лиц, использующих подобные единицы как средство транспорта (United Nations, 2008, para. 2.510).

657. «Число жильцов» связано с количеством записей о заселении конкретной жилищной единицы, и оба показателя должны быть идентичны. Если это не так, то должны быть приняты меры по внесению корректив в число жильцов или в «количество записей о заселении». Как правило, корректируется число жильцов, с тем чтобы оно равнялось числу лиц в данной жилищной единице. Этот элемент не следует указывать как «неизвестно», и его не следует исчислять.

23. Тип здания (основная тема)

658. Организация Объединенных Наций (United Nations, 2008, para. 2.514) для строений, в которых хотя бы часть площади используется как жилье, рекомендует применять следующую классификацию:

1. Строения, состоящие из одной жилищной единицы:
 - 1.1. отдельные;
 - 1.2. примыкающие
2. Строения, имеющие более одной жилищной единицы:
 - 2.1. до двух этажей;
 - 2.2. от трех до четырех этажей;
 - 2.3. от пяти до десяти этажей;
 - 2.4. одиннадцать и более этажей
3. Строения для лиц, проживающих в специализированных учреждениях
4. Все прочие строения

659. В случае недостоверности данного показателя, если не применяется динамическое вменение, следует указывать значение «неизвестно». Страны, применяющие динамическое вменение в отношении недостоверных показа-

телей, должны использовать по меньшей мере две характеристики, например строительный материал для наружных стен, период строительства и/или тип жилищных единиц в строении, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

24. Год или период строительства (дополнительная тема)

660. Эта тема относится к возрасту строения, в котором находятся комплексы жилых помещений. Рекомендуется точно указывать год строительства для строений, возведенных в период между предыдущей и проводимой переписями, если этот период не превышает десяти лет. Если период между переписями превышает десять лет или если в данном районе еще никогда не проводилась перепись, точный год строительства нужно указать для строений, возведенных в течение предшествующих десяти лет. Что касается ранее построенных домов, то сведения следует собирать за периоды, которые дадут возможность оценить срок службы имеющегося жилого фонда. При сборе данных по этой теме могут возникнуть трудности, поскольку в некоторых случаях жильцы могут не знать года возведения строения (United Nations, 2007, 2.519).

661. Некоторые страны, даже используя динамическое вменение, указывают ответ «неизвестно» для элемента данных, относящегося к году или периоду строительства. В таком случае страна может и не использовать динамическое вменение для этого элемента, даже если используются матрицы вменения для других переменных. Страны, применяющие метод динамического вменения в отношении недостоверных величин, должны использовать по меньшей мере две характеристики, в том числе тип строения, строительный материал для наружных стен и/или тип жилищных единиц в рассматриваемом строении для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

25. Число жилищных единиц в здании (дополнительная тема)

662. Схема редактирования данных о числе жилищных единиц в здании разъясняется в главе III в рамках редактирования структуры.

26. Строительный материал для наружных стен (основная тема)

663. Эта тема касается строительного материала для наружных (внешних) стен строения, где расположены группы жилых помещений. Если стены построены из нескольких материалов, следует указать преобладающий материал. Выбор типа материалов, по которым собираются сведения (кирпич, бетон, дерево, необожженный кирпич и т. д.), будет зависеть от тех материалов, которые наиболее часто используются в строительстве в данной стране, и от их значения с точки зрения долговечности постройки или оценки ее прочности (United Nations, 2008, para. 2.525).

664. В случае недостоверности данного показателя, если не применяется динамическое вменение, следует указывать значение «неизвестно». Страны, применяющие динамическое вменение в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, такие как период строительства и/или тип жилищных единиц в рассматриваемом строе-

нии, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

27. Строительный материал для пола и крыши (дополнительная тема)

665. В некоторых случаях материалы, применяемые для крыш и полов, могут представлять особый интерес и использоваться для дополнительной оценки качества жилых помещений в строении. Эта тема касается материалов, используемых для крыши и/или пола (хотя в зависимости от конкретных потребностей страны, он может также относиться и к другим частям строения, таким как каркас или фундамент). Единицей учета является строение. Регистрируются только наиболее распространенные материалы, а что касается крыши, то это может быть черепица, бетон или кровельное железо, пальмовые листья, солома, бамбук или аналогичные растительные материалы, а также грунт, пластиковое покрытие или некоторые другие материалы (United Nations, 2008, para. 2.528).

666. Иногда ответ на вопрос о строительном материале для наружных стен не согласуется с ответом на вопрос о строительном материале для крыши; это может произойти, например, если строительный материал, определенный для стен, недостаточно прочный для того, чтобы поддерживать крышу. Как отмечалось выше, в этом случае специалисты должны решить, какой из двух показателей изменить, или указать значение «неизвестно». В случае недостоверности того или иного показателя, если не используется динамическое вменение, следует указывать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, например тип строения, тип жилищной единицы, строительный материал для стен, условия владения и сведения о занятости помещений, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

667. Данные о строительном материале для пола могут согласовываться или не согласовываться с данными о строительных материалах для крыш и стен. Если группа редактирования в стране обнаруживает несогласованность или недостоверные комбинации, она должна решить, указывать ли значение «неизвестно» или использовать матрицы вменения для корректировки одного или нескольких ответов. В случае недостоверности данного показателя, если не используется динамическое вменение, следует указывать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, например тип строения, строительный материал для наружных стен, тип жилищной единицы, условия владения и сведения о занятости помещений, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

28. Наличие лифта (дополнительная тема)

668. Эта тема касается наличия лифта (встроенной поднимающейся или опускающейся платформы для транспортировки людей и грузов) в многоэтажном строении. Сведения собираются о лифте, работающем в течение большей части времени: иными словами о лифте, который функционирует большую

часть времени при условии его регулярного технического обслуживания (United Nations, 2008, para 2.529).

669. Если строение одноэтажное или это отдельно стоящая обособленная единица, то лифт присутствовать не должен. Если все же он указан, группа редактирования должна решить, чему отдать предпочтение: числу этажей или тому факту, что там есть лифт. Если предпочтение отдается лифту, то необходимо изменить число этажей, указав «неизвестно» или используя метод динамического вменения для получения другого значения. Если предпочтение отдается числу этажей, а строение имеет всего один этаж, то ответ на вопрос «наличие лифта» нужно изменить на «нет».

670. При наличии лифта, который требует электроэнергии, необходимо проверить, электрифицировано ли данное строение.

671. И наконец, в случае недостоверности данных о лифте, если не используется динамическое вменение, следует указать значение «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, такие как тип строения и строительный материал для наружных стен, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

29. Сельскохозяйственное строение (дополнительная тема)

672. В ряде стран сочли необходимым уточнять при проведении своих национальных переписей, имеют ли регистрируемые строения сельскохозяйственное назначение или нет. Строение сельскохозяйственного назначения представляет собой часть сельскохозяйственного владения и используется для сельскохозяйственных целей и/или проживания (United Nations, 2008, para. 2.531).

673. Данные о строениях сельскохозяйственного назначения не зависят от других жилищных показателей. Страны могут по своему усмотрению проверить их на соответствие с данными о населении в отношении рода занятий и отрасли. В случае недостоверности данного показателя, если не используется динамическое вменение, следует указать значение «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географической районе.

30. Эксплуатационное состояние (дополнительная тема)

674. Данная тема касается того, нуждается ли рассматриваемое строение в ремонте, и определяет характер необходимого ремонта. Единицей учета является строение. Классификация строений по их эксплуатационному состоянию может включать такие показатели, как «ремонт не требуется», «требуется незначительный ремонт», «требуется средний ремонт», «требуется серьезный ремонт» и «ремонту не подлежит». Незначительный ремонт относится главным образом к регулярному поддержанию эксплуатационного состояния строения и его компонентов, например замена разбитых стекол. Средний ремонт предусматривает устранение более серьезных дефектов, таких как недостающие во-

досточные желоба на крыше, большие участки осыпавшейся штукатурки или отсутствие надежных перил у лестниц. Серьезный ремонт необходим в случае значительных структурных дефектов строения, таких как отсутствие дранки или черепицы на крыше, трещины и отверстия в наружных стенах или недостающие ступеньки. Термин «ремонт не подлежит» указывает на то, что строение отремонтировать невозможно: это означает, что у строения так много серьезных структурных дефектов, что представляется более целесообразным его снести, чем ремонтировать. Чаще всего этот термин относится к строениям, от которых остался только каркас с полуразрушенными наружными стенами и/или крышей (United Nations, 2008, para. 2.532).

675. Данные об эксплуатационном состоянии строения не зависят от других жилищных показателей. Следовательно, в случае недостоверности этого показателя, если не используется динамическое вменение, следует указывать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, например тип строения, строительный материал для наружных стен и тип жилищной единицы, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

31. Характеристики главы или другого основного члена домохозяйства (основная тема)

676. Характеристики главы домохозяйства берутся, как правило, из записей о населении, с тем чтобы легче было включать информацию в сводную таблицу в целях планирования и проведения анализа. Эти данные, включая этническое происхождение, вероисповедание или доход, помогают определить различие в социальном статусе или потребностях. Поскольку эти характеристики уже редактировались как элементы, относящиеся к населению, то никакого дополнительного редактирования здесь не требуется (United Nations, 2008, para. 2.533).

32. Условия владения жильем (основная тема)

677. Согласно документам Организации Объединенных Наций (United Nations, 2007, para. 2.536), условия владения определяются договоренностями, в соответствии с которыми домохозяйство занимает всю или часть жилищной единицы. Единицей учета здесь является домохозяйство, занимающее ту или иную жилищную единицу. По условиям владения домохозяйства классифицируются следующим образом:

1. Член домохозяйства является владельцем жилищной единицы
2. Член домохозяйства арендует всю жилищную единицу или ее часть:
 - 2.1. член домохозяйства является главным арендатором жилищной единицы или ее части;
 - 2.2. член домохозяйства является субарендатором части жилищной единицы
3. Жилищная единица заселяется без оплаты аренды
4. Прочие договоренности

678. Условия владения могут соотноситься с формой собственности, поэтому группа редактирования при составлении программы редактирования, возможно, должна будет учесть соотношение между этими двумя элементами данных. В случае недостоверности данного показателя условий владения, если не используется динамическое вменение, следует указать значение «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, такие как тип жилищной единицы, арендная плата и сведения о занятости помещений, для получения «известной» информации, исходя из аналогичных жилищных единиц в данном географическом районе.

33. Арендная плата за жилье и жилищные расходы владельца (дополнительная тема)

679. Данные об арендной плате за жилье и жилищные расходы владельца не зависят от других жилищных переменных за исключением того, что конечно расходы на арендную плату должны указываться только для сдаваемых в аренду единиц, а расходы владельца — только для единиц, занимаемых самим владельцем. Группа редактирования должна в каждом конкретном случае определять наиболее приемлемые взаимосвязи между этими переменными. В случае недостоверности данного показателя, если не используется динамическое вменение, следует присваивать значение «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных величин, должны использовать по меньшей мере две характеристики для получения «известной» информации, исходя из аналогичных жилищных единиц в данном географическом районе (United Nations, 2008, para. 2.540).

34. Меблированные и немеблированные (дополнительная тема)

680. Элемент данных касательно наличия или отсутствия мебели в жилищной единице является новым. Группы редактирования должны рассмотреть возможность тестирования этого элемента, если он включен в набор, для определения наилучших элементов для использования в процессе динамического вменения, если этот метод применяется для устранения недостоверных или несогласующихся данных (United Nations, 2008, para. 2.542).

35. Наличие информационно-коммуникационных устройств (основная тема)

681. В современном обществе существенно возросла важность наличия устройств, использующих информационно-коммуникационные технологии (ИКТ). Такие устройства предоставляют комплекс услуг, которые меняют структуру и закономерности крупных социальных и экономических явлений. Перепись населения обеспечивает прекрасную возможность оценки наличия таких устройств в домохозяйствах. Выбор тем должен быть достаточным для понимания роли, которую информационно-коммуникационные технологии играют в конкретном домохозяйстве, а также для их использования в целях планирования правительством и частным сектором, с тем чтобы расширить и улучшить предоставление соответствующих услуг и оценить их воздействие на общество. Рекомендуется следующая классификация:

1. домохозяйство, в котором имеется радиоприемник
2. домохозяйство, в котором имеется телевизор
3. домохозяйство, в котором имеется стационарный телефон
4. домохозяйство, в котором имеется мобильный(е) телефон(ы)
5. домохозяйство, в котором имеется персональный(е) компьютер(ы)
6. домохозяйство, в котором имеется домашний выход в Интернет
7. домохозяйство с доступом в интернет вне дома
8. домохозяйство, не имеющее выхода в Интернет

682. Устройства, использующие информационно-коммуникационные технологии, являются новыми элементами данных. Устройства, требующие для своей работы электричества, должны присутствовать только там, где имеется электроснабжение. Однако по мере расширения использования солнечной и ветровой энергии и других «возобновляемых» источников энергии данный фактор должен учитываться при разработке схем редактирования для этого элемента. Группам редактирования в различных странах необходимо тщательно протестировать этот элемент данных и соответствующие матрицы вменения до проведения переписи или обследования. Полезные элементы для «hot decks» включают социальный уровень домохозяйства (например, определяемого индексом благосостояния) и возраст главы домохозяйства (United Nations, 2008, para. 2.543).

683. Эти темы относятся к наличию конкретного устройства в жилищной единице. Например, телефон — это прежде всего телефонная линия, а не тот или иной телефонный аппарат, поскольку к одной телефонной линии может быть подключено несколько телефонных аппаратов (United Nations, 2008, paras. 2.547–548). В ходе редактирования данные о наличии телефона не связываются с другими элементами данных по жилищной единице. Однако, если в определенных географических районах телефонная связь отсутствует, группа редактирования должна принимать это во внимание при разработке схем редактирования. В случае недостоверности данных о наличии телефона, если не используется динамическое вменение, следует указывать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, такие как тип жилищной единицы, строительный материал для наружных стен или условия владения, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

36. Число автомобилей (дополнительная тема)

684. Показатель «число автомобилей» относится к числу автомобилей или легковых автофургонов, имеющих обычно в пользовании у членов домохозяйства. Как правило, термин «обычно имеющиеся в пользовании» означает наличие автомобилей и легковых автофургонов, которые находятся либо во владении жильцов, либо в пользовании, в соответствии с более или менее постоянным соглашением, например договором аренды, а также предоставленных работодателем, при их наличии, для использования домохозяйством, исключая при этом автофургоны, используемые только для перевозки товаров (United Nations, 2008, para. 2.551).

685. Число транспортных средств не зависит от других жилищных показателей. Если в стране есть районы, где нет никаких транспортных средств,

специалисты могут по своему усмотрению разработать особые приемы редактирования для конкретных географических районов. Если этого не сделано, то в случае недостоверности данного показателя, если не используется динамическое вменение, следует указать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, такие как тип жилищной единицы, строительный материал для наружных стен, условия владения и форма собственности, или в данном конкретном случае число взрослых жильцов, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

37. Наличие бытовой техники длительного пользования (дополнительная тема)

686. Сведения собираются о наличии такой бытовой техники длительного пользования, как стиральные машины, посудомоечные агрегаты, холодильники, морозильные камеры и т. д., в зависимости от условий в той или иной стране (United Nations, 2008, para. 2.552).

687. Для функционирования большей части бытовой техники необходимо наличие электропитания в жилищной единице. При указании данных о такой бытовой технике группа редактирования должна разработать соответствующую программу, которая проверяла бы наличие электропитания (с возможными исключениями для холодильника, который может работать от бензинового двигателя или в виде «ящика со льдом»). Кроме того, если в какой-либо стране, для того чтобы работала стиральная машина или посудомоечный агрегат, необходим водопровод, то при редактировании это также необходимо учитывать. Программу редактирования можно использовать для того, чтобы определить, требуется ли наличие какого-то отдельного элемента данных в зависимости от наличия электричества и воды, и следует ли произвести соответствующие действия, если возникают несоответствия. Кроме того, в отдельных частях страны может не быть электричества или водопровода, и специалистам это, видимо, необходимо учитывать при разработке программ редактирования. В случае недостоверности или несогласованности данного показателя, если не используется динамическое вменение, следует указать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, такие как тип жилищной единицы, наличие электричества, строительный материал для наружных стен и условия владения, для получения «известной» информации на базе сведений об аналогичных единицах в данном географическом районе.

38. Наличие свободной площади вне помещения (дополнительная тема)

688. Эта тема относится к наличию свободной площади вне помещений, предназначенной для отдыха членов домохозяйства, занимающих жилищную единицу. Классификация в данном случае может относиться к свободной площади вне помещения, которой пользуются как частью жилищной единицы (например, задний двор в отдельных домах), свободные площади, прилегающие к

строению (например, задний двор и площадка для игр, находящиеся вблизи многоквартирных строений), свободные площади, входящие в общую зону отдыха в пределах 10 минут ходьбы от жилищной единицы (например, парки, спортивные площадки и т. п.) или свободные площади вне помещений, находящиеся на удалении, то есть не в пределах 10 минут ходьбы (United Nations, 2008, para. 2.553).

689. Размер свободных площадей вне помещений, которой пользуются члены домашнего хозяйства, не зависит от других жилищных элементов. Однако в некоторых географических районах или определенных типах строений никаких свободных площадей вне помещений может и не быть. В этом случае группы редактирования, возможно, должны учитывать конкретные обстоятельства при разработке своих программ. В случае недостоверности данного показателя, если не используется динамическое вменение, следует указать «неизвестно». Страны, применяющие метод динамического вменения в отношении недостоверных показателей, должны использовать по меньшей мере две характеристики, например тип строения и тип жилищной единицы, для получения «известной» информации на базе сведений об аналогичных жилищных единицах в данном географическом районе.

В. Занятые и незанятые жилищные единицы

690. Описанные выше приемы редактирования касаются данных о занятых жилищных единицах. Однако у занятых жилищных единиц и свободных жилищных единиц зачастую разные характеристики и одни и те же приемы редактирования к ним не подходят. Группа редактирования в национальном бюро переписи/статистическом управлении должна будет разработать разные программы редактирования для каждого типа единицы, если, как это обычно и бывает, для незанятых жилищных единиц собраны не все элементы данных по жилому фонду. Группе редактирования надо будет уделить особое внимание переменным в матрице вменения, поскольку, вероятнее всего, они будут разными.

691. В настоящей главе рассматривались переменные величины по жилому фонду, рекомендуемые во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда*. Ни одна из стран не должна использовать все эти переменные, а выбранные переменные и их пространственные взаимосвязи с другими переменными необходимо тщательно протестировать в кабинетных условиях и в проводимых перед переписью обследованиях в целях обеспечения полноты и надежности ответов. Переменные по жилому фонду важны сами по себе в качестве составной части индекса благосостояния для оценки уровня благосостояния во всей стране или в ее частях.

Приложение I

Производные переменные

1. Для того чтобы наилучшим образом использовать данные по результатам переписи или обследования, в странах зачастую нужны переменные, представляющие собой комбинации и вариации других переменных. Например, элемент данных, касающийся статуса экономической активности (глава IV, раздел D.1), — это уже комбинация нескольких переменных, собранных во время переписи. Вместо того чтобы разрабатывать специальную программу для повторного кодирования информации каждый раз, когда национальному бюро переписи/статистическому управлению понадобится специальная таблица, специалисты по обработке данных могут записать программу для повторного кодирования один раз, хранить повторно закодированную информацию в файле конкретного лица и затем использовать ее для последующих таблиц. Национальные бюро переписи/статистические управления должны решить, как часто использовать повторно закодированные данные в таблицах, и насколько необходимы будут конкретные повторно закодированные данные, когда они будут определять, нужно ли создавать и хранить эту информацию. Следует помнить о том, что повторно закодированные данные также занимают определенное место в записях, касающихся конкретного лица. Чем больше численность населения, тем больше места будет использоваться.

2. Таким образом, можно создать множество переменных. Например, если сообщается дата рождения, а не возраст, то возраст можно определить путем вычитания даты рождения из даты проведения переписи, и эта информация будет храниться в записях. Аналогичным образом можно вычислить совокупный доход домохозяйства, суммируя доход каждого лица и указывая эту сумму в записи о жилом фонде для дальнейшего использования.

3. Иногда производные переменные получаются из комбинации одного или нескольких показателей в одной записи, а иногда из нескольких записей. Например, классификация «Экономически неактивное население, посещающее учебные заведения» может потребовать просмотра ответов на четыре вопроса. При разработке формата таблиц или планировании дополнительных таблиц использование производных переменных облегчит программирование и сделает его более эффективным, а также будет способствовать тому, чтобы данные оставались сопоставимыми с течением времени. Ниже приводятся некоторые примеры возможных производных записей.

A. Производные переменные для записей о жилом фонде

1. Доход домохозяйства

4. Производные переменные для дохода домохозяйства складываются из суммы дохода, включающего все категории дохода всех лиц в домохозяйстве.

Категории информации о доходе могут включать заработную плату, доход от собственного дела, проценты и дивиденды, социальное обеспечение и пенсии, денежные переводы, гонорары и доходы от аренды. Если собираются также сведения об общем доходе, то при редактировании следует проверить общий доход каждого лица путем суммирования отдельных категорий. Затем эта общая сумма проверяется по зарегистрированному общему доходу. Если суммарный общий доход не совпадает с зарегистрированным общим доходом, то группы редактирования должны разработать план корректировки. При этом необходимо или изменить общий итог, чтобы отразить сумму отдельных частей, или одну или более отдельных категорий дохода. Когда устанавливается общая сумма дохода для всех членов домохозяйства, переменная величина «доход домохозяйства» складывается из суммы доходов его отдельных членов.

5. Группа редактирования должна учитывать ситуацию, при которой один или более членов домохозяйства имеет отрицательный доход из-за неудачного бизнеса или по другим причинам. В этом случае общий доход домохозяйства не увеличится, а уменьшится на сумму дохода этого конкретного лица.

2. Семейный доход

6. Производная переменная для семейного дохода складывается из суммы дохода, получаемого по всем категориям дохода всеми членами семьи. Семьи, в отличие от домохозяйств, как правило, состоят только из родственников, но это определение будет зависеть от ситуации в конкретной стране. Для некоторых стран домохозяйства и семьи — это одно и то же, поэтому производная переменная для семейного дохода будет необязательна. Сведения о семейном доходе могут включать такие категории, как заработная плата, доход от собственного дела, проценты и дивиденды, социальное обеспечение и пенсии, денежные переводы, гонорары и доходы от аренды. Если собираются также сведения об общем доходе, то при редактировании следует проверить общий доход каждого лица путем суммирования отдельных категорий. Затем эта итоговая сумма сверяется с зарегистрированным общим доходом. Если суммарный общий доход не совпадает с зарегистрированным общим доходом, то группа редактирования должна разработать план корректировки. Здесь необходимо изменить или общий итог, чтобы отразить сумму отдельных категорий, или одну или более отдельных категорий дохода. Когда устанавливается общий доход для всех членов семьи, то семейный доход складывается из суммы отдельных доходов в семье.

7. Группа редактирования должна учитывать ситуацию, когда один или более членов семьи имеют отрицательный доход из-за неудачного бизнеса или по другим причинам. В этом случае общий семейный доход не увеличится, а уменьшится на сумму отрицательного дохода этого конкретного лица.

3. Нуклеарная семья

8. Для данных по составу домохозяйства во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда* разработан код для нуклеарной семьи, который определяется следующим образом (предложения по перекодированию даны в скобках):

1. Супружеская пара (или пара, живущая в гражданском браке) без детей (глава семьи и супруга/супруг или совместные главы семьи или пара, живущая в гражданском браке)
2. Супружеская пара (или пара, живущая в гражданском браке) с одним или несколькими не состоящими в браке детьми (как и позиция выше, однако с помощью обследования домохозяйства или перекодирования числа не состоящих в браке детей в жилищной единице или по крайней мере одного не состоящего в браке ребенка)
3. Отец с одним или несколькими не состоящими в браке детьми (мужчина — глава семьи без супруги и по крайней мере с одним не состоящим в браке ребенком, в соответствии с приведенным выше определением)
4. Мать с одним или несколькими не состоящими в браке детьми (женщина — глава семьи без супруга и по крайней мере с одним не состоящим в браке ребенком, в соответствии с приведенным выше определением).

4. Тип домохозяйства

9. Во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда* содержатся общие положения по различным типам домохозяйств для помощи в разработке повторных кодов по составу домохозяйства. Страны могут пойти по пути введения единого повторного кода или серии повторных кодов в зависимости от возможных способов использования данных.

10. Первый повторный код мог бы определять тип домохозяйства, представленный одним из указанных ниже пунктов, включая определения. Предлагаемые повторные коды приводятся в следующем разделе.

1. Домохозяйство, состоящее из одного лица
2. Нуклеарное домохозяйство: состоящее из одной нуклеарной семьи, то есть супружеской пары или партнеров в гражданском браке с ребенком/детьми или без них или одинокого родителя с ребенком/детьми или без них
3. Расширенное домохозяйство: состоящее из одной нуклеарной семьи и других лиц, имеющих родственные связи с главой домохозяйства; двух или более нуклеарных семей или двух и более лиц, имеющих родственные связи между собой, но не входящих в нуклеарную семью
4. Составное домохозяйство (другие типы домохозяйств)

5. Состав домохозяйства

11. *Домохозяйства, состоящие из одного лица*, — это домохозяйства, а не семьи, и поэтому их следует включать в отдельную категорию в перекодированную запись о составе домохозяйства.

12. *Домохозяйства, состоящие из нуклеарной семьи*. Домохозяйства, состоящие из нуклеарной семьи, можно подразделить на следующие категории (и им могут быть присвоены индивидуальные коды): 1) семьи, состоящие из супружеской пары с детьми; 2) семьи, состоящие из супружеской пары без детей; 3) пары, живущие в гражданском браке, с детьми; 4) пары, живущие в гражданском браке, без детей; 5) отцы с детьми; и 6) матери с детьми. Для определения надлежащего кода устанавливается пол главы домохозяйства, после чего домохозяйство исследуется на наличие супруги и детей. Тип домохозяйства, состоящего из

нуклеарной семьи, может отражаться двузначным кодом с цифрой 2 в качестве первого из двух знаков (код 1 резервируется для домохозяйств, состоящих из одного лица); следовательно, код 21 будет обозначать супружескую пару с детьми.

13. *Домохозяйства, состоящие из расширенной семьи.* Расширенные семьи также можно разделить на категории, которые будут включать (исходя из указанных выше обозначений): 31 — одна нуклеарная семья и другие лица, имеющие родственные связи с этой нуклеарной семьей; 32 — две и более нуклеарных семьи, имеющие родственные связи друг с другом, без каких-либо других лиц; 33 — две и более нуклеарных семьи, имеющие родственные связи друг с другом, плюс другие лица, имеющие родственные связи с этой нуклеарной семьей; и 34 — два и более лиц, имеющих родственные связи друг с другом, ни одно из которых не относится к нуклеарной семье. Фактические коды будут определяться на основе обследования домохозяйства на предмет числа нуклеарных семей и родственных связей между лицами в домохозяйстве. Если то или иное домохозяйство уже получило код как нуклеарная семья, то такая процедура проводиться не будет.

14. *Составные домохозяйства.* Все прочие домохозяйства будут являться составными домохозяйствами. Использование такой же схемы, как и раньше, даст следующие результаты: 41 — одна нуклеарная семья плюс другие лица, причем некоторые из них имеют с ней родственные связи, а некоторые не имеют; 42 — одна нуклеарная семья плюс другие лица, ни одно из которых не имеет с ней родственных связей; 43 — две и более нуклеарных семьи плюс другие лица, причем некоторые из них имеют родственные связи по крайней мере с одной нуклеарной семьей, а некоторые не имеют родственных связей ни с одной нуклеарной семьей; 44 — две и более нуклеарных семьи, имеющие родственные связи друг с другом, плюс другие лица, ни одно из которых не имеет родственных связей ни с одной нуклеарной семьей; 45 — две и более нуклеарных семьи, не имеющие родственных связей друг с другом, с другими лицами или без них; 46 — два и более лиц, имеющих родственные связи друг с другом, ни одно из которых не принадлежит к нуклеарной семье, плюс другие лица, не являющиеся родственниками; и 47 — лица, не имеющие родственных связей. И вновь необходимо провести серию обследований и обобщений, в результате которых будет определен надлежащий код для каждого типа домохозяйства.

6. Состав семьи

15. Семьи представляют собой подсовкупность домохозяйств; таким образом, повторное кодирование данных по составу семьи будет включать в себя категории, подходящие для описанных выше типов семей. Поскольку домохозяйства, состоящие из одного лица, не являются семьей, они не включаются в повторное кодирование данных по составу семьи. Аналогичным образом, составные домохозяйства представляют собой именно домохозяйства, а не семьи, поэтому они также не включаются в повторное кодирование. Отдельные страны могут затем принять решение о том, вводить единый повторный код для всех семей (нуклеарных и расширенных семей совместно) или отдельные повторные коды для нуклеарных и расширенных семей при понимании того, что такие повторные коды не должны накладываться друг на друга (хотя можно выдвинуть аргументы в пользу включения домохозяйств, состоящих из нуклеарной семьи, вместе с расширенными семьями во все семьи).

7. Статус домохозяйства и семьи

16. Статус домохозяйства и семьи указывает на то, в каком родстве то или иное лицо состоит с другими членами домохозяйства или семьи. Методы установления статуса домохозяйства и семьи отличаются от традиционного подхода к классификации членов домохозяйства исключительно в соответствии с их родственной связью с главой или основным членом домохозяйства.

17. Во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда* предлагается следующая схема кодирования статуса домохозяйства. Первый набор кодов предназначен для членов домохозяйств, содержащих по крайней мере одну нуклеарную семью (иными словами, домохозяйства, которые также являются семьями). Предлагаются следующие определения повторного кодирования:

- 1.1. Муж (мужчина — глава домохозяйства или супруг)
- 1.2. Жена (женщина — глава домохозяйства или супруга)
- 1.3. Партнер в гражданском браке или сожитель (данные взяты из кодов родства, если таковые имеются, или из комбинации кодов родства и брачного статуса)
- 1.4. Одинокая мать (определяемая исходя из факта отсутствия мужа этой женщины, но наличия детей)
- 1.5. Одинокий отец (определяемый исходя из факта отсутствия жены этого мужчины, но наличия детей)
- 1.6. Ребенок, живущий с обоими родителями (ребенок главы домохозяйства с обоими родителями в доме)
- 1.7. Ребенок, живущий с одинокой матерью (ребенок главы домохозяйства, но отец ребенка отсутствует)
- 1.8. Ребенок, живущий с одиноким отцом (ребенок главы домохозяйства, но мать ребенка отсутствует)
- 1.9. Лицо, не являющееся членом нуклеарной семьи (любой другой родственник). Во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда* этот элемент данных делится на две группы: 1) живущие с родственниками; и 2) живущие с лицами, не являющимися родственниками

18. Второй набор повторных кодов предназначен для лиц в домохозяйствах без создания какой-либо нуклеарной семьи: лиц, проживающих в одиночку, а также лиц, проживающих с другими родственниками и/или не родственниками, за исключением супруга(и) или ребенка главы домохозяйства. Предусмотрены следующие категории:

- 2.1. Лица, проживающие в одиночку (домохозяйство, состоящее из одного человека)
- 2.2. Лица, проживающие с другими лицами (лицо, проживающее в жилищной единице без супруги(а) или ребенка главы домохозяйства). Эта категория далее подразделяется на лиц, проживающих: 1) с братьями/сестрами; 2) с другими родственниками, не являющимися братьями/сестрами; или 3) с лицами без родственных связей

19. Из этих категорий необходимо разработать одну переменную, поскольку они являются взаимно исключающими. Такая переменная должна быть двузначной. Некоторые статистические управления могут сделать первый знак

независимым от второго знака; в этом случае первый знак будет указывать, является ли домохозяйство нуклеарной семьей или нет, а второй — будет определять статус того или иного лица в домохозяйстве.

20. Во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда* также изложены категории для классификации лиц по семейному положению. К ним относятся: 1) мужчина или женщина в составе проживающей в домохозяйстве супружеской пары с детьми или без детей; 2) одинокий родитель в разбивке по признаку пола; 3) ребенок главы домохозяйства, ребенок супружеской пары или ребенок одинокого родителя в разбивке по признаку пола родителя; и 4) лицо, не являющееся членом нуклеарной семьи (родственник или не родственник и если родственник, то степень родства). Указанные выше определения статуса домохозяйства могут использоваться и для статуса семьи.

8. Влияние ВИЧ/СПИДа на структуру домохозяйства

21. С учетом воздействия эпидемии ВИЧ/СПИДа на структуру домохозяйств во многих странах повторное кодирование может помочь в описании различных типов жилищных единиц. Например, повторное кодирование с описанием домохозяйств с каким-либо отсутствующим поколением (только дедушки/бабушки и внуки/внучки), домохозяйств, возглавляемых лицами моложе 18 лет, домохозяйств, возглавляемых вдовцами/вдовами и т. д. можно использовать для оценки, хотя и весьма косвенной, социально-экономических последствий этой эпидемии. Данные о детях, участвующих или не участвующих в составе рабочей силы, о структуре рабочей силы в рамках таких домохозяйств и т. д. способны помочь плановым органам правительства дать полную характеристику последствий ВИЧ/СПИДа.

9. Лица, связанные родственными отношениями

22. Связанные родственными отношениями лица — это те, кто связан какими-либо родственными отношениями с главой домохозяйства. Производная переменная для связанных родственными отношениями лиц складывается из суммы переменных для всех лиц, связанных родственными отношениями с главой домохозяйства. Этот показатель особенно важен в ситуациях, когда большое число лиц, не связанных между собой родственными отношениями, проживают совместно в жилищных единицах. Когда много лиц, не связанных родственными отношениями, проживают совместно таким способом, их зачастую классифицируют как лица, живущие в «помещениях для коллективного проживания» или «помещениях для группового проживания».

23. При составлении наборов данных национальные статистические управления часто разрабатывают производные переменные для различных групп, связанных родственными отношениями лиц в разбивке по возрасту. Например, производные переменные можно разработать для связанных родственными отношениями детей от 0 до 5 лет, детей от 5 до 17 лет, детей от 6 до 17 лет, детей от 0 до 17 лет, лиц в возрасте 65 лет и старше и лиц в возрасте 75 лет и старше.

24. Понятие «связанные родственными отношениями дети» в семье может включать, например, собственных детей главы домохозяйства и других лиц моложе 18 лет в домохозяйстве, независимо от брачного статуса, которые

связаны родственными отношениями с главой домохозяйства, за исключением супруги(а) главы домохозяйства. В категорию «связанные родственными отношениями дети» могут включаться или не включаться приемные дети, поскольку они не связаны родственными отношениями с главой домохозяйства, но это решение зависит от конкретной ситуации в стране.

10. Работники в семье

25. Иногда страны по своему усмотрению могут сопоставлять переменные показатели домохозяйств по числу работников, например показатели распределения дохода по размеру домохозяйства и числу работников на иждивенца. Страна может получать производную переменную для числа работников в семье, суммируя число лиц, которые работали по меньшей мере один час в отчетный период, например неделю или год (или календарный год, или последние 12 месяцев). Страна может использовать показатель, относящийся к числу лиц, выполняющих работу в течение «последней недели», если данные собираются только за этот период.

11. Полный комплект сантехнического оборудования

26. В вопроснике переписи есть несколько элементов, предназначенных для получения данных о сантехническом оборудовании. Эти элементы, как правило, связаны с наличием водопровода, туалета со смывом и ванной или душа и обычно собираются как по занятым, так и незанятым жилищным единицам. Производная переменная для полного комплекта сантехнического оборудования может помочь в сопоставлении социально-экономических условий между районами или группами населения в определенный момент времени или в динамике по времени. Производная переменная для полного комплекта сантехнического оборудования может быть получена, например, когда есть три вида удобств: водопровод (с горячей или холодной водой), туалет со смывом и ванна или душевая установка (в самой жилищной единице или за пределами строения, где расположена эта единица). Группа редактирования должна будет определить наиболее приемлемый набор переменных для полного комплекта сантехнического оборудования.

27. В данном примере производная переменная может быть получена, если все три элемента данных запрашиваются отдельно, и при редактировании будет определяться наличие всех трех элементов. Если в жилищной единице есть водопровод, туалет со смывом и ванна или душ, то следует указывать «имеется полный комплект сантехнического оборудования». Без всех трех элементов следует указывать «полный комплект сантехнического оборудования отсутствует».

12. Полный комплект кухонного оборудования

28. Во время переписи собираются данные о кухонном оборудовании по элементам вопросника, касающимся оборудования для приготовления пищи, холодильника и раковины; эти данные собираются как по занятым, так и незанятым жилищным единицам. Можно считать, что в данной жилищной единице есть «полный комплект кухонного оборудования», когда оборудование для приготовления пищи (электрическая, керосиновая или газовая плита, мик-

роволновая печь и стационарные горелки или кухонная плита), холодильник и раковина с водопроводной водой расположены в том же строении, в котором зарегистрированы жилые помещения. Все это не обязательно должно быть в одной комнате.

29. Производная переменная выводится, если вышеперечисленные три элемента запрашиваются отдельно, и при редактировании определяется наличие всех трех элементов. Запись «отсутствует полный комплект кухонного оборудования» включает такие условия, когда имеются в наличии все три указанных кухонных удобства, но оборудование расположено в другом строении; есть некоторые, но не все удобства; или нет ни одного из трех указанных кухонных удобств в строении, в котором зарегистрированы жилые помещения.

13. Валовая квартирная плата

30. Как правило, страны собирают данные о квартирной плате за наличный расчет или договорной квартирной плате. Из квартирной платы за наличный расчет, как правило, исключаются расходы на коммунальные услуги. Иногда странам нужны также сведения о валовой квартирной плате. Валовая квартирная плата — это договорная квартирная плата или плата за наличный расчет плюс среднемесячная стоимость коммунальных услуг (электричество, газ и вода) и топливо (включая нефтепродукты, уголь, керосин и дрова), если оплату этих услуг должен производить съемщик. Сведения о валовой квартирной плате используются для того, чтобы ликвидировать разницу, которая возникает в связи с различиями в практике в отношении включения коммунальных услуг и топлива в арендную плату. Единицу, занимаемую съемщиком без оплаты наличными, можно в таблицах показать отдельно в графе «без оплаты наличными».

31. Производная переменная для валовой квартирной платы складывается из суммы арендных платежей и оплаты за коммунальные услуги, если эти данные собираются отдельно.

14. Индекс благосостояния

32. Индекс благосостояния используется для измерения уровня благосостояния страны или частей страны. В большинстве случаев этот индекс составляется на основе активов домохозяйств. Часто для получения наилучшего набора элементов данных и их вариантов используется факторный анализ. Обычно этим элементам данных присваиваются величины двузначного кода: 1 — для записи «имеется» и 0 — для записи «отсутствует», а затем величины суммируются. Чем больше величина, тем выше уровень благосостояния. Например, наличие телевизора может кодироваться как 1, его отсутствие — 0. С другой стороны, наличие туалета может кодироваться как 1 для «внешнего туалета», 2 — для «туалета с принудительным сливом» или 3 — для «туалета с водопроводным сливом» (применяя три набора переменных в двоичной форме). В ходе суммирования разным элементам данным могут придаваться различные веса.

33. Затем можно сформировать квинтили, составляющие каждую пятую часть в распределении величин индекса благосостояния. В самый нижний квинтиль будут входить беднейшие домохозяйства, а в самый верхний квинтиль — наиболее обеспеченные домохозяйства.

В. Производные переменные для записей о населении

1. Статус экономической активности

34. Производная переменная для статуса экономической активности может оказаться весьма полезной при составлении таблиц, но она требует информации по нескольким переменным. В соответствии категориями, указанными во втором пересмотренном издании *Принципов и рекомендаций в отношении переписей населения и жилого фонда*, необходима переконфигурация нескольких переменных. Производная переменная может состоять из двух категорий, которые подразделяются на шесть подкатегорий:

1. Экономически активное население
 - 1.1. Работающее
 - 1.2. Неработающее
2. Экономически неактивное население
 - 2.1. Учащиеся
 - 2.2. Лица, ведущие домашнее хозяйство
 - 2.3. Получатели пенсионного пособия или доходов на капитал
 - 2.4. Прочие

35. Поскольку во многих связанных с этой тематикой таблицах используются различные классификации экономической активности, группа редактирования должна подумать над тем, чтобы ввести в записи данных производную переменную, вместо того чтобы обработчики данных проводили при составлении таблиц повторную классификацию статуса экономической активности. Переклассификация во время составления таблиц может привести к ошибкам, поскольку различные разработчики данных могут немного по-разному произвести эту переклассификацию; даже одна программа может произвести переклассификацию по-разному, в зависимости от конкретных требований к редактированию или составлению таблиц. Специалисты по экономическим характеристикам должны подготовить спецификации для производной переменной.

2. Собственные дети

36. Иногда страны хотят получить сведения о «собственных детях», которые являются биологическими детьми главы домохозяйства и/или супруги(а). В таблицах можно показать данные о «собственных детях» с дальнейшей разбивкой на детей, живущих с двумя родителями или только с одним родителем.

37. Производная переменная для понятия «собственные дети» может складываться из данных о суммарном числе собственных детей того или иного конкретного лица, как правило женского рода, следуя определениям, отобранной группой редактирования. Иногда пользователям нужны более подробные сведения о собственных детях в разбивке по возрасту. Например, для Соединенных Штатов производные переменные разрабатываются для числа собственных детей младше 6 лет и для детей в возрасте от 6 до 17 лет. Эти данные указываются в записях обо всех лицах женского пола. Такая информация используется, в частности, для определения характеристик лиц женского пола в составе трудоспособного населения с собственными детьми.

3. Родители в доме

38. Эти данные показывают характеристики детей в семьях с одним родителем в сопоставлении с жилищными единицами, где проживают оба родителя. Программа редактирования получает эту производную переменную, определяя, сколько родителей того или иного конкретного лица есть в доме, используя коды степени родства. Программа проверяет код степени родства для каждого ребенка и использует эту информацию в сочетании с информацией о субсемьях, чтобы определить, сколько родителей живут в данной жилищной единице.

4. Текущий год в учебном заведении

39. В некоторых странах задается два вопроса об образовании:

- a) посещает ли данное лицо в настоящее время учебное заведение;
- b) высший уровень образовательной подготовки.

40. В этих странах группы редактирования зачастую находят несоответствие между этими двумя элементами, когда какое-либо лицо фактически посещает учебное заведение во время регистрации. Иногда это приводит к тому, что высший уровень образовательной подготовки данного лица будет указываться на год меньше, чем с учетом года фактического посещения учебного заведения. Если данное лицо находится в середине ряда ступеней или уровней, то на статистику это не повлияет. Однако если это лицо посещает первую ступень из целого ряда ступеней для конкретного уровня образования, то согласование с данными из других источников может оказаться невозможным. Например, лицо, посещающее первый класс, будет зарегистрировано как посещающее учебное заведение, но без образовательной подготовки. Аналогичным образом, лицо, посещающее среднюю школу, будет зарегистрировано как посещающее учебное заведение, но уровень образовательной подготовки при этом будет соответствовать последней ступени (или уровню) начальной школы.

41. Для такого сочетания элементов данных можно разработать производную переменную под названием «текущий год в учебном заведении». Если данное лицо на текущий момент не посещает учебное заведение, то код будет таким же, как для высшего уровня образовательной подготовки. Если же данное лицо посещает учебное заведение на текущий момент, то программа редактирования добавит единицу к ступени (или уровню) образовательной подготовки и отнесет это к показателю «текущий год в учебном заведении».

42. Некоторые страны задают три вопроса об образовании, два из которых отмечены выше, а третий вопрос касается того, завершено ли обучение на высшей ступени. Если такие сведения также собираются, то их следует использовать и при определении показателя «текущий год в учебном заведении».

5. Число месяцев с даты последних родов

43. Если ведется сбор информации по вопросу о дате рождения последнего ребенка, можно создать схему перекодирования для получения косвенных оценок годовых показателей рождаемости в разбивке по возрасту и общему

уровня рождаемости. При перекодировании берется дата регистрации (обычно месяц и год) и эти данные переводятся в общее число месяцев, после этого берется дата последних родов, а затем производится вычитание для получения числа месяцев с даты последних родов. Соответствующая цифра сохраняется в записи для этой женщины с целью оказания помощи в определении годовых оценок показателей рождаемости.

Приложение II

Взаимосвязь между форматом переписного листа и вводом данных с клавиатуры

1. Двумя самыми распространенными форматами переписного листа для элементов данных о населении при проведении переписи или обследования являются персональные страницы и страницы по домохозяйствам.

2. Персональные страницы содержат одну или две титульные страницы с информацией о населении и отдельные страницы для каждого лица. Такой метод полезен потому, что все сведения, касающиеся одного лица, записываются на одной странице, что облегчает сбор данных. Кроме того, такой формат облегчает проверку на внутреннюю согласованность во время регистрации. Персональные страницы можно объединять под одним переплетом для облегчения обработки данных на местах, как показано на рисунке А.И.1.

3. Кодирование и введение информации с клавиатуры для различных элементов данных с персональных страниц — это в основном механическая операция, при которой не предполагается, что кодировщик/оператор по вводу данных будет оценивать достоверность представленной информации, а будет лишь присваивать соответствующий код и правильно нажимать на клавиши. На рисунке А.И.2 показана последовательность информации для данного лица, записанной на одной странице. Гораздо проще вводить данные с одной страницы, чем работать на клавиатуре, одновременно переворачивая страницы. Проверка на достоверность производится позднее во время компьютерного редактирования.

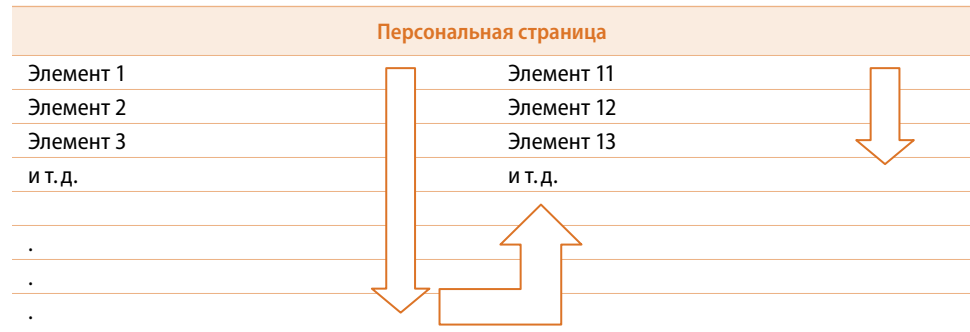
4. На страницах, посвященных домохозяйствам, содержатся все сведения о домохозяйстве по возможности на одной странице или на нескольких страницах, причем все члены домохозяйства перечисляются на каждой странице. Такой метод перечисления членов домохозяйства полезен, поскольку элементы данных переписного листа не надо печатать для каждого лица отдельно, в связи с чем экономится место. Кроме того, счетчик может сопоставить записи о членах домохозяйства после сбора данных.

Рисунок А.И.1

Образец бланка вопросника с персональными страницами

Персональная страница для лица X		Персональная страница для лица X+1	
Элемент 1	Элемент 10	Элемент 1	Элемент 10
Элемент 2	Элемент 11	Элемент 2	Элемент 11
.	.	.	.
.	.	.	.
.	.	.	.

Рисунок А.И.2

Пример последовательности информации в вопроснике с персональными страницами

5. Третий метод заключается в том, чтобы иметь отдельные бланки для каждого лица, из которых счетчик во время или после регистрации составляет буклет с вкладными страницами. Такой метод является эффективным, поскольку счетчик собирает лишь такое количество бланков (страниц), которое необходимо для данного домохозяйства. Недостаток здесь заключается в том, что при передаче или во время каких-то других операций бланки могут отсоединиться, что чревато множеством возможных проблем с точки зрения редактирования и доли обследованного материала, если бюро переписи не сможет вновь соединить их для правильного отражения данных о домохозяйстве (рисунок А.И.3).

6. Физический размер страниц вопросника также имеет значение не только для регистрации, но и для ввода данных с клавиатуры. Во время кодирования или ввода данных документ должен ровно лежать на поверхности рабочего стола, и кодировщики или операторы по вводу информации должны легко находить и обрабатывать элементы данных на бланке.

7. Когда вся информация содержится на одной странице, сотрудники могут легко вводить данные со страниц по домохозяйству, и, несомненно, это будет быстрее, поскольку вводящему данные оператору не нужно будет переворачивать страницы. На рисунке А.И.4 показана последовательность информации на странице, посвященной домохозяйству.

Рисунок А.И.3

Образец вопросника по домохозяйству с данными по всем лицам на одной странице

Страница по домохозяйству					
Лицо	Элемент 1	Элемент 2	Элемент 3	Элемент 4	и т.д.
1					
2					
3					
4					
5					
.					
.					
.					

Рисунок А.И.4

Пример последовательности для вопросника со страницами по домохозяйству, причем на одной странице содержатся данные о нескольких лицах

Страница по домохозяйству				
Лица	Элемент 1	Элемент 2	Элемент 3	и т. д.
1	→			↻
2	→			
3	→			
4				
5				
•				
•				
•				

8. Проблемы могут возникнуть при вводе с клавиатуры информации о населении или жилом фонде, которая занимает больше одной страницы. Для разрешения этой проблемы национальное статистическое управление может использовать один из двух подходов, изложенных ниже.

9. Данные об одном лице можно вводить за один раз. Оператор по вводу данных может набирать строку информации для какого-либо лица на первой странице из ряда страниц, а затем переходить на вторую и последующие страницы. Когда страницы, посвященные первому лицу, заканчиваются, оператор по вводу информации возвращается к первой из посвященных данному домохозяйству страниц и набирает данные о втором лице, потом о третьем лице и т. д. Такой способ введения данных с клавиатуры действует до тех пор, пока оператор может оставаться на нужной строке в течение всего процесса введения данных. Хотя могут быть созданы такие программы компьютерного редактирования, которые восстановят информацию, если данные о том или ином лице будут ошибочно внесены в строку, посвященную другому лицу, но подготовить такую программу очень сложно.

10. Данные на одной странице можно вводить за один раз. Оператор по вводу данных может набрать целую страницу информации, а затем перейти на следующую страницу. Здесь оператор набирает всю информацию, изложенную на первой странице, независимо от числа лиц. Затем оператор переворачивает страницу и набирает следующую часть информации в отношении всех лиц. При этом схемы пропуска тех или иных данных могут быть включены или не включены, в зависимости от способа набора данных (с компьютерным редактированием или без него). В любом случае во время компьютерного редактирования записи из различных наборов введенных данных необходимо будет собирать, а затем уже рассматривать все неправильно набранные номера рассматриваемых лиц.

11. В следующем примере (рисунок А.И.5) демографическая информация о домохозяйстве не представляет никаких особых проблем в связи с вводом данных с клавиатуры, поскольку при проведении переписи были получены ответы по всем элементам данных в отношении всех лиц.

Рисунок А.И.5

Пример страницы по домохозяйству с указанием нескольких лиц, без проблем в отношении ввода данных с клавиатуры

Страница по домохозяйству				
Лица	Степень родства	Пол	Возраст	и т.д.
1	Глава домохозяйства	М	40	
2	Супруг(а)	Ж	35	
3	Ребенок	Ж	18	
4	Ребенок	М	12	
5	Брат/сестра	М	35	
6	Брат/сестра супруга(и)	Ж	30	
7	Ребенок брата/сестры	М	5	
8	Ребенок брата/сестры	Ж	3	
и т.д.				

12. Вместе с тем вторая страница для этого же домохозяйства (рисунок А.И.6) может вызвать некоторые проблемы, связанные с вводом данных с клавиатуры. Например, если страна предпочитает собирать данные о языке по одному и тому же домохозяйству только в отношении лиц в возрасте 5 лет и старше, то вместо информации о восьмом лице (в возрасте 3 лет) будет пропуск. Оператор по вводу данных должен оставить эту ячейку для данного ребенка незаполненной, а компьютерная программа редактирования попытается позднее скорректировать данные для этой ячейки.

13. Аналогичным образом, незаполненными должны быть и другие элементы данных, например касающихся лиц, не достигших минимального возраста для участия в составе рабочей силы; женщин, не достигших минимального возраста для деторождения; и рождаемости в отношении всех мужчин. На рисунке А.И.6 показано, что оператор по вводу данных мог неправильно ввести с клавиатуры информацию о шестом лице в отношении детей, родившихся

Рисунок А.И.6

Пример страницы по домохозяйству с указанием нескольких лиц с возможными проблемами в отношении ввода данных с клавиатуры

Страница по домохозяйству 2				
Лица	Язык	Участие в составе рабочей силы	Дети, родившиеся живыми	и т.д.
1	Язык 1	Да		
2	Язык 1	Нет	3	
3	Язык 1	Нет	0	
4	Язык 1			
5	Язык 1	Да		
6	Язык 1	Нет	4	
7	Язык 1			
8				
и т.д.				

живыми (в данном случае 4), и поставить ее по ошибке в ячейку пятого лица. Компьютерная программа редактирования затем устранил данные о рождаемости мужчин и вычислит данные о рождаемости для женщин, но она может неправильно вменить данные значения.

14. Ввиду стоимости и ограниченности места страна вынуждена использовать бланк по домохозяйству множество раз. Однако если население страны небольшое или страна может позволить себе дополнительные расходы, то в персональных страницах, вероятно, будет меньше ошибок согласования из-за неправильного ввода данных с клавиатуры, чем в бланках по домохозяйству.

Приложение III

Сравнение сканирования и ввода данных с клавиатуры

1. Многие страны используют сканирующие устройства либо с оптическим распознаванием меток (ОПМ), либо с оптическим распознаванием символов (ОПС). Каждый из этих методов имеет свои преимущества по сравнению с вводом данных с клавиатуры, если операция проходит без каких-либо проблем, эффективна и если затраты при этом небольшие. Однако многие страны, даже те, которые доверяют сканированию, не могут себе позволить первоначальные затраты или постоянные расходы по обслуживанию во время и после регистрации. В позитивном плане многие страны используют сканеры, приобретенные для выполнения ряда других задач в ходе проведения переписи, включая другие обследования и ведение таких административных записей, как бланки ввода и вывода информации. Страны могут также пойти по пути передачи во внешний подряд операций сканирования или взять в аренду сканеры на тот период, на который они требуются для проведения переписи.

2. Одно из преимуществ ввода данных с клавиатуры заключается в том, что навыки, приобретенные во время ввода данных с клавиатуры, используются при выполнении других работ в национальных бюро переписи/статистических управлениях и других государственных ведомствах. После того как операторы по вводу данных приобретут соответствующие навыки при проведении переписи, они работают с данными и для различных последующих обследований. К таким обследованиям относятся послерегистрационные (ПРО) и другие обследования, например обследования по вопросам рождаемости или доходов и расходов домохозяйств. Сотрудники могут также вводить с клавиатуры административные записи, такие как записи актов гражданского состояния и записи, касающиеся торговли, иммиграции и эмиграции и таможенных процедур.

А. Ввод данных

1. Сканирование

3. Страны, использующие оптические или другие сканирующие устройства для сбора данных, как правило, их после этого не корректируют, хотя изменения могут зависеть от встроенных в систему схем пропуска тех или иных элементов данных. Однако у стран, предпочитающих вводить данные с клавиатуры, есть возможности выбора в зависимости от того, как быстро им нужно ввести данные и какой потребуется объем ручных проверок. Каждая возможность зависит от требований групп редактирования, навыков операторов по вводу данных и степени сложности программы редактирования.

4. Количество и тип требуемого оборудования по вводу данных зависят от выбранного метода сбора данных, времени, выделенного для этой фазы пере-

писи, размеров страны, степени децентрализации операций по сбору данных и ряда других факторов. При вводе данных с клавиатуры средние темпы ввода варьируются обычно в пределах от 5 до 10 тысяч ударов в час. При этом одни операторы работают с гораздо меньшей скоростью по сравнению с ее указанным диапазоном, а другие значительно превышают этот диапазон. К числу факторов, которые влияют на скорость работы оператора, относятся: *a)* вспомогательная программа и программное обеспечение; *b)* сложность задач оператора; *c)* эргономические характеристики, надежность и скорость срабатывания оборудования; *d)* степень загрузки оператора; *e)* уровень подготовки и навыки нанятого персонала; и *f)* мотивация работников (United Nations, 2008, para. 1.193).

2. Ввод данных с клавиатуры

5. Существуют две формы ввода данных с клавиатуры в режиме «heads down». При использовании первой формы вводятся все встречающиеся элементы данных без каких-либо схем пропуска элементов данных. В этом случае ввод данных с клавиатуры происходит быстрее, поскольку операторы по вводу данных не прекращают свою работу, когда встречаются недостоверные или несогласующиеся сведения. Возможно, такая форма и более точная, поскольку операторы выполняют свою работу более механически. При использовании второй формы ввода данных с клавиатуры в режиме «heads down» они вынуждены останавливаться для проверки недостоверных или несогласующихся ответов в переписных листах, поэтому процесс идет более медленно и требует участия гораздо более опытных операторов. С точки зрения скорости цена здесь очень высокая и требует серьезного рассмотрения. Как это ни парадоксально, при таком методе точность может даже повыситься, если оператор увидит, что данные были внесены правильно, но им был присвоен неправильный код. Неправильный ввод данных с клавиатуры сам по себе может быть иногда незамедлительно поставлен под сомнение, поскольку пакет редакторских программ обеспечивает автоматическую проверку.

a) Ввод данных в режиме «heads down» без схем пропуска элементов

6. Когда все данные вводятся или пропускаются вручную, можно поддерживать определенный ритм работы, и некоторые схемы пропуска элементов не обойдут достоверные, но временно несогласующиеся сведения. Например, если то или иное лицо зарегистрировано в качестве мужчины, большинство групп редактирования потребует пропустить весь раздел о рождаемости. В этом случае оператор по вводу данных будет набирать сквозным методом (используя клавишу пробела или стрелку, чтобы пройти через запись, касающуюся мужчины или молодой девушки), поскольку все поля будут пустыми. Однако это требует времени, и промежуток может быть не совсем точным. Например, оператор по вводу данных может пройти слишком далеко или, наоборот, не дойти до нужного места, и тогда другие элементы могут быть введены неправильно, поскольку будут выстроены не так, как нужно. Если все поля вводятся с клавиатуры таким образом, то тогда эту информацию можно вводить, не имея схем пропуска элементов. Например, если оператор встречает данные о рождаемости в отношении взрослой женщины (женщина, в отношении которой были собраны и закодированы такие данные, как дети, родившиеся живыми, дети, живущие на момент переписи или дети, родившиеся за последний год), то вводятся все элементы. Если вводится информация о рождаемости, то компьютерная программа редак-

тирования может определить, какой элемент или набор элементов достоверен, а какой необходимо изменить. Если программа редактирования определяет, что данным лицом является взрослая женщина, но информация о рождаемости отсутствует, то в этом случае следует применить метод динамического вменения или другие соответствующие средства, для того чтобы получить информацию о рождаемости для составления таблиц. Если фактическая информация утеряна из-за схем пропуска данных, то группа редактирования должна решить, компенсирует ли эта потеря повышение эффективности и скорости. При наличии схем пропуска элементов операторы по вводу данных могут, кроме того, идти назад по экрану до соответствующей позиции, требующей исправления. Хотя операторы по вводу данных теряют какое-то время, проходя через элементы, которые не вводятся с клавиатуры, при такой форме ввода данных несоответствие между данными, касающимися пола, возраста и рождаемости, лучше устранять во время редактирования, чем во время ввода с клавиатуры.

b) Ввод данных в режиме «heads down» со схемами пропуска элементов

7. Второй метод ввода данных с клавиатуры в режиме «heads down» предусматривает наличие схем пропуска. И вновь, если группа редактирования настаивает на применении схем пропуска, как правило, для того, чтобы представить, как счетчики собирали эти данные, то клавиатурный ввод будет проходить легче и быстрее, если схемы пропуска элементов будут удобны в обращении и если операторы по вводу данных быстро освоят схемы ввода. Если схемы пропуска очень сложные, операторы по вводу данных могут запутаться и постоянно вводить данные неправильно. Ввод со схемами пропуска элементов будет проходить наиболее эффективно, если использовать ограниченные схемы, которые покрывают большие разделы вводимых записей.

8. Группа редактирования должна определить соответствующие схемы пропуска тех или иных элементов для проведения переписи или обследования у себя в стране. Например, имеет смысл пропустить все данные о занятости для детей, то есть для лиц младше определенного в стране возраста для потенциальной занятости. Зачастую эти сведения составляют половину элементов данных о населении, поэтому целесообразно пропускать их в отношении детей, за исключением особых ситуаций, например, когда возраст детей уже приближается к граничным показателям или когда страна может быть заинтересована в детском труде.

9. Группа редактирования решает по каждому элементу отдельно, какие из них будут включаться в те или иные возрастные группы. Сотрудники могут сгруппировать эти элементы, с тем чтобы легче было пользоваться схемами пропуска конкретных элементов.

10. Не всегда просто принять четкие решения относительно схем пропуска. Например, рассмотрим следующую последовательность:

1. Каково гражданство данного лица?
 - Родился в этой стране (пропустить до элемента 3)
 - Натурализованный
 - Не гражданин
2. Каков год прибытия данного лица?
3. СЛЕДУЮЩИЙ ЭЛЕМЕНТ

11. Схему пропуска можно составить, чтобы переходить от 1 к 3, то есть пропускать элемент о годе прибытия для лиц, родившихся в данной стране. Однако иногда операторы по вводу данных нарушают схему пропуска или потому, что счетчик или кодировщик делает ошибку, или из-за неправильного ввода данных с клавиатуры. Это связано со многими факторами, включая уровень мастерства операторов по вводу данных, уровень культуры, компоновку переписного листа и формат экранов. Группы редактирования часто работают вместе, чтобы определить, обоснованной ли является схема пропуска элементов в таких случаях, как этот.

3. Интерактивный ввод данных с клавиатуры

12. Интерактивный ввод данных с клавиатуры можно использовать при проведении переписи, но более целесообразно применять этот метод при проведении обследований, особенно небольших обследований, когда выделенные элементы могут повлиять на результаты обследования. Интерактивный ввод данных с клавиатуры может включать ручную или автоматическую корректировку в зависимости от наличия информации, которую можно использовать для внесения изменений или корректировок.

13. Рассмотрим пример небольшого обследования. Для небольших обследований значение имеет каждый ответ. Если, например, страна проводит однопроцентное выборочное обследование, то каждый ответ представляет сто человек, жилищных единиц или сельскохозяйственных владений. Несколько недостоверных или несогласующихся ответов могут существенно повлиять на результаты данного обследования. В этих случаях демографы и другие специалисты в области общественных наук, как правило, хотят вести серьезный контроль над обработкой данных.

14. Контроль можно устанавливать по-разному. Демографы и другие специалисты могут сами вводить данные с клавиатуры, проверяя посторонние, недостоверные или несогласующиеся ответы по мере их появления, используя информацию, записанную на бланках для сбора данных. Они могут, зачастую незамедлительно, разрешить противоречия, случаи неправильного присвоения кодов или другие несоответствия при рассмотрении непосредственно собранной информации. Иногда они могут отправить назад сотрудникам на местах не полностью заполненные или недостоверные переписные листы. Такой вид интерактивного ввода данных с клавиатуры дает наилучшие результаты, поскольку демограф работает и как оператор по вводу данных, но этот метод намного дороже, и не так много стран могут себе его позволить.

15. Группы редактирования могут разработать очень подробные правила редактирования для определения того, что должны делать операторы по вводу данных в каждом конкретном случае, который возникает во время набора с клавиатуры. В отношении каждого недостоверного кода, если этот вопрос не разрешается, они могут решить, что должен вводить оператор с клавиатуры. Группа редактирования может рассматривать случаи, которые не охватываются подробными правилами, а также видоизменять правила (хотя при этом возникает опасность возникновения несоответствий между первой частью и последующими частями процесса ввода данных с клавиатуры).

16. Схемы пропуска элементов данных, которые играют важную роль при вводе данных с клавиатуры в режиме «heads down» здесь также имеют важ-

ное значение. Как и при вводе в режиме «heads down», операторы должны знать и освоить все схемы пропуска элементов, которые применяются в данном случае. Как упоминалось выше, схемы пропуска могут увеличить скорость ввода данных с клавиатуры, но обычно с некоторой потерей качества. Для интерактивного ввода данных с клавиатуры общий практический подход заключается в том, что чем меньше пропусков, тем выше качество.

17. Разработав инструкции по вводу данных с клавиатуры, национальные бюро переписи/статистические управления должны организовать реальную проверку инструкций по вводу данных с клавиатуры операторами перед принятием решения о практической работе, независимо от того, используется или нет ввод данных в режиме «heads down». После проверки инструкций по вводу данных из системы можно устранить дефекты и добиться оптимального режима ввода данных с клавиатуры.

В. Проверка данных

18. Национальное бюро переписи/статистическое управление должно также решить вопрос о том, какой уровень проверки является достаточным. Для вводимых с клавиатуры данных многие эксперты рекомендуют проводить стопроцентную проверку. В этом случае все элементы данных снова вводятся с клавиатуры (или вводятся поверх имеющейся информации) для того, чтобы убедиться, что собранные данные — это те данные, которые вводятся в машину для компьютерной обработки. Однако во многих случаях общая проверка практически неосуществима, поскольку у определенной страны нет времени вновь вводить с клавиатуры все данные или не хватает для этого финансовых или людских ресурсов. Процентная доля проверяемой выборки должна быть выше для начинающих операторов, работающих на клавиатуре, но ниже для более опытных операторов. Кроме того, если проверенный коэффициент ошибок очень низок и операторы по вводу данных совершают весьма незначительное количество ошибок, то, вероятно, полная проверка и не обязательна.

19. В рамках любой процедуры проверки прежде всего нужно определить, какая информация необходима. Желает ли та или иная страна отслеживать работу отдельных операторов или группы операторов? Желает ли страна определить, какие навыки приобретаются или сохраняются? Важными также могут быть единицы учета, включая ежедневную, еженедельную или ежемесячную отчетность и т. д. для определения рабочего процесса и приобретаемых навыков.

20. И наконец, чрезвычайно важно, чтобы проверка была независимой и чтобы проверку данных при вводе выполняли различные группы операторов или по крайней мере различные члены одной группы. Использование различных групп операторов позволяет добиться независимости в ходе процедур проверки и, таким образом, лучших результатов.

21. В отношении отсканированных данных некоторые страны также осуществляют процедуру проверки для обеспечения того, чтобы сканирование было всеобъемлющим и полным. Поскольку технология сканирования до сих пор является достаточно новой, даже после тщательного тестирования систем с помощью пилотных или предварительно проверенных данных такие факторы, как изменение качества бумаги, фактическая печать бланков в различных мес-

тах, хранение материалов и т. д., могут создать проблемы, которые необходимо будет решать с помощью проверки.

22. Если ошибки носят систематический характер и могут устраняться с помощью программы редактирования, операторы по вводу данных и контролеры не должны принимать решения по поводу корректировки информации. При этом, однако, операторы по вводу данных и контролеры несут ответственность за поиск ошибок. Эти ошибки могут привести к неадекватному тестированию сканирующего оборудования, что станет причиной систематических ошибок по некоторым элементам или комбинациям элементов данных, путаницы в считывании некоторых знаков (например, подмена 2s на 3s или 8s на 9s), неправильное считывание последовательности «галочек» в нужных ячейках и т. д.

23. Неправильное считывание последовательности «галочек» в последние годы было постоянной проблемой, причем решить ее с помощью редактирования можно лишь в некоторых случаях. Если бланки не имеют непрерывной структуры заполнения, то для решения возникающих вопросов требуются иные процедуры, вероятнее всего в ходе редактирования структуры данных. Как отмечалось выше, необходимо создать полностью проверенный структурированный файл до начала проверки содержательной части данных.

1. Зависимая проверка

24. Методы проверки могут быть либо зависимыми, либо независимыми. При зависимой проверке операторы вводят данные с клавиатуры, уже введенные ранее таким же способом другими сотрудниками. Если количество нажатий клавиш разное, то программный пакет сообщает об этом оператору по вводу данных, и в зависимости от программы оператор или корректирует предыдущие данные, или отмечает несоответствие. Поскольку данные вводятся с клавиатуры на основе оригинальных переписных листов, то, как правило, оператор может сам принять информированное решение о том, было ли первоначальное введение данных ошибочным.

2. Независимая проверка

25. При независимой проверке операторы могут начать повторное введение данных с нуля; они создают полностью независимый файл вводимых данных, используя первоначальные переписные листы. Два получившихся в результате файла — набор первоначально введенных данных и набор данных, введенных при проверке, — затем сопоставляются между собой с помощью компьютерной программы, для того чтобы проверить их на наличие несоответствий. Предположительно, при этом могут использоваться некоторые ручные операции для исправления недостоверных или несогласующихся нажатий клавиш.

С. Соображения по вопросам редактирования отсканированных данных

26. В настоящее время все больше и больше стран сканируют свои данные. В начале 2000-х годов заинтересованные лица во многих из этих стран с удивлением узнали, что сканирование дает такие виды ошибок, которые не возникают в процессе ввода данных с клавиатуры. Частично проблема с редактированием отсканированных данных связана с отсутствием контроля качест-

ва в ходе процесса сканирования. Из-за новизны технологии в начале 2000-х годов многие статистические управления не располагали опытом или оборудованием, необходимым для разработки надлежащих процедур контроля для всех элементов данных. Многие страны, которые действительно разработали соответствующие процедуры контроля качества, в конечном счете не сформировали их в отношении всех элементов; вследствие этого некоторые элементы данных в конце переписного листа — в частности, элементы по рождаемости — оказались недостоверными или несогласующимися.

27. Конечно, многие несоответствия, которые можно обнаружить во введенных с клавиатуры данных, случаются и с отсканированными данными. Тем не менее полезно обсудить некоторые специфические проблемы, возникающие в связи с использованием отсканированных данных. Поскольку считываемые сканером переписные листы требуют особых отметок для помощи в их машинном считывании, зачастую элементы данных расположены таким образом, что у регистраторов и респондентов могут возникнуть проблемы в ходе сбора данных. Проблемы, связанные с такими элементами, необходимо решать на системной основе. Когда те или иные элементы данных тесно связаны с другими элементами, например связь вероисповедания с этнической принадлежностью, то можно использовать обычные схемы редактирования, описанные выше.

28. При этом, однако, следует проявлять особую осторожность, когда проблемы могут возникать в тех элементах данных, которые необходимы для планирования и политики. Обычно элемент по признаку пола не создает проблем, поскольку в этом случае могут быть только две возможности. Однако, как отмечено выше, хотя оператор ввода данных с клавиатуры обычно ограничен вводом только величин 1 или 2 (или кода для ответа «неизвестно»), в графах по полу может появиться любая величина в виде других знаков, букв или иных символов. Вследствие этого необходимо добавить некоторую схему редактирования к тому редактированию, которое уже один раз осуществлялось в отношении введенных с клавиатуры данных для учета таких различных величин.

29. Хорошей иллюстрацией этой проблемы являются коды степени родства. Если для кодов степени родства используется один знак, как показано выше в тексте, обычно никаких проблем не возникает. Однако при использовании двузначных кодов иногда в ходе сканирования возникает проблема, когда первый знак или закодирован, или считан сканером неправильно. Обычно при использовании кодов от 1 до 12 оператор будет ограничен введением в клавиатуры только этих кодов, и программный пакет ввода обнаружит ввод неправильного кода. При сканировании будет приниматься практически любая величина (хотя пакеты сканирования можно запрограммировать также и для выявления неправильных кодов). В таких случаях ошибочные коды должны быть изменены в процессе редактирования; в противном случае они могут стать причиной самых разных проблем на этапе составления таблиц.

30. Иногда проблемой становится элемент данных по возрасту, особенно при использовании трех граф (чтобы предусмотреть возможность указания возраста для людей старше 100 лет); в этом случае может понадобиться анализ по каждому разряду, то есть такой анализ, который позволяет отдельно рассматривать разряды единиц, десятков и сотен для достижения надлежащих результатов редактирования. После того как установлено, что возраст введен надлежащим образом, можно использовать обычные схемы редактирования.

31. При этом, однако, в случаях, когда при наличии данных и о возрасте, и о дате рождения вводящая в заблуждение информация может вызвать проблемы, если один из этих элементов начинает превалировать над другими элементами. Как правило, специалисты по содержательной части переписи предпочитают использовать дату рождения совместно с датой проведения переписи или обследования для получения (путем вычитания) точного возраста для его сравнения с заявленным возрастом. В случае отсутствия одной или нескольких цифр необходимо обеспечить, чтобы все остальные цифры использовались надлежащим образом для получения наилучшей оценки вычисленного возраста для целей сравнения. Когда сканер, например, не распознает единственную цифру, при редактировании необходимо учитывать это для предоставления наиболее точной оценки того, какая цифра была пропущена. Такого рода проблемы обычно не возникают при вводе данных с клавиатуры.

32. Элементы данных, с которыми в начале 2000-х годов возникали наиболее серьезные проблемы, имели отношение к данным по рождаемости — как число детей, родившихся живыми, и детей, живущих на момент переписи, так и детей, рожденных за последний год или иной период. Для большинства стран основной проблемой было отсутствие контроля качества в процессе сканирования, в результате чего при вводе данных появлялись странные показатели. Например, если в той или иной стране по этому элементу были представлены такие данные, как 17, 18 или 19 умерших детей женского пола, эти данные, при отсутствии их редактирования, бесполезны для целей планирования.

33. В сканированных данных могут возникнуть проблемы также с информацией по смертности. При вводе данных с клавиатуры, если существует статистический ряд элементов данных по смертности, например за год до переписи (пол и возраст умерших лиц вне зависимости от того, умерли ли данные лица от естественных причин или имела место смерть матери вследствие родов), ввод данных с клавиатуры осуществляется даже в отношении стертых и вычеркнутых данных. Однако в случае со сканированием стертые данные обычно не считываются, и сканер оставит эти элементы незаполненными и продолжит считывание данных. Программа редактирования должна помещать информацию в надлежащие графы для составления таблиц и последующего анализа. Следует отметить, что более современные технологии сканирования могут осуществлять такое перемещение данных как в течение, так и непосредственно после считывания.

D. Выводы

34. Поскольку проблемы каждой страны, к сожалению, зависят от конкретных особенностей программирования и функционирования отдельных сканеров, возникают сложности с разработкой общих руководящих принципов. Однако практически во всех рассмотренных на данный момент случаях проблемы сканирования носили систематический характер; иными словами, когда специалисты разработали алгоритм, необходимый для решения таких проблем, появилась возможность составления полностью отредактированных наборов данных.

Приложение IV

Образцы рабочих блок-схем

1. Одной из задач группы редактирования является разработка реляционной структуры для переменных, используемых в процессе редактирования. Блок-схемы облегчают выявление различных связей между переменными и помогают в разработке ясных и сжатых спецификаций для редактирования. Эти спецификации для реляционных связей помогают специалистам по содержательной части переписи и специалистам по обработке данных отчетливо представлять себе процесс редактирования и способствуют взаимодействию между этими двумя группами специалистов.

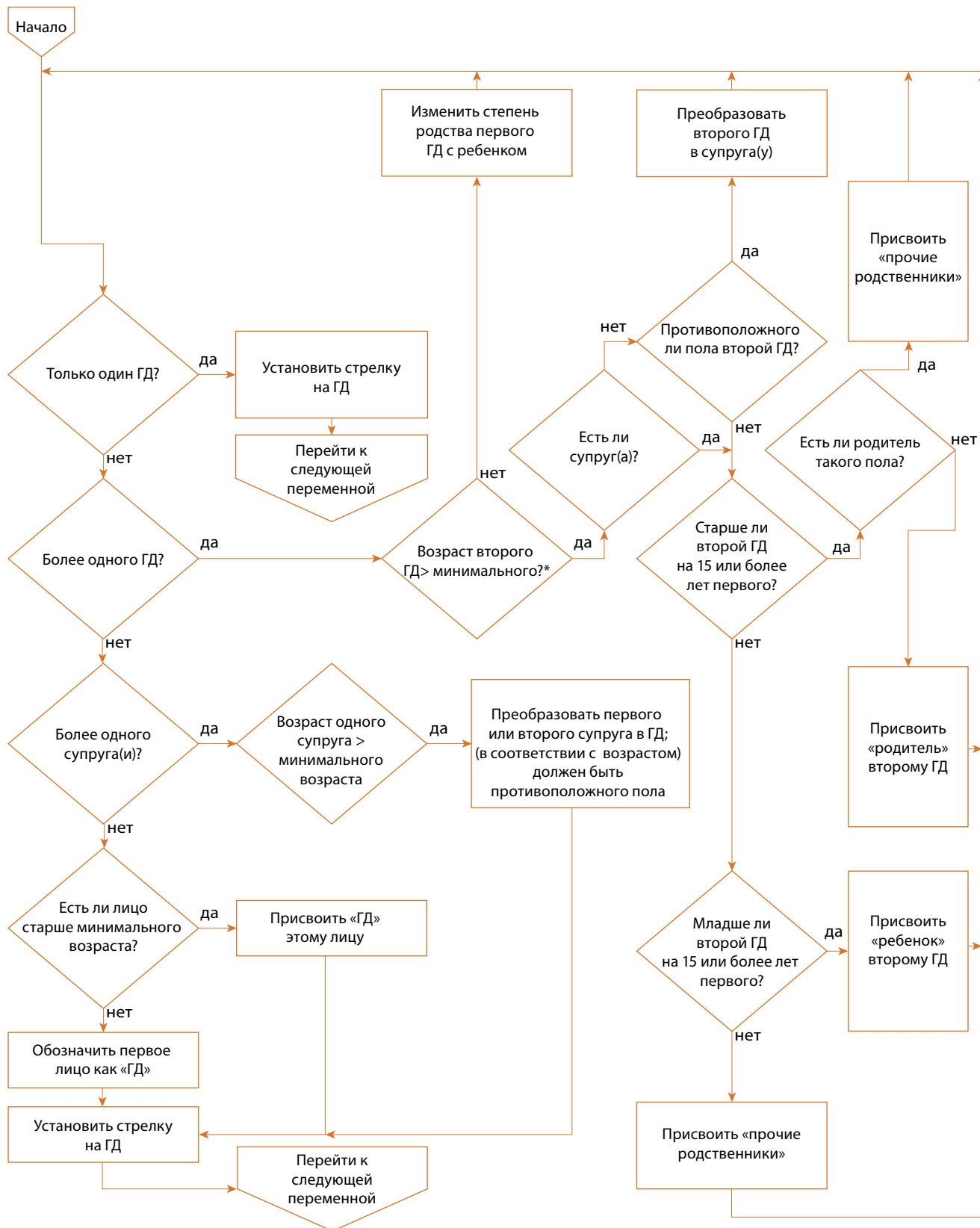
2. На следующих страницах представлены три образца блок-схем:

- a) блок-схема для определения главы домохозяйства;
- b) блок-схема для определения супруги (супруга) в домохозяйстве;
- c) блок-схема для редактирования переменной, относящейся к полу главы домохозяйства и супруги (супруга).

Эти образцы блок-схем представлены лишь для целей иллюстрации и к ним следует относиться соответствующим образом. Группа редактирования может по мере необходимости вносить дополнительные изменения в такие образцы блок-схем, исходя из ситуации в стране.

3. Блок-схемы для редактирования должны быть составлены для каждой переменной в переписи. Группа редактирования должна разрабатывать блок-схемы совместно, а специалисты по обработке данных должны использовать их вместе со спецификациями по редактированию в целях разработки компьютерных программ для редактирования данных переписи. Блок-схемы и спецификации для редактирования должны надлежащим образом документироваться, с тем чтобы их можно было использовать при обработке данных в будущих переписях и обследованиях.

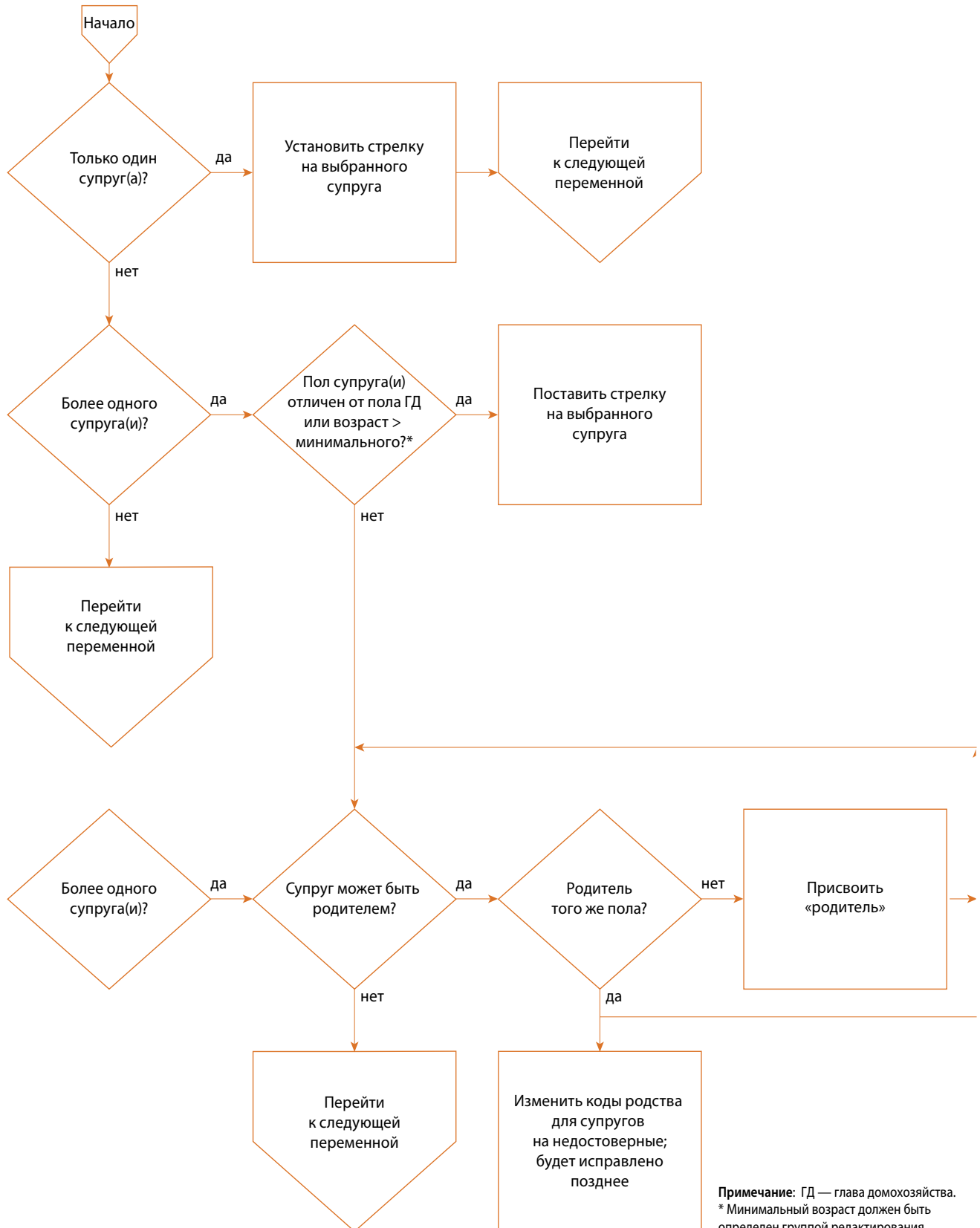
Рисунок А.IV.1
Образец блок-схемы для определения главы домохозяйства (ГД)



Примечание: ГД — глава домохозяйства.

* Минимальный возраст должен быть определен группой редактирования.

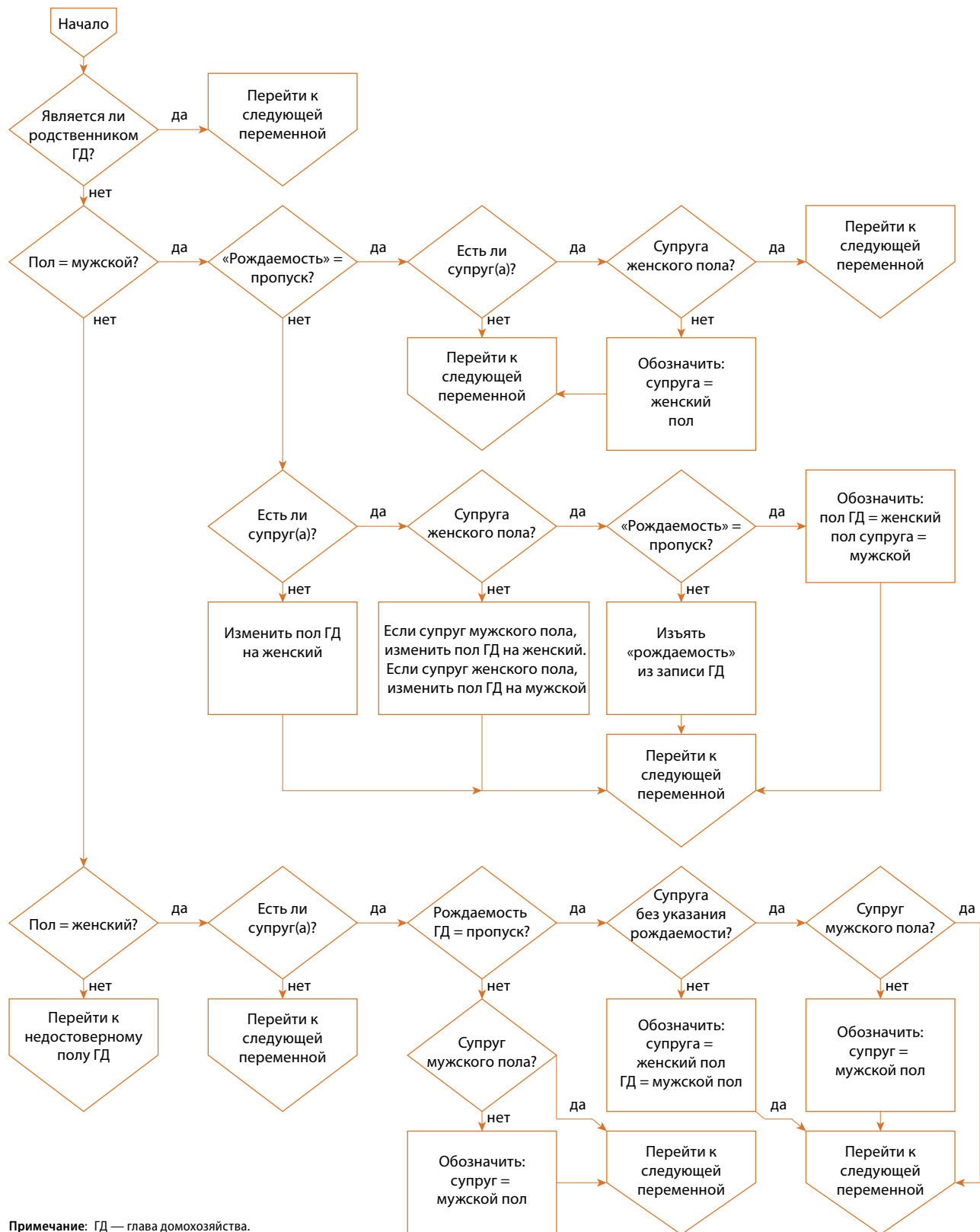
Рисунок А.IV.2
 Образец блок-схемы для определения супруги (супруга) в домохозяйстве



Примечание: ГД — глава домохозяйства.
 * Минимальный возраст должен быть определен группой редактирования.

Рисунок А.IV.3

Образец блок-схемы для редактирования переменной для пола главы домохозяйства и супруги (супруга)



Примечание: ГД — глава домохозяйства.

Приложение V

Методы вменения

1. На данный момент разработан целый ряд методов вменения. Большинство методов, описанных ниже, рассматриваются в работах Кальтона и Каспржыка (Kalton and Kasprzyk, 1982, 1986); Санде (Sande, 1982); и Серндаля, Свенссона и Вретмана (Särndal, Swensson and Wretman, 1992).

2. Методы вменения можно подразделить на стохастические или детерминистические в зависимости от степени случайности в исчисляемых данных.

3. Детерминистические методы вменения включают дедуктивное вменение; методы вменения, основанные на использовании моделей, такие как вменение по средним значениям и вменение по уравнению регрессии; и (в необходимых случаях) вменение по ближайшим соседним значениям.

4. Метод дедуктивного вменения позволяет вывести с определенностью недостающую или несогласующуюся величину. В его основе зачастую лежит схема ответов, даваемых на другие вопросы в переписном листе.

5. В большинстве случаев метод вменения применяется для замены величины, которая не выглядит достоверной. Некоторые общие процедуры вменения излагаются ниже.

6. За исключением алгоритмов динамического вменения по единому донору, описанные ниже методы связаны с вменением одного элемента данных за один раз. Так, в каждом классе вменения содержащиеся в записи элементы данных рассматриваются последовательно один за другим. Как правило, при этом рассматриваются только те подлежащие редактированию сведения, которые относятся исключительно к данному элементу или к небольшому набору тесно связанных переменных. Поскольку могут быть как явные, так и скрытые редакционные изменения, увязывающие данный(е) элемент(ы) с другими элементами, которые предстоит рассмотреть позднее, эта процедура может дать вмененную величину при обработке тех показателей, которые редактируются в данный момент, без редактирования других показателей, которые должны рассматриваться позднее. Только когда учтен весь набор редактируемых показателей, включая все подразумеваемые показатели, можно быть уверенным в том, что вмененные величины пройдут все этапы редактирования. Подразумеваемый показатель, подлежащий редактированию, — это величина, которую можно вывести путем логического совмещения двух или более явно редактируемых величин.

7. В последующих описаниях понятие «записи, прошедшие редактирование» относится к тем записям, которые прошли все этапы редактирования данного элемента (элементов). Понятие «записи, не прошедшие редактирование» относятся к тем записям, которые не прошли по крайней мере один этап редактирования в связи с рассматриваемым элементом (элементами) данных.

8. При вменении по общим средним значениям среднее значение того или иного элемента для прошедших редактирование записей присваивается всем недостающим или несогласующимся элементам данных во всех записях, отредактировать которые не удалось. При таком методе можно получать приемлемые точечные оценки, но он менее привлекателен, если оценки дисперсии вычисляются с использованием стандартной формулы оценки дисперсии. Оценки дисперсии могут быть серьезно занижены, если только коэффициент вменения не является очень низким или если не изменена формула оценки дисперсии для учета произведенного вменения.

9. При вменении по средним значениям в разбивке на классы используются классы вменения, определяемые для создания групп записей с определенной степенью сходства. В каждом классе среднее значение того или иного элемента данных по прошедшим редактирование записям исчисляется для недостающего или несогласующегося элемента данных по всем записям, отредактировать которые не удалось. Этот метод очень похож на метод вменения по общим средним значениям, однако последствия для распределения данных и проблемы с оценками дисперсии здесь, скорее всего, будут менее серьезными.

10. При вменении по уравнению регрессии или, в более общем смысле, вменении, основанном на моделях, используются данные из прошедших редактирование записей для вычисления путем регрессии конкретной переменной, которую необходимо исчислить по ряду предикторных переменных. Предикторами в регрессии могут быть элементы данных из переписного листа или вспомогательные переменные. Затем используется уравнение регрессии с целью вменения значений для недостающих или несогласующихся значений элементов данных. Это особый случай вменения на основе моделей. Такой метод часто используется в отношении непрерывных переменных при обследовании деловой активности, когда данные о предыдущем событии зачастую могут на вполне удовлетворительном уровне предсказать значения, касающиеся текущего события.

11. При вменении по ближайшим соседним значениям или подборе по функции расстояния значение элемента данных для не прошедшей редактирование записи присваивается по «ближайшей» записи, прошедшей редактирование, где понятие «ближайшая» определяется с использованием функции расстояния исходя из других известных переменных. Этот метод можно применять в классах вменения. Он обычно считается приемлемым для непрерывных переменных, но его можно применять и с нечисловыми переменными.

12. Методы стохастического вменения включают использование уравнений регрессии или любого другого детерминистического метода с прибавлением случайных остатков и использованием методов динамического («hot deck») или статического («cold deck») вменения.

13. Для каждого детерминистического метода есть свой стохастический аналог. Этого можно добиться, прибавляя случайный остаток от соответствующего распределения к исчисленному значению из детерминистического вменения. Такая процедура поможет лучше сохранить структуру частотности файла данных. Кальтон и Каспржык (Kalton and Kasprzyk, 1986) рассматривают некоторые подходы к применению такого метода.

14. При вменении методами «hot deck» и «cold deck» делается попытка создать более реалистичную вариабельность исчисляемых значений, чем это

можно сделать при детерминистических методах. В ходе процедур вменения «hot deck» недостающие или несогласующиеся величины заменяются значениями, отобранными (произвольно) из прошедших редактирование записей текущего обследования или переписи. В ходе процедур вменения «cold deck» исчисление данных производится на основе других источников, зачастую по прошлым данным, таким как более ранние события того же обследования или переписи. Есть целый ряд различных форм вменения «hot deck» и «cold deck».

15. Самой простой формой вменения «hot deck» является случайное общее вменение. Для каждой не прошедшей редактирование записи случайным образом выбирается одна прошедшая редактирование запись из всех записей такого рода, и ее сообщенное значение для рассматриваемого элемента данных вменяется вместо не прошедшей редактирование записи.

16. При случайном вменении в рамках отдельных классов для ограничения случайного отбора донорских записей в набор данных, который имеет некоторое сходство с записью, требующей исчисления, вновь используются классы вменения.

17. При последовательном вменении «hot deck» также используются классы вменения, и преимущество этого метода заключается в том, что одного прохода через файл данных достаточно для завершения процесса вменения. Процедура начинается со значения «cold deck» для каждого класса вменения, и записи в файле данных рассматриваются по очереди. При обнаружении прошедшей редактирование записи ее значение для рассматриваемого элемента данных заменяет хранящееся в памяти значение для определенного класса вменения. При обнаружении не прошедшей редактирование записи ее недостающее или несогласующееся значение заменяется хранящейся в памяти величиной. Число классов вменения не может быть чрезмерно большим, поскольку необходимо, чтобы доноры имелись в каждом классе вменения. Если записи в файле данных расположены в случайном порядке, этот метод будет почти эквивалентен случайному вменению в границах классов. Недостаток такой процедуры заключается в том, что она зачастую приводит к многократному использованию доноров и может отрицательно повлиять на оценки распределения элементов данных и дисперсии.

18. Иерархическое вменение «hot deck» — это усиление последовательного вменения «hot deck», в процессе которого используется большое количество классов вменения. Когда в первоначальном классе вменения донора найти невозможно, классы свертываются в иерархическом порядке до тех пор, пока донор не находится.

19. Задача алгоритмов вменения «hot deck» с использованием одного донора заключается в том, чтобы исчислять данные для не прошедшей редактирование записи по одному донору. Следовательно, они предусматривают совместное исчисление всех значений элемента данных в записи, которая при редактировании была определена как сомнительная. На практике зачастую ставится цель использовать одного донора для каждого раздела тесно связанных переменных в одной записи. Существенное преимущество такого подхода заключается в том, что он позволяет лучше сохранять не только маргинальные распределения, как в случае вышеупомянутых методов вменения «hot deck», но и плотность совместных распределений. Еще одно преимущество методов вменения «hot deck» по одному донору заключается в том, что при их использо-

вании становится меньше проблем с вменением величин, не поддающихся редактированию в последующих разделах переменных. С точки зрения методов вменения «hot deck» с использованием одного донора прошедшая редактирование запись — это запись, которая прошла все этапы редактирования по отношению к данному разделу. Не прошедшая редактирование запись — это запись, которая не прошла по крайней мере один из этих этапов.

20. Метод вменения и редактирования Фелледжи–Холта (Fellegi and Holt, 1976) учитывает одновременно все приемы редактирования. Главная особенность метода вменения и редактирования Фелледжи–Холта заключается в том, что правила вменения выводятся из соответствующих приемов редактирования без явной спецификации. Что касается каждой не прошедшей редактирование записи, то она сначала проходит через этап локализации ошибки с определением минимального набора переменных для вменения, а также приемлемый(е) диапазон(ы) величин для вменения, а затем производится сама процедура исчисления. В большинстве операций единый донор выбирается из прошедших редактирование записей путем согласования на основе других переменных, используемых программой редактирования, но не требующих вменения. При такой методике отыскивается один точный элемент согласования и ее можно расширить, с тем чтобы учесть другие переменные, не задействованные явным образом при редактировании. Иногда подходящего донора найти не удается, и тогда необходимо использовать метод вменения по умолчанию.

21. Метод вменения по ближайшему соседу (МВБС) (Bankier and others, 1996; Bankier, Lachance and Poirier, 1999) похож на метод Фелледжи–Холта тем, что в нем учитываются одновременно все этапы редактирования, не уточняются конкретно приемы вменения и исчисление данных происходит по одному донору. При этом для каждой не прошедшей редактирование записи определяются действия по вменению с минимальными изменениями в зависимости от имеющихся потенциальных доноров. Такой подход гарантирует наличие донора. В отличие от метода Фелледжи–Холта при использовании МВБС сначала ведется поиск доноров, а затем определяются действия по вменению с минимальными изменениями. При такой методологии поиск доноров ведется путем согласования с использованием всех переменных (включая те, которые, возможно, придется исчислять), задействованных в редактировании, и он может быть успешным, если производить согласование числовых переменных по ближайшему соседу плюс согласование большинства, но не обязательно всех других переменных. При этом определяются действия по вменению с учетом каждого потенциального донора и такие приемы, которые позволяют выявлять действия по вменению с минимальными изменениями. При такой методике учитываются также действия по вменению по ближайшему соседу с минимальными изменениями; иногда они могут дать более достоверные вмененные записи. И наконец, случайным образом выбирается одно из действий по вменению с минимальными изменениями и по вменению по ближайшему соседу с минимальными изменениями и производится исчисление рассматриваемого элемента данных.

22. Хотя метод Фелледжи–Холта и МВБС требуют большого объема вычислений, эффективный алгоритм для этого имеется, и поэтому с современными компьютерами эти методы вполне можно применять. Особенно это касается метода МВБС, при использовании которого можно легко решать несколько бо-

лее сложные проблемы вменения и редактирования по сравнению с методикой Фелледжи–Холта.

23. Все вышеперечисленные методы вменения дают одно исчисленное значение для каждой недостающей или несогласующейся величины. Все эти методы в какой-то степени искажают обычное распределение значений для рассматриваемого элемента данных и могут привести к неправильным оценкам дисперсии, если используются стандартные формулы ее оценки. Степень погрешности может меняться в значительной мере в зависимости от количества произведенных вменений и используемого метода.

24. Многократное вменение — это метод, предложенный Рубином (Rubin, 1987), при применении которого эта проблема решается путем исчисления несколько раз (m) каждого значения, требующего вменения. Затем из собранного набора данных для соответствующего элемента можно получить m оценок. Из них получается одна комбинированная оценка вместе с совокупной оценкой дисперсии, которая выражает степень неопределенности в отношении того, какую величину исчислять. Недостаток метода многомерного вменения заключается в том, что он требует больше усилий по обработке данных и исчисления оценок.

25. В большинстве систем вменения используется сразу несколько методов вменения; как правило, по возможности используется дедуктивное вменение, за которым следует применение одной или более других процедур. Большинство национальных статистических управлений используют ту или иную форму метода динамического вменения при редактировании и исчислении данных переписи. В настоящее время наиболее широкое применение находит последовательное вменение «hot deck» и метод Фелледжи–Холта. Из тех национальных статистических управлений, которые в настоящее время пользуются методом Фелледжи–Холта, одно переходит на метод МВБС, а ряд других рассматривают этот вопрос. Однако с учетом предполагаемого круга основных читателей в данном *Руководстве* основное внимание обращается на ту или иную форму последовательного вменения «hot deck».

Приложение VI

Программные пакеты компьютерного редактирования

1. В условиях распространения относительно недорогих микрокомпьютеров страны могут тщательно и своевременно редактировать данные переписей и обследований^a. До недавнего времени каждая страна вынуждена была сама составлять собственную программу редактирования, что требует больших затрат времени на отладку и обработку. Однако с появлением стандартных программных пакетов компьютерного редактирования странам сейчас значительно легче удовлетворять свои потребности в редактировании, и это требует меньше знаний и опыта в области обработки данных.

2. Одно из преимуществ применения программного пакета компьютерного редактирования заключается в том, что при правильном использовании данные будут согласованными и достоверными, поэтому появится возможность более своевременного составления таблиц. Для написания программ редактирования можно использовать многие компьютерные программные пакеты, например SAS и SPSS, а также другие языки высокого уровня. Кроме того, та или иная страна может использовать один из компьютерных программных пакетов, составленных специально для целей редактирования данных по результатам переписей и обследований. Для большинства стран общий процесс редактирования идет быстрее при использовании программного пакета, чем специально составленных программ, поскольку готовый пакет не требует такого уровня знаний в области обработки данных, как при работе со специально составленными программами.

3. Хороший компьютерный программный пакет редактирования предусматривает общение между специалистами по содержательной части переписи и программистами. Хороший пакет редактирования должен допускать установку описательного кода или символического кода рядом с кодом программирования, если только сам код программирования не будет очевиден или прозрачен для специалистов по содержательной части переписи. Демографы и другие специалисты должны иметь возможность проходить через программу строка за строкой и четко понимать, что делает данная программа.

4. Любой имеющийся в наличии компьютерный программный пакет редактирования, который та или иная страна могла бы использовать, должен давать возможность составления отчетов по различным проверкам, тестам и вменениям, которые необходимы для редактирования данных переписи. Эти требования действуют даже тогда, когда специалисты по обработке данных составляют программы редактирования по индивидуальному заказу. Такой пакет должен отвечать следующим требованиям:

- a) давать возможность вводить данные с клавиатуры и/или проверять вводимые данные. Этот пакет должен обеспечивать возможность до-

^a Важно отметить, что основной упор в настоящем *Руководстве* делается на редактирование данных переписи. В силу того что в обследованиях присутствует меньше респондентов и, как правило, больше вопросов, для них часто разрабатываются более детальные схемы редактирования. Некоторые программные пакеты, которые рассматриваются в данном приложении, предназначены для проведения обследований в более населенных странах. Однако такие пакеты могут применяться и в небольших странах, проводящих переписи. При увеличении размера файлов в них несколько сложнее использовать некоторые статистические методы, такие как регрессия и многовариантный анализ. С другой стороны, при наличии полного файла влияние фактора неполучения ответов будет несколько ниже, чем в случае обследования, в котором каждое вошедшее в выборку лицо «подсчитывается» за множество других. Следовательно, статистическое управление в каждой стране должно протестировать различные программные пакеты для определения того, какой из них в наибольшей мере соответствует их потребностям.

бавления схем пропуска тех или иных элементов данных. Например, группа редактирования может решить, что сведения о рождаемости в отношении лиц мужского пола должны быть пропущены;

- b) выполнять редактирование структуры данных, которое позволит определить, имеются ли фактически те типы записей, которые должны быть в наличии, включая, например, запись по жилищной единице для каждого порядкового номера;
- c) производить записи, если они пропущены, и/или добавлять весовые коэффициенты к существующим записям;
- d) определять, что каждая переменная имеет достоверное значение;
- e) хранить все или часть уже отредактированных записей;
- f) проверять согласованность между двумя или более характеристиками в одной и той же записи и между записями. При сокращенном варианте этого пакета должна производиться проверка на согласованность внутри домохозяйств, сверка ответов с ответами членов предыдущего домохозяйства; исчислять значения методом вменения «hot deck», если страна предпочтет использовать этот метод;
- g) использовать несколько значений в той или иной записи или из нескольких записей для построения производной переменной и вводить производную переменную в соответствующую запись;
- h) выявлять и устранять двойные записи;
- i) составлять дневник ошибок и изменений в разбивке по небольшим географическим районам.

5. Обычно с помощью готовых пакетов или программ редактируется одна запись за один раз, однако современные программные пакеты позволяют также производить проверку между записями, особенно в рамках жилищных единиц.

6. Как отмечено в тексте, до появления метода Фелледжи–Холта (Fellegi-Holt, 1976) и его последующих версий практически во всех схемах редактирования использовался нисходящий принцип; иными словами, элементы данных редактировались в определенном порядке — обычно, но не всегда, в том же порядке, в котором осуществлялся сбор данных. Например, поскольку первым элементом данных о населении обычно является «степень родства», то показатель «пол» будет редактироваться на основе этого элемента, затем показатель «возраст» — на основе пола и степени родства, и так далее.

7. За последние десятилетия было разработано несколько систем вменения, предполагающих внесение минимальных изменений и основанных на методе Фелледжи–Холта. Среди них можно выделить систему CANEDIT и «Общую систему редактирования и вменения (General Edit and Imputation System) (GEIS), разработанные Статистическим управлением Канады, за которыми последовала «Канадская система редактирования и вменения данных по результатам переписей» (CANadian Census Edit and Imputation System) (CANCEIS) (Bankier, 2005; Chen, 2007)^b. В Бюро переписей Соединенных Штатов Америки разработана система DISCRETE (Chen and others, 2000; Winkler, 1997a; Winkler, 1997b; Winkler and Chen, 2002) а также «Структурированные программы редактирования и ссылок в отношении экономических данных (Structured Programs for Economic Editing and Referrals) (SPEER) (Greenberg and Surdi, 1984; Winkler and Draper, 1997). Ковар и Уинклер (Kovar and Winkler, 2000) провели довольно подробное сравнение

^b В обзоре системы CANCEIS Банкьер, Лашанс и Пурье (Banquier, Lachance and Poirier, 2000, p. 10) дают ей следующую краткую характеристику: система CANCEIS с ее в высшей степени эффективными алгоритмами редактирования и вменения является весьма перспективной в плане решения общих проблем вменения, в которых задействовано множество правил редактирования и большое число качественных и количественных переменных, когда более подходящей является схема вменения по данным донора с минимальными изменениями. Тем не менее применение алгоритма редактирования и вменения с минимальными изменениями в рамках концепции Фелледжи–Холта все же является предпочтительным методом для незначительных проблем вменения, если будет недостаточное число доноров или если более предпочтительно использовать другой метод для осуществления вменения.

указанных систем Канады и Соединенных Штатов. К другим программным пакетам редактирования, основанным на методе Фелледжи–Холта можно отнести пакет редактирования, разработанный Статистическим управлением Нидерландов под названием CHERRYPI (De Waal and Van de Pol, 1997).

8. В ходе переписи в Канаде в 1996 году (и в других переписях) был использован другой подход, а именно метод вменения по ближайшему соседу (МВБС). В версии 1996 года вменение проводилось в отношении ответов, касающихся пола, возраста, брачного статуса и степени родства одновременно для всех проживающих в доме лиц (Bankier, 1999). Этот метод был улучшен и расширен для переписи 2001 года в Канаде и в ходе последующих статистических мероприятий (Bankier, Lachance and Poirier, 2000; Banquier, 2001).

9. При использовании метода вменения по ближайшему соседу сначала ведется поиск доноров — «ближайших соседей», а затем, исходя из данных по этим донорам, определяются возможности вменения с минимальными изменениями. Хотя метод Фелледжи–Холта предполагает исчисление минимального числа переменных и сохранение целостности групп населения, МВБС в этом плане дает преимущество с точки зрения объема вычислительных операций, поскольку он ставит операции в обратном порядке, начиная с поиска доноров, а затем внося изменения в минимальное число переменных. При этом, однако, если МВБС может осуществлять вменение, используя исключительно данные доноров, то метод Фелледжи–Холта может использоваться и с другими методологиями (например, нисходящим методом). Статистическое управление Канады включило МВБС в свою «Канадскую систему редактирования и вменения данных по результатам переписей» для переписей 2001 и 2006 годов.

10. На состоявшихся после 2000 года серии совещаний многие европейские статистики собирались вместе для обсуждения различных аспектов редактирования и вменения данных по результатам переписей и обследований. Дискуссии концентрировались вокруг темы «Развитие и оценка новых методов редактирования и вменения» (EUREDIT) (Chambers, 2000; Poirier, 2000). Кроме того, Статистическое управление Канады разработало методы измерения и уменьшения дисперсии в процессе вменения. Канадская система SIMPVAR, предназначенная для обследований, способна производить вменение по четырем основным методам вменения (соотношения, средние значения, «hot deck» и «поиск ближайшего соседа») (Rancourt and others, 1997). Для переписи в Италии 2001 года Национальный институт статистики (Istituto Nazionale di Statistica) (ISTAT) разработал систему редактирования и вменения с использованием итальянского программного пакета DIESIS (Data Imputation and Editirs System — Italian Software) и других методов (Di Zio, 2002; Bianci and others, 2005).

11. Существуют и другие методы вменения неизвестных величин — а именно использование фактических примеров. Иногда используются методы вменения по средним значениям. В некоторых странах используются модели по уравнению регрессии (Государственный комитет по статистике Российской Федерации, 2000 год). Кроме того, методы по уравнению регрессии использовались для исчисления возраста в кратком переписном листе переписи 2000 года в Соединенных Штатах (Williams, 1998).

12. В силу преимуществ интерактивного ввода данных с клавиатуры, в некоторых системах кодирование, ввод с клавиатуры и редактирование объединены в единую систему, особенно применительно к обследованиям. К таким системам

можно отнести бразильскую систему CRIPTAX (Hanono and Barbosa, n. d.), которая использует метод редактирования при вводе данных. Другие системы, например «Система обработки данных по результатам переписей и обследований» (Census and Survey Processing System) (CSPro), имеют некоторые функции интерактивного редактирования. Как отмечалось выше, национальные статистические управления должны самостоятельно принимать решения по поводу окупаемости затрат на оборудование, программное обеспечение, персонал, трудозатраты и т. д.

13. Что касается нисходящего подхода, Бюро переписи Соединенных Штатов разработало «Комплексную микрокомпьютерную систему обработки данных» (Integrated Microcomputer Processing System) (IMPS) для использования в переписях с 1980 по 2000 год. Этот программный пакет написан в системе DOS и включает ввод данных, редактирование и составление таблиц, а также ряд других функций, и многие страны до сих пор используют этот пакет. В период конца 1990–2000-х годов Бюро переписи разработало версию пакета для Windows под названием «CSPro» (как указано выше), который выполняет многие из тех же задач, но при этом совместим с Windows^c. Пакет CSPro вполне пригоден для использования во всех переписях и обследованиях в небольших странах; при этом, однако, темпы обработки данных настолько медленны, что процесс редактирования занимает больше времени, чем при применении IMPS. Тем не менее страны, начинающие работы по редактированию данных «с нуля», должны использовать версию для Windows. Прикладные программы как в IMPS, так и в CSPro могут разрабатываться для метода Фелледжи–Холта, в то время как метод «по ближайшему соседу» потребует гораздо больше работы.

^c CSPro — это программный пакет для ввода, редактирования, табулирования и распространения данных по результатам переписей и обследований. Пакет CSPro объединяет функции «Комплексной микрокомпьютерной системы обработки данных» (IMPS) и «Комплексной системы для анализа результатов обследований» (ISSA), работающей в программной среде Windows.

14. Пакет CSPro дает возможность пользователю создавать, изменять и запускать прикладные программы ввода данных, пакетного редактирования и составления таблиц из единой интегрированной среды проектирования. Он осуществляет обработку данных по каждому конкретному варианту (для одного или более переписных листов), где один вариант может состоять из одной или многих записей данных. Данные хранятся в «Американском стандартном коде для обмена информацией» (ASCII) в форме текстовых файлов, описываемых с помощью словарей данных. CSPro содержит общий процедурный язык с широкими возможностями для контроля ввода данных и введения правил редактирования.

15. В конкретном плане функция пакетного редактирования CSPro выявляет и сообщает о содержащихся в переписных листах ошибках в структуре, значениях и согласованности данных. Этот пакет может изменять (вменять) значения данных с применением простых или сложных методов. Он может составлять краткие или подробные отчеты об ошибках и исправлениях, иметь доступ к многочисленным поисковым файлам и считывать вторичные файлы или делать в них записи.

16. Пакет CSPro также содержит инструменты для просмотра данных и других текстовых файлов, просмотра таблиц и тематических карт, создаваемых этим же пакетом, перевода словарей данных в формате IMPS и ISSA в формат CSPro и обратно, а также перевода файлов форм (карт) из формата Института исследований экологических систем (ESRI) в файлы карт в формате CSPro. Пакет CSPro был разработан совместно Бюро переписи США и компаниями Masco International и Serpro при значительном финансировании со стороны Агентства международного развития Соединенных Штатов. Пакет CSPro находится в общем пользовании, доступен бесплатно и свободно распространяется, его можно скачать по адресу: www.census.gov/ipc/www/cspro.

Глоссарий

- Автоматическая корректировка** — Корректировка ошибок в данных с помощью компьютера без вмешательства человека. Один из аспектов автоматизированного редактирования данных (Pierzchala, 1995).
- Блок-схема** — Изображение в виде диаграммы всех функций, которые должны быть выполнены.
- Ввод данных в режиме «heads down»** — Стилль ввода данных, при котором компьютер не обнаруживает ошибки в данных по мере их ввода, позволяя оператору быстро вводить данные (Pierzchala, 1995).
- Ввод данных в режиме «heads up»** — Стилль ввода данных, при котором компьютер обнаруживает ошибки в данных по мере их ввода, позволяя оператору немедленно вносить исправления (Pierzchala, 1995). См. «[Интерактивный ввод данных с клавиатуры](#)».
- Весовые коэффициенты** — По методу редактирования и вменения Феллджи–Холта весовые коэффициенты присваиваются полям на основе надежности. Чем выше весовой коэффициент, тем больше вероятность вменения того или иного поля (при прочих равных факторах). Весовые коэффициенты могут также присваиваться редактируемым данным (Pierzchala, 1995).
- Вменение** — Присвоение того или иного значения полю в связи с отсутствием ответа или замена указанного значения, которое по определению не согласуется с набором отредактированных данных (Pierzchala, 1995).
- Вменение «hot deck»** — Метод вменения, при котором донорские записи берутся из текущей «колоды» выборочных данных (в отличие от метода вменения «cold deck», когда донорская запись берется из данных прошлого обследования) (Pierzchala, 1995).
- Вменение «hot deck» по одному донору** — Вменяются данные для не прошедшей редактирование записи по одному донору, что позволяет проводить совместное вменение всех значений рассматриваемого элемента в записи, которая при редактировании определена как сомнительная.
- Вменение на основе модели** — Используются данные из прошедших редактирование записей для регрессного вычисления переменной, которую необходимо исчислить по набору предикторных переменных.
- Вменение по ближайшим соседним значениям** — Приписывает то или иное значение элемента данных для не прошедшей редактирование записи из «ближайшей» прошедшей редактирование записи, где понятие «ближайшая» запись определяется с использованием функции расстояния исходя из других известных переменных.
- Вменение по данным донора** — Метод, при использовании которого каждая запись, требующая вменения, запись о получателе или кандидате соединяется попарно с одной записью из определенной совокупности доноров, как, например, при вменении «hot deck» (Pierzchala, 1995).

Вменение по общим средним значениям — Приписывается среднее значение данного элемента для прошедших редактирование записей вместо недостающего или несогласующегося элемента по всем не прошедшим редактирование записям.

Вменение по средним значениям в разбивке на классы — Используются классы вменения, определяемые для создания групп записей с определенной степенью сходства.

Вменение по уравнению регрессии — Используются данные из прошедших редактирование записей для регрессного вычисления переменной, которую необходимо исчислить по набору предикторных переменных.

Внутренняя согласованность — Этот термин относится к отношениям между переменными для данной единицы выборки и является основанием для редактирования в большинстве процедур обследования (Ford, 1983; Pierzchala, 1995).

Дедуктивное вменение — Метод, при использовании которого можно с определенностью вывести недостающую или несогласующуюся величину, зачастую на основе структуры ответов, сообщаемых на другие вопросы переписного листа.

Детерминистическое вменение — Такая ситуация возникает, когда только одно значение того или иного поля дает запись, удовлетворяющую всем требованиям редактирования. Возникает в некоторых случаях (когда часть от целого не прибавляется к целому). Первое решение необходимо проверить при автоматизированном редактировании и вменении данных обследования (Pierzchala, 1995).

Детерминистическое редактирование — Редактирование, которое в случае нарушения указывает на ошибку в данных с вероятностью до единицы. Например: возраст = 5 и статус = мать. Следует отличать от стохастического редактирования (Pierzchala, 1995).

Достоверная запись — Запись, в которой нет недостающих величин и которая проходит все этапы редактирования (Pierzchala, 1995).

Журнал редактирования — См. «[Контрольный журнал](#)».

Запись — Хранящееся на магнитном носителе машиночитаемое отображение данных обследования. Обычно имеется одна запись для каждого переписного листа, хотя можно разбить данные из одного переписного листа на более чем одну запись, например, по населению и по жилому фонду (Pierzchala, 1995).

Интерактивный ввод данных с клавиатуры — Стилль ввода данных, при котором компьютер обнаруживает ошибки в данных по мере их ввода, позволяя оператору немедленно вносить исправления. См. «[Ввод данных в режиме «heads up»](#)».

Количественное редактирование — Процедуры редактирования применяются к полям, измеряемым по непрерывной шкале (Pierzchala, 1995).

Контрольный журнал — Метод отслеживания изменений в значениях поля и причин и источников каждого изменения. Контрольный журнал, как правило, начинают вести после завершения первоначального опроса.

Косвенное редактирование — Незаявленные приемы редактирования, выводимые логическим путем из явных приемов редактирования, написанных специалистом по содержательной части переписи (Pierzchala, 1995).

Макроредактирование — Обнаружение отдельных ошибок путем 1) проверки агрегированных данных или 2) проверок, используемых по отношению ко всей совокупности записей. Эти проверки основаны на оценках (Granquist, 1987; Pierzchala, 1995).

Массив — Набор числовых величин. Иногда называется матрицей и может использоваться для хранения периодически повторяющихся числовых данных.

Матрица «cold deck» — Первоначальная статическая матрица; база корректировки, для которой все элементы даются до начала корректировки и не меняются в процессе этой процедуры. Например, базой корректировки могут быть данные за предыдущий год. С помощью модифицированной матрицы «cold deck» можно корректировать значения «cold deck» в соответствии с текущей (возможно, агрегированной) информацией.

Метод Фелледжи–Холта для автоматической корректировки — Метод автоматической корректировки, при котором меняется наименьшее число элементов данных, и модель Фелледжи–Холта используется для определения приемлемых наборов значений или диапазонов для вменяемых элементов данных. Может применяться последовательное или одновременное вменение методом «cold deck» или «hot deck».

Методология вменения по ближайшим соседним значениям — Аналогична методу Фелледжи–Холта в том, что учитывает одновременно все редактируемые данные, не определяет явным образом действия по вменению и проводит исчисления по одному донору. МВБС просматривает каждую не прошедшую редактирование запись для определения действий по вменению с минимальными изменениями в зависимости от наличия потенциальных доноров.

Микро/макро-редактирование — Процедура редактирования, когда подробные редактируемые микроданные заменяются комбинацией редактируемых микро- и макростатистических данных. Редактируемые микроданные при комбинированной процедуре представлены менее подробно, чем в первом случае. Идея заключается в том, чтобы выработать приемы редактирования данных по результатам обследований не на основе подхода по принципу «уловить все несоответствия в данных», а по принципу «воздействовать на оценки» (Granquist, разные даты, Pierzchala, 1995).

Микроредактирование — Традиционное редактирование данных на уровне записи. Логический антоним макроредактирования (Pierzchala, 1995).

Минимальный набор — Наименьший набор полей, требующих вменения, который гарантирует, что будут пройдены все этапы редактирования (Pierzchala, 1995).

Многократное вменение — Каждое значение, требующее вменения, исчисляется несколько раз, а затем выводится оценка для рассматриваемого элемента.

Многомерное редактирование — Тип статистического редактирования, при котором многомерные распределения используются для оценки данных и обнаружения отклоняющихся значений (Pierzchala, 1995).

- Не прошедшие редактирование записи** — При редактировании и вменении записи, которые не прошли по меньшей мере одно редактирование того или иного рассматриваемого элемента (элементов данных)
- Отклоняющиеся значения** — Значения элементов данных, которые выходят за некоторые границы согласно соответствующему определению таких границ (Pierzchala, 1995).
- Отсутствие ответа** — Не полностью заполненный переписной лист или отсутствующий переписной лист (Pierzchala, 1995).
- Ошибки качества** — Ошибки, которые могут исказить качество данных: например, систематические ошибки, которые ведут к погрешностям (Granquist, 1984; Pierzchala, 1995).
- Перечень кодов** — Перечень всех разрешенных (допустимых) величин того или иного элемента данных.
- Поиск** — При применении процедуры вменения «hot deck» — мероприятие по поиску донорской записи (Pierzchala, 1995).
- Полный набор редактируемых данных** — Объединение явных и косвенных отредактированных данных. Необходим для создания возможных участков для вменения (если возникнет необходимость произвести вменение в целях редактирования) (Pierzchala, 1995).
- Последовательное вменение «hot deck»** — Вменение, при котором статистический ряд переменных редактируется последовательно и только отредактированные величины используются в качестве последующих переменных «hot deck».
- Проверка достоверности** — Контрольное редактирование, которое производится между полями в конкретной записи. Сюда входит проверка каждого поля каждой записи, чтобы убедиться, что она содержит достоверные входные данные, и проверка входных данных в определенной заранее комбинации полей, чтобы удостовериться, что эти данные согласуются между собой (Pierzchala, 1995).
- Прошедшие редактирование записи** — В процессе редактирования и вменения — записи, прошедшие все этапы редактирования, относящиеся к рассматриваемым.
- Рабочий прогон** — Мероприятие по обработке больших объемов данных после устранения первоначальных «дефектов» из программы редактирования или составления таблиц.
- Редактирование (определение 1)** — Логические ограничения значений, которые может приобрести каждая переменная (Pierzchala, 1995).
- Редактирование (определение 2)** — Правила, с помощью которых выявляются недопустимые комбинации ответов (Pierzchala, 1995).
- Редактирование в рамках записи** — Другое название редактирования достоверности (Pierzchala, 1995).
- Редактирование вручную** — Редактирование, осуществляемое сотрудниками до ввода данных в компьютер (см. «[Ручное редактирование](#)») (Pierzchala, 1995).
- Редактирование для проверки согласованности** — Проверка определяющих отношений, например долей, прибавляемых к общим или «уборочным площадям», всегда меньше «засеянных площадей» (Pierzchala, 1995).

Редактирование между записями — Редактирование, производимое на полях, где есть более одной записи в ходе обследования. Статистически редактируемые данные являются примером редактирования между записями, поскольку распределения составляются на совокупностях полей по всем записям в обследовании (Pierzchala, 1995).

Ручное редактирование — Редактирование, осуществляемое сотрудниками до ввода данных в компьютер (см. «[Редактирование вручную](#)») (Pierzchala, 1995).

Символический код — Письменные инструкции или спецификации, касающиеся редактирования.

Система Феллджи–Холта — Касается допущений и целей редактирования и вменения, изложенных Феллджи и Холтом в их публикации 1976 года в «Журнале Американской статистической ассоциации» (*Journal of the American Statistical Association*). Основная особенность модели Феллджи–Холта заключается в том, что она демонстрирует необходимость использования косвенных приемов редактирования для обеспечения того, чтобы набор значений в полях данных, которые не всегда вменяются, давал конечные (вмененные) записи, удовлетворяющие требованиям всех ситуаций.

Случайное вменение в рамках отдельных классов — Используются классы вменения для ограничения случайного отбора донорской записи в тот или иной набор, имеющий некоторое сходство с записью, требующей вменения.

Случайное общее вменение — Для каждой не прошедшей редактирование записи случайным образом выбирается одна прошедшая редактирование запись из набора всех прошедших редактирование записей и ее указанное значение для рассматриваемого элемента данных вменяется для не прошедшей редактирование записи.

Согласование — При проведении вменения «hot deck» акт согласования донорской записи с записью получателя (кандидата) (Pierzchala, 1995).

Согласование функции расстояния — Приписывает то или иное значение элемента данных для не прошедшей редактирование записи из «ближайшей» прошедшей редактирование записи, где «ближайшая» запись определяется с использованием функции расстояния исходя из других известных переменных.

Согласующие переменные — Переменные, используемые для согласования записи получателя (кандидата) и записи донора (Pierzchala, 1995).

Статистическое вменение — Примером статистического вменения может быть применение модели с уравнением регрессии, согласно которой нужно исчислить зависимую переменную, а коэффициенты независимых переменных выводятся из предположительно достоверных ответов (Pierzchala, 1995).

Статистическое редактирование — Набор проверок, основанных на статистическом анализе данных респондента: например, соотношение двух полей лежит в пределах, установленных путем статистического анализа этого соотношения для предположительно достоверных ответов (Greenberg and Surdi, 1984; Pierzchala, 1995).

- Статическое согласование (при вменении «hot deck»)** — Действие по согласованию донорской записи с записью получателя (кандидата) в соответствии с определенными статистическими критериями, с тем чтобы перевести данные от донора к получателю (Pierzchala, 1995).
- Стохастическое редактирование** — Редактирование, которое в случае нарушения указывает на ошибку в данных с вероятностью меньше единицы (Pierzchala, 1995).
- Стрелка (указатель)** — Стрелка означает переменную, используемую для отметки того или иного элемента данных, или другую переменную для последующих ссылок. Например, стрелки используются для того, чтобы отметить номера строк «главы (домохозяйства)» и «супруги(а)» для последующего использования, чтобы быть уверенными, что супруги противоположного пола и оба состоят в браке.
- Структурное вменение** — Структурное вменение используется, когда между несколькими переменными есть структурная взаимосвязь. Например, итоговое значение должно равняться сумме его составляющих: поэтому если вопрос касается матери, число детей, родившихся живыми, должно равняться числу живущих на момент переписи детей плюс число умерших детей (Pierzchala, 1995).
- Структурное редактирование** — Проверки на основе логической взаимосвязи между двумя или более полями редактирования. Например, итоговое значение должно равняться сумме его составляющих или из-за предусмотренной в переписном листе схемы пропусков две переменные, лежащие на непересекающихся дорожках, не могут обе не равняться нулю. Структурное редактирование обеспечивает сохранение структуры переписного листа в записи данных (Pierzchala, 1995).
- Сходство** — В числовых данных — понятие близости двух записей, основанных на заданных согласующихся переменных. Здесь используется функция расстояния для числового выражения этого понятия в соответствии с некоторыми критериями (Pierzchala, 1995).
- Считывание данных** — Процесс, с помощью которого собранные данные переводятся в машиночитаемую форму. В подмодулях программного обеспечения, куда вводятся данные, часто проводится элементарное контрольное редактирование.
- Флажок** — Переменная, используемая для того, чтобы отметить полезную информацию о другой переменной или переменных. Например, если тот или иной элемент данных меняется с недостоверного на достоверный, такой флажок можно использовать для того, чтобы отметить первоначальную информацию или отметить, что значение этого элемента изменилось.
- Функция расстояния** — Для числовых данных функция, определяемая по согласующимся переменным в записях получателя (кандидата) и донора и используемая для числового выражения понятия схождения. Используется для нахождения согласующихся записей при вменении «hot deck» (Pierzchala, 1995).
- Явные приемы редактирования** — Приемы, явно написанные специалистом по содержанию переписи (в отличие от явных приемов редактирования с косвенными приемами) (Pierzchala, 1995).

Библиография

- Banister, J. (1980). Use and abuse of census editing and imputation. *Asian and Pacific Census Forum*, vol. 6, No. 3, pp.1–20.
- Bankier, M. (1999). «Experience with the New Imputation Methodology used in the 1996 Canadian Census with Extensions for Future Censuses», Proceedings of the Workshop on Data Editing, UNECE, Italy (Rome).
- _____ (2005). Edit and imputation for the 2006 Canadian Census. Статья представлена на рабочей сессии по вопросам редактирования статистических данных, Оттава, 16–18 мая 2005 года.
- _____, A.-M. Houle and M. Luc (n. d.). Canadian census demographic variables imputation. Manuscript.
- Bankier, M., M. Lachance and P. Poirier (1999). A generic implementation of the new imputation methodology. In *Proceedings of the Section on Survey Research Methods*. Alexandria, Virginia: American Statistical Association forthcoming.
- _____ (2000). 2001 Canadian Census minimum change donor imputation methodology. Статья представлена на рабочей сессии UN/ЕСЕ по вопросам редактирования статистических данных, Кардифф, Соединенное Королевство, 18–20 октября 2000 года.
- Bankier, M., P. Poirier, and M. Lachance (2001). Efficient methodology within the Canadian Census Edit and Imputation System (CANCEIS). Proceedings of the Annual Meeting of the American Statistical Association, 5–9 August 2001.
- Bankier, M., and others (1996). Imputing numeric and qualitative census variables simultaneously. In *Proceedings of the Section on Survey Research Methods*. Alexandria, Virginia: American Statistical Association, pp. 287–292.
- Bianchi, G., and others (2005). New procedures for editing and imputation of demographic variables. Статья представлена на рабочей сессии по вопросам редактирования статистических данных, Оттава, 16–18 мая 2005 года.
- Boucher, L. (1991). Micro-editing for the annual salary of manufacturers: what is the value added? In *Proceedings of the Annual Research Conference*. Washington, D.C.: United States Bureau of the Census, pp. 765–781.
- Chambers, Ray (2000). Evaluation criteria for editing and imputation in EUREDIT. Статья представлена на рабочей сессии UN/ЕСЕ по вопросам редактирования статистических данных, Кардифф, Соединенное Королевство, 18–20 октября 2000 года.
- Chen, Bor-Chung (2007). CANCEIS Experiments of edit and imputation with 2006 Census test data. Statistical Research Division Study Series, No. 2007-1. Washington, D.C.: U.S. Bureau of the Census.

- _____ and others. (2000). Using the DISCRETE Edit System for ACS Surveys. Bureau of the Statistical Research Division Statistical Research Report Series, No. RR2000/03. Washington, D.C.: U.S. Bureau of the Census.
- De Waal, Tom, and Frank van de Pol (1997). A recipe for applying CHERRYPI in the edit process. *Статья представлена на рабочей сессии по вопросам редактирования статистических данных, Прага, 14–17 октября 1997 года.*
- Di Zio, M (2002). Evaluating editing and imputation processes: the Italian experience. *Статья представлена на рабочей сессии UN/ECE по вопросам редактирования статистических данных, Хельсинки, 27–29 мая 2002 года.*
- Fellegi, I. P., and D. Holt (1976). A systematic approach to automatic edit and imputation. *Journal of the American Statistical Association*, vol. 71, No. 353 (March), pp. 17–35.
- Ford, Barry L. (1983). An overview of hot deck procedures. In *Incomplete Data in Sample Surveys*, vol. 2, *Theory and Bibliographies*. William G. Madow. Ingram Olkin and Donald B. Rubin, eds.
- Granquist, L. (1984). Data editing and its impact on the further processing of statistical data. *Статья представлена на семинаре-практикуме по вопросам статистического расчета, Будапешт, 12–17 ноября 1984 года.*
- _____ (1987). The short-term developing program for computer-supported editing at Statistics Sweden. *Доклад представлен на совещании Объединенной группы по вопросам редактирования данных, Мадрид, 22–24 апреля 1987 года. Stockholm: Statistics Sweden.*
- _____ (1997). The new view on editing. *International Statistical Review*, Vol. 65, No. 3, New York: Academic Press, pp. 381–387.
- _____, and J. G., Kovar (1997). Editing of survey data: how much is enough? In *Survey Measurement and Process Quality*, Lyberg and others, eds. New York: Wiley and Sons, pp. 415–435.
- Greenberg, Brian, and Rita Surdi (1984). A flexible and interactive edit and imputation system for ratio edits. In *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, Virginia: American Statistical Association pp. 421–426.
- Hanono, Reina Marta, and Dulce Maria Rocha Barbosa (n.d.). Generalized environment for application development for capturing, editing and coding statistical survey's data. Rio de Janeiro: Brazilian Institute of Geography and Statistics (IBGE).
- Ireback, H. (2000). The impact of new information technologies on data collection at Statistics Sweden. *Статья представлена на рабочей сессии UN/ECE по вопросам редактирования статистических данных, Кардифф, Соединенное Королевство, 18–20 октября 2000 года.*
- Kalton, G., and D. Kasprzyk (1982). Imputing for missing survey responses. In *Proceedings of the Section on Survey Research Method*. American Statistical Association, pp. 23–31.

- _____ (1986). The Treatment of Missing Survey Data. *Survey Methodology*, vol. 12, pp. 1–16.
- Kovar, J., and W. Winkler (2000). Comparison of GEIS and SPEER for editing economic data. Bureau of the Census Statistical Research Division Statistical Research Report Series, No. RR 2000/04. Washington, D.C.: U.S. Bureau of the Census.
- Naus, J. I. (1975). *Data Quality Control and Editing*. New York: Marcel Dekker.
- Nordbotten, S. (1963). Automatic editing of individual statistical observations. Conference of European Statisticians, Statistical Standards and Studies, No. 2. New York: United Nations.
- Pierzchala, M. (1995). Editing systems and software. In *Business Survey Methods*. B. G. Cox, and others, eds. New York: John Wiley and Sons, pp. 425–411.
- Poirier, C. (2000). A prototype knowledge base on data editing and imputation. Статья представлена на рабочей сессии UN/ECE по вопросам редактирования статистических данных, Кардифф, Соединенное Королевство, 18–20 октября 2000 года.
- Pullum, T. W., T. Harpham and N. Ozsever (1986). The machine editing of large-sample surveys: the experience of the World Fertility Survey. *International Statistical Review*, vol. 54, 311–326.
- Rancourt, E., and others (1997). Estimation of variance in presence of imputation. Proceedings of Symposium 1997: New Directions in Surveys and Censuses. Ottawa: Statistics Canada, pp. 273–279.
- Rubin, D. B. (1987). *Multiple Imputation for Non-response in Surveys*. New York: Wiley.
- Russian Federation, State Committee on Statistics (2000). Data imputation based on regression models with variations of entropy. Статья представлена на рабочей сессии UN/ECE по вопросам редактирования статистических данных, Кардифф, Соединенное Королевство, 18–20 октября 2000 года.
- Sande, I. G. (1982). Imputation in surveys: coping with reality. *American Statistician*, vol. 36, pp. 145–152.
- Särndal, C. E., B. Swensson and J. Wretman (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Statistics Canada (1998). *Statistics Canada Quality Guidelines*, 3rd Edition. Ottawa: Statistics Canada.
- United Nations (1992a). *Справочник по проведению переписей населения и жилого фонда, часть I: Планирование, организация и управление переписями населения и жилого фонда*. Методологические исследования, № 54, в продаже под № R.92.XVII.8.
- _____ (1992b). *Справочник по проведению переписей населения и жилого фонда, часть II: Демографические и социальные характеристики*. Методологические исследования, № 54, в продаже под № R.91.XVII.9.

- _____ (1999). *Standard Country or Area Codes for Statistical Use*. Statistical Papers, Series M, No. 49/Rev.4. Sales No. M.98.XVII.9.
- _____ (2008). *Принципы и рекомендации в отношении переписей населения и жилого фонда, второе пересмотренное издание*. Статистические документы, серия М, № 67/Rev.2, в продаже под № R.07.XVII.8.
- United Nations Statistical Commission and Economic Commission for Europe (1994). *Statistical Data Editing*, vol. No.1: *Methods and Techniques*. Conference of European Statisticians, Statistical Standards and Studies Series, No. 44. Sales No. 94.II.E.36.
- _____ (1997). *Statistical Data Editing*, vol. 2 : *Methods and Techniques*. Conference of European Statisticians, Statistical Standards and Studies Series, No. 48. Sales No. 96.II.E.30.
- Williams, Todd R. (1998) Imputing person age for the 2000 Census Short Form: a model-based approach. Washington, D.C. : Bureau of the Census Statistical Research Division, Statistical Research Report Series, No. RR98/07.
- Winkler, W. E. (1997a). Edit/imputation System for the U.S. Decennial Census. Статья представлена на рабочей сессии по вопросам редактирования статистических данных, Прага, 14–17 октября 1997 года.
- _____ (1997b). Set-covering and editing discrete data. Technical report. Washington, D.C.: U.S. Bureau of the Census.
- _____ (2006). Data quality: automated edit/imputation and record linkage. U.S. Census Bureau Statistical Research Division Research Report Series, No. 2006-7. Washington, D.C.: U.S. Bureau of the Census.
- _____, and B. C. Chen (2002). Extending the Fellegi-Holt Model of Statistical Data Editing. U.S. Bureau of the Census Statistical Research Division, Statistical Research Report Series, No. 2002-02.
- Winkler, W.E., and L.R. Draper (1997). The SPEER edit system. In U.N. Statistical Commission and Economic Commission for Europe, *Statistical Data Editing*, vol. 2, *Method and Techniques*. Conference on European Statisticians Statistical Standards and Studies, No. 48. Sale No. E.96.II.E.30, pp. 56–62.

