



Экономический и Социальный Совет

Distr.: General
19 August 2009
Russian
Original: English

Европейская экономическая комиссия

Конференция европейских статистиков

Группа экспертов по переписям населения
и жилищного фонда

Двенадцатое совещание

Женева, 28-30 октября 2009 года

Пункт 4 предварительной повестки дня

**Качество переписи и предупреждение
идентификации**

Оценки качества комплексной переписи 2008 года - Израиль

Записка Центрального бюро статистики Израиля

I. Введение

1. Ход и результаты переписи оценивались по ряду составляющих компонентов качества, которые включали в себя охват, точность, непротиворечивость, полноту, надежность и достоверность^{1,2}. В публикации Европейской экономической комиссии Организации Объединенных Наций "Рекомендации Конференции европейских статистиков по проведению переписей населения и жилищного фонда 2010 года" перечисляются некоторые элементы качества применительно к трем этапам переписного процесса: планирования, осуществления и оценки³. Эти элементы касаются: точности, своевременности, доступности, интерпретируемости и непротиворечивости. Израиль провел комплексную перепись в 2008 году без послепереписного обследования, поскольку предполагалось, что комплексная перепись благодаря своим двум выборочным совокупностям обеспечит необходимую информацию для оценки переписных мероприятий. В процессе оценки также учитывался тот факт, что перепись носила ком-

¹ Willem de Vries (2002), Dimensions of Statistical Quality, Inter-agency meeting on coordination of statistical activities, SA/2002/6/add.1.

² Leo, Pipnio, Yang W. Lee and Ricahrd Y., Wang (2002), Data Quality Assessment, Communication of the ACM April 2002/vol. 45, no, 4ve.

³ Рекомендации Конференции европейских статистиков по проведению переписей населения и жилищного фонда 2010 года, Европейская экономическая комиссия Организации Объединенных Наций, 2006 год.

плексный характер и предусматривала как использование административных источников, так и сбор данных на местах. Сбор переписных данных был полностью компьютеризирован благодаря использованию переносных компьютеров для сбора данных и ежедневной передаче данных в центральную базу. Эти характеристики позволили осуществлять мониторинг за полевыми операциями с использованием специальных инструментов, использовавших преимущества технологии. На настоящий момент этот процесс оценки еще не завершен; он будет закончен к концу 2009 года. Тем не менее большинство процедур оценки уже завершены.

II. Методологии и процедуры комплексной переписи (резюме)

2. Комплексная перепись в значительной мере опирается на административные источники, главным образом на регистр населения, а также другие административные источники для обновления итогов переписи в отношении "неактивных" лиц, а также в качестве подспорья для построения "административных семей". Качество этих данных зависит от потребностей, процедур, действующих норм и законов, в соответствии с которыми собираются и ведутся административные данные. Они необязательно совпадают с определениями и потребностями переписи. Целями двух выборочных обследований являлась корректировка недостатков административных источников. Основными целями этих обследований являлись:

а) оценка качества основных административных источников и соответствующая корректировка оценки численности "постоянных жителей" страны и их географического распределения;

б) сбор данных по социально-демографическим характеристикам населения и домохозяйств на уровне "переписных участков" (ПУ) и малых населенных пунктов. Базой комплексной переписи служит регистр населения⁴. Были сформированы две независимые выборочные совокупности: территориальная выборка на основе географической базы данных, созданная конкретно для целей переписи, в соответствии с которой территория страны была разбита на 40 000 ячеек (которые охватывают всю территорию страны), в каждую из которых в среднем входило 50 домохозяйств (от 30 до 70 домохозяйств), для оценки охвата регистра населения, и выборка на основе регистра населения для оценки пропусков в регистре. Институциональное население подвергалось полному учету в соответствии с классической процедурой переписи.

III. Полевые мероприятия

3. Для сбора данных по территориальной выборке использовался компьютеризированный вопросник, установленный на переносных компьютерах (АЛОК). Опрос по выборке из регистра населения производился по телефону (АЛОТ). Для сбора данных по территориальной выборке страна была разделена на 12 регионов. Полевые операции заняли три месяца, причем первый месяц был посвящен профподготовке, ориентации и маркировке включенных в выборку строений и жилищ. Период фактического сбора данных продлился семь

⁴ http://www.cbs.gov.il/www/mifkad/census2008_e.pdf

недель. Контрольной датой переписи являлось 27 декабря 2008 года, а сбор данных был начат 28 декабря. Счетчики были наняты на трехмесячный период (с декабря 2008 года по февраль 2009 года) и оплачивались дифференцированно в зависимости от качества собранных данных и степени трудности сбора данных. В качестве нижнего предела использовался общенациональный минимальный размер оплаты труда.

IV. Усовершенствованный регистра населения

4. В регистре населения ведутся данные о лицах, являющихся как гражданами, так и "постоянными жителями" страны с момента создания регистра в ноябре 1948 года, вскоре после создания Государства Израиль. Лица, которые определенно более не проживают на постоянной основе, получают юридический "неактивный" статус (главным образом это касается умерших или официальных уведомивших министерство внутренних дел о проживании в другой стране). Лица весьма преклонного возраста (старше 110 лет) в отсутствие поступления новых данных в регистр на протяжении длительного периода времени рассматриваются в качестве "неактивных", что означает, что они, вероятнее всего, были из состава населения. Записи о лицах, которые, как предполагается, отсутствуют в стране на протяжении длительного непрерывного периода времени без обновления данных в регистре с момента их последнего заявленного убытия из страны, помечаются в регистре населения как "возможно неактивные" с удалением их постоянного адреса из регистра населения. Эти пометки вносятся в административной порядке Министерством внутренних дел. Все типы "неактивных" резидентов удаляются из "списков избирателей" для парламентских и муниципальных выборов. Вспомогательными источниками, которые были включены в усовершенствованный регистр населения (УРН), являлись:

а) общенациональные файлы страхования (включая медицинское страхование), которые используются для представления права на медицинское страхование и государственные социальные пособия, такие, как детские пособия, пособия по инвалидности и т.д.;

б) перечень электросчетчиков;

с) регистрация транспортных средств.

5. Вспомогательные файлы используются для обновления адресных данных и создания "административных семей". Качество усовершенствованного регистра населения и конкретно качество административных семей все еще находятся в процессе оценки.

V. Качество географической инфраструктуры

6. Оценка географического охвата в совокупности с обновлениями осуществлялась в рамках трех допереписных процессов:

а) были разработаны усовершенствованные средства ориентации в отношении населенных пунктов без зарегистрированных адресов (в основном арабских деревень без названий улиц). Данный процесс осуществлялся в течение 12 месяцев до начала сбора данных; в компьютеризированный слой "строения и улицы" были добавлены межевые знаки;

б) еще одна оценка охвата строений была проведена в течение месяца, предшествующего процессу переписи. Специально нанятые для осуществления

задач обновления и ориентации счетчики вручную внесли в напечатанные бумажные карты межевые знаки. Что касается компьютеризированного слоя, то в 50 населенных пунктах было добавлено 3 694 межевых знака, а в бумажные карты – около 900, причем последние в настоящее время включаются в цифровую географическую базу данных;

с) слой строений потребовал обновлений в отношении новых строений во включенных в выборку ячейках. Выборка определялась на основе цифровой географической базы данных, которая была заморожена за 15 месяцев до начала сбора сведений. Компьютеризированные данные относились к базе данных за предыдущий год. Новые строения, которые не были включены в базу данных на стадии формирования выборки, добавлялись на предшествующей переписи стадии, в ходе которой счетчики переписывали все жилища во включенных в выборку строениях и добавляли строения, которые отсутствовали на бумажных картах. Во включенные в выборку ячейки было добавлено 9 032 строения (5%) (в общей сложности 181 001 строение в исходном перечне). Число новых строений находилось в ожидаемых пределах в заданном временном интервале (за два года), определенных на основе аэрофотоснимков, сделанных до начала фактической переписи. Данный лаг являлся основным предметом озабоченности: еще одной задачей, которую выполнили вручную географические координаторы, стало обновление бумажных карт, подготовленных для решения всех задач, поставленных перед полевым персоналом.

VI. Качество прямого сбора данных (выборочное обследование на неполноту учета)

7. Оценка качества комплексной переписи проводится по пяти измерениям качества и с использованием различных инструментов.

а) Охват - в среднем было выявлено около 5% отсутствующих жилищ в ходе процесса переписи по сравнению с a priori ожидаемым числом домохозяйств на включенных в выборку территориях (от +6% до -21% жилищ в различных населенных пунктах);

Результаты переписи

Опрошенные домохозяйства	83%
Полный вопросник	78%
Частичный вопросник	5%
Жилища, не представившие ответов (закрытые, незанятые или коммерческие)	9%
Отказ	4%

Каждое жилище посещалось как минимум пять раз до определения окончательного статуса.

б) повторный опрос по телефону для оценки надежности (короткие проверочные вопросы) руководителями счетчиков – примерно 5% выборки было повторно опрошено для выяснения того, было ли домохозяйство опрошено лично. Выявленные случаи обмана были редким явлением;

с) онлайн-проверки на непротиворечивость – компьютеризированные вопросники (программное обеспечение BLAISE) содержали комплексные тесты на непротиворечивость. Использовались два типа параметров непро-

тиворечивости: первым являлось "предупреждение", которое позволяло продолжить опрос после ответа на предупредительное замечание, такое как "супруг имеет тот же пол, что и опрашиваемое лицо, просьба проверить" или "проверьте возраст ребенка"; а вторым – "ошибка", которая препятствовала продолжению опроса, например "год иммиграции предшествует году рождения" – проверьте ваши данные. Число "ошибок" и "предупреждений", задокументированных в компьютеризированных вопросниках, еще не было подсчитано;

d) онлайн-оценка достоверности и надежности. Была разработана компьютеризированная система "онлайн-контроля" сбора переписных данных. Потребность в этой системе возникла в ходе последней пилотной переписи (генеральная репетиция в 2006 году), когда возникли подозрения, что некоторые счетчики нашли пути снизить трудозатраты при сохранении полной оплаты труда. Ежедневная передача компьютеризированных данных позволяет практически онлайн-контроль передаваемых данных и работы счетчиков в дополнение к встроенным тестам на непротиворечивость в вопроснике. Данная процедура позволяет выявлять случаи обмана, немедленное сообщение о них вышестоящим органам как на полевом уровне, так и в центральные органы. Ежедневно утром в отношении переданных накануне вечером данных осуществлялся прогон набора заранее определенных запросов. Эти запросы не тестировались до переписи, поскольку необходимость в них и соответствующие возможности были определены после окончательной пилотной переписи. Новые запросы добавлялись в процессе.

Примеры системы онлайн-запросов

<i>Тип запроса</i>	<i>Параметр</i>	<i>Тип проверок</i>	<i>Рабочее определение</i>
Целенаправленное занижение числа членов домохозяйства	Число членов домохозяйства	Сопоставление размера домохозяйства со средним размером домохозяйств, расположенных по соседству	Более чем одно СО от среднего размера домохозяйства
	Частое обновление числа членов домохозяйства	Частое исключение лиц из перечня домохозяйства	2-5% подозрительных >5% невероятных
Информация, не полученная от опрашиваемого	Дублирование идентификационных номеров	Более одной записи с одним и тем же идентификационным номером во всех предыдущих записях одного и того же счетчика	5-10% подозрительных >10% невероятных
	Допустимый ошибочный идентификационный номер	Идентификационный номер отсутствует в регистре населения или принадлежит неактивному лицу	<5% подозрительных >5% невероятных
	Даты рождения с нерегулярным распределением	Распределение каждого компонента даты рождения	30-40% подозрительных >40% невероятных

Пропуск вопросов	Частота "не обучается" или "никогда не обучался"	Число лиц с такими ответами	10-20% подозрительных >20% невероятных
	Частота "не знаю" или "отказался"	Число вопросников с более чем 4 такими ответами	10%-20% подозрительных >20% невероятных
Отсутствующие критические переменные	Отсутствует имя	Процент лиц с отсутствующей переменной	1%-5% подозрительных >5% невероятных
	Отсутствие фамилии	Процент лиц с отсутствующей переменной	1%-5% подозрительных >5% невероятных
	Отсутствует возраст/дата рождения	Процент лиц с отсутствующей переменной	1%-5% подозрительных >5% невероятных

Отчет с указанием имени счетчика, типа обмана, уровня каждого случая обмана, числа вопросников и числа записей передавался в электронной форме региональным руководителям и центральным органам ежедневно. Система была несовершенной, но адекватной и весьма гибкой и, как было уже указано, в процессе добавлялись новые и модифицировались уже существующие запросы;

е) увязка собранных данных со сведениями усовершенствованного регистра населения (УРН) и проверка достоверности результатов для определения качества полученных идентификационных параметров. Результаты увязки иллюстрирует приводимая ниже диаграмма. Увязка собранных счетчиками данных с данными регистра населения имеет чрезвычайно важное значение для методологии комплексной переписи. Увязка априори возможна только в отношении постоянно проживающих жителей Израиля. Успешная увязка была проведена в отношении 98,3% постоянных жителей Израиля, проживающих в домохозяйствах, и в отношении 95,8% жителей Израиля, проживающих в институциональных хозяйствах. Эти результаты доказывают, что в ходе переписи было обеспечено адекватное качество идентификационной информации.

VII. Обследование полноты данных (выборочное обследование на неполноту учета – телефонный опрос); элементы оценки качества

8. Оценка качества в связи с обследованием полноты данных проводилась с использованием различных процедур:

а) первая процедура оценки заключалась в заслушивании выборки счетчиков с речевой записью представителями центральных органов;

б) непрерывный контроль инкогнито по выборке счетчиков, проводивших опрос по телефону, специалистами центральных органов для оценки достоверности и надежности ответов. Результаты этих оценок качества еще не подведены;

с) в-третьих, оценивалась пропорция лиц и домохозяйств, отсутствовавших в ходе опроса. Ряд лиц, не переписанных счетчиками, были найдены по зарегистрированному адресу на этапе опроса по телефону. Результаты представляют собой оценку числа лиц, не опрошенных в ходе переписи. Из

42 337 лиц, представивших ответы в рамках обследования полноты данных (опрос по телефону) и сообщивших в качестве адреса на контрольную дату переписи адрес, по которому они зарегистрированы в регистре населения, примерно 1% лиц не был охвачен переписью. К их числу относятся лица, чье домохозяйство было опрошено, но чьи имена отсутствуют в вопросниках, а также лица, домохозяйство которых не было учтено вообще.

VIII. Автоматизированные процедуры

9. Комплексная перепись опирается на использование автоматизированных процедур, и только остаточные операции производятся вручную. Процессы редактирования и условного расчета являются полностью автоматизированными. Автоматизированные процедуры были разработаны и внедрены с целью обеспечения своевременности и непротиворечивости.

а) Ответы, присваиваемые категориям "другое" на основе текста, введенного в соответствии с устным ответом респондента на вопросник, – специальная автоматизированная процедура, опирающаяся на текстовые стринги, проверяет отсутствие конкретных категорий, которые могли бы быть присвоены. 15% таких категорий "другое" были присвоены конкретные категории с использованием данного процесса;

б) использовалась процедура автоматизированного кодирования экономических отраслей и основного занятия. 75% записей были закодированы в автоматизированном режиме, из которых 5% были перекодированы вручную с целью оценки качества автоматизированного кодирования. Оставшиеся 25% кодировались вручную с использованием компьютеризированной системы кодирования (25 кодировщиков в течение 8 месяцев), причем 10% кодированных вручную записей были случайно перекодированы различными кодировщиками. Несовместимые коды перекодировались экспертами. Результаты оценки качества автоматизированного и ручного кодирования планируется получить в ноябре 2009 года;

с) процедуры автоматизированного кодирования и условного расчета с использованием NIM и CANCEIS в настоящее время применяются в отношении социально-экономической и демографической информации. На совещании ЕЭК ООН в декабре 2009 года будет представлен документ, описывающий соответствующие правила и процедуры.

IX. Выводы

10. Результаты процесса увязки (еще не полностью завершено, поскольку увязка все еще продолжается) в отношении лиц с предполагаемым потенциалом увязки составляют 98,3% в случае обследования домохозяйств и 95,8% в случае обследования учреждений. Такие результаты личного учета являются более высокими, чем ожидаемый нижний уровень в 95% совпадений. Результаты увязки в отношении с трудом поддающихся опросу подгрупп населения (с опорой на результаты предыдущей переписи 1995 года) также находятся в диапазоне 90% совпадений. Предполагается, что данные результаты были достигнуты благодаря проведению масштабной пропагандистской кампании, конкретно ориентированной на каждую из таких подгрупп населения. Комплексная перепись, которая опирается на две крупные выборки (примерно 20% населения) и полный учет институционального населения, позволила сосредоточить усилия на вклю-

ченных в выборку домохозяйств и снизить нагрузку по предоставлению ответов на население.

10. Возможность распространения окончательных итогов через год после проведения переписи обусловлена широким использованием автоматизированных процедур на всех этапах переписных операций. Двойная система оценки позволяет оценивать качество переписи без использования "послепереписного обследования".

График.

Результаты увязки результатов выборочного обследования на неполноту учета

