



---

## **Европейская экономическая комиссия**

### **Конференция европейских статистиков**

#### **Группа экспертов по переписям населения и жилищного фонда**

##### **Семнадцатое совещание**

Женева, 30 сентября – 2 октября 2015 года

Пункт 2 b) предварительной повестки дня

**Инновации, запланированные для раунда  
переписей 2020 года и результаты тестирования**

### **Первый раунд скользящей комплексной переписи Израиля – методология, результаты и недостатки**

#### **Записка Центрального статистического бюро Израиля**

##### *Резюме*

В 2012 году Израиль провел первый раунд плановой скользящей комплексной переписи, контрольной датой для которой стало 31 декабря 2011 года. Этот раунд планировался как экспериментальный, цель которого заключалась в том, чтобы изучить возможность скорейшего перехода к скользящей переписи после переписи 2008 года. Организационное решение было принято до переписи 2008 года, методология была разработана в 2010 году, а реализация была запланирована на 2011 год.

В настоящем документе описываются использовавшаяся методология и полученные результаты.



## I. История вопроса

1. Комплексная перепись (КП) 2008 года позволила получить качественные оценки<sup>1</sup>. Методология КП 2008 года подтвердила обоснованность избранного теоретического подхода и, таким образом, могла повторно использоваться в будущем. При проведении КП 2008 года был применен метод двойных оценок (МДО), который предполагает использование двух разных основ для построения выборки, о чем подробно рассказывалось в предыдущих документах<sup>2,3,4</sup>. В качестве основы для расчета переписных оценок численности населения, демографическо-географических оценок по всей стране в целом и по всем группам населения служил Центральный регистр населения (ЦРН). Эта методика опирается на использование двух не связанных между собой основ выборки для оценки избыточного и неполного охвата ЦРН; территориальной выборки для оценки неполноты охвата и выборки ЦРН для оценки чрезмерности охвата. Регистр населения удалось усовершенствовать, увязав его с различными внешними административными источниками, важнейшими из которых являются:

- данные пограничного контроля, позволяющие отсеять резидентов, находящихся за рубежом на протяжении длительного времени;
- данные национальной системы страхования, позволяющие идентифицировать резидентов, не имеющих право на медицинское обслуживание, что является показателем того, что они более не проживают в стране на постоянной основе.

2. Переход на скользящую комплексную перепись (СКП) вместо проведения второй комплексной переписи (КП) был обусловлен организационными трудностями и стратегическими планами на будущее. Объявленная стратегия заключалась в том, чтобы максимально эффективно использовать знания, комплексную методологию, навыки и опыт, унаследованные от КП 2008 года, с одной стороны, и более своевременно готовить обновленные данные по стране, даже если ежегодные оценки будут иметь более высокие показатели вариации (ВП). Иными словами, СКП ведет к снижению качества данных по небольшим группам населения, но повышает своевременность и актуальность данных о более крупных группах при сохранении того же уровня качества. Этот переход будет сопровождаться движением в направлении полностью регистровой переписи.

3. Основными принципами СКП являются:

- a) сохранение статистической методологии КП;
- b) максимально широкое использование результатов текущих обследований, проводимых Центральным статистическим бюро (ЦСБ);
- c) более широкое использование административных источников;
- d) интеграция процедур переписи в обычные (повседневные) работы ЦСБ;

<sup>1</sup> Центральное статистическое бюро Израиля (2009 год) *Оценки качества комплексной переписи 2008 – Израиль*, Европейская экономическая комиссия Организации Объединенных Наций, ECE/CES/GE.41/2009/12.

<sup>2</sup> H. Glickman, R. Nirel and D. Ben Hur (2003) *False Captures in Capture-Recapture Experiments with Application to Census Adjustment*. Bulletin of the International Statistical Institute, 54<sup>th</sup> Session, Contributed Papers, Vol. LX (2003), pp. 413–414.

<sup>3</sup> R. Nirel, H. Glickman and D. Ben Hur (2004), *A Strategy for a System of Coverage Samples for an Integrated Census*. Proceedings of Statistics Canada Symposium 2003: Challenges in Survey Taking for the Next Decade.

<sup>4</sup> Nirel, R. & Glickman, H. (2009) Chapter 21 – Sample Surveys and Censuses. In: Rao, C.R. (ed.) *Handbook of Statistic*, Elsevier.

- e) децентрализация организационной структуры и сохранение профессиональных кадров и их опыта;
- f) гарантирование бюджета следующей переписи;
- g) уменьшение нагрузки на респондентов благодаря использованию данных, полученных в ходе других проводимых ЦСБ текущих обследований.

4. Первый раунд СКП был проведен в 2012 году. Контрольной датой переписи стало 31 декабря 2011 года. Этот раунд планировался как экспериментальный, цель которого заключалась в том, чтобы изучить возможность скорейшего перехода к скользящей переписи после переписи 2008 года. Организационное решение было принято еще до того, как методология переписи 2008 года была приведена в полное соответствие с требованиями СКП. На момент начала сбора данных (начало 2012 года) методология оценки результатов переписи теоретически еще не была полностью проработана. В 2011 году было проведено небольшое предварительное тестирование, которое коснулось главным образом технологических аспектов. Это было вызвано необходимостью изменить некоторые платформы, которые вышли из использования (например, провести обновление операционных систем). В качестве общеметодологического на вооружение был взят метод двойных оценок (МДО). В соответствии с требованиями к данным переписи, касающимися обследования с избыточным охватом (ИО), необходимые изменения были внесены в методику проведения обследований рабочей силы (ОРС) и обследований расходов домохозяйств (ОРДХ). Данные этих двух обследований должны были использоваться в качестве дополнения при расчете оценок численности населения.

Таблица 1  
**Отличия от процедур 2008 года**

<i>СКП 2011 года</i>	<i>КП 2008 года</i>
Одиннадцатимесячный период сбора данных в рамках обоих обследований (обследование ИО и обследование с неполным охватом (НО))	Девятинедельный период сбора данных
Две независимые выборки	Две зависимые выборки (независимые основы двух выборок)
Основой выборки для переписи НО являлся регистр жилых зданий (РЖС)	Основа выборки – территориальные ячейки (каждая ячейка насчитывает в среднем 50 домохозяйств)
Размер выборки – 10% жилищ/строений	Размер выборки – 20% территориальных ячеек
Одновременный сбор данных в рамках обоих обследований (предполагалась минимальная нагрузка на респондентов)	Сбор данных в рамках ИО после завершения сбора данных в рамках НО и их увязки с данными ЦРН (в целях снижения нагрузки на респондентов и повышения эффективности)
Данные ОРДХ и ОРС увязываются с результатами переписи	Данные обследований используются как вторичный источник для повышения качества анализа результатов переписи

5. В качестве основы выборки для целей НО впервые использовался РЖС<sup>5</sup>.
6. Некоторые изменения были внесены и в базовые процедуры 2008 года, исходя из того понимания, что они окажут минимальное влияние на результаты переписи.

## II. Методология

7. В охват первого раунда переписи вошли лишь муниципалитеты с высоким качеством географических кодов в РЖЗ. Муниципалитеты были разбиты на две основные группы. Выборка муниципалитетов с численностью населения свыше 70 000 человек составляется ежегодно. Все остальные муниципалитеты делятся на отдельные группы по размеру и географическому положению, и по каждой группе ежегодно составляется десятипроцентная выборка, а на местном уровне – десятипроцентная выборка жилищ.

### A. Обследуемые группы

8. В Израиле насчитывается около 3 000 статистических районов (СР). СР эквивалентен «переписному району» и насчитывает в среднем 3 000 жителей (около 1 000 домохозяйств). 1 099 муниципалитетов включают в себя один СР (сельские поселения и населенные пункты, насчитывающие менее 10 000 жителей). 114 муниципалитетов разбиты на несколько СР. «Обследуемая группа» – это население СР, разбитое, как и в ходе КП 2008 года, на четыре возрастные группы: 0–19 лет, 20–29 лет, 30–39 лет, 40+ лет. Эти группы идентичны в плане вероятности быть правильно учтенными в ЦРН. Обследуемые группы, по которым в тот или иной год не составляется выборка, сохраняют веса предыдущего раунда или КП 2008 года. Считается, что веса в этих обследуемых группах быстро не меняются.
9. Вес присваивается каждому человеку в зависимости от его возраста и СР, в котором он постоянно проживает. Чем больше времени проходит с момента проведения переписи 2008 года, тем меньше доверия вырывает предположение о том, что возможные изменения будут минимальными. Исходя из этого будет разработана процедура выявления тех СР, в которых ситуация быстро меняется. Такие СР будут шире представлены в выборке и будут чаще отбираться для выборочного обследования в ходе СМП, поскольку в короткие временные лаги их вес утрачивает значение.
10. Основная логика этой методологии заключается в том, что для более многочисленных групп населения требуются более точные оценки с менее существенной относительной погрешностью.
11. Поскольку изменения в характере учета данных в ЦРН можно игнорировать, в долгосрочном плане СМП должна давать хорошие, достоверные оценки по более многочисленным группам населения и адекватные оценки по меньшим группам.
12. Данные по СР будут калиброваться по данным муниципалитетов, которые считаются более надежными.

---

<sup>5</sup> P. Zadka (2012), *Dwelling and Building Register Based on Municipal Taxation Lists – Quality and Distinctiveness*, United Nations, Economic Council for Europe, ECE/CES/GE.41/2012/15.

## **В. Основа выборки, процедуры формирования выборки и размер выборки**

13. Существуют две основные группы:
- а) Муниципалитеты с числом жителей н 70 000 человек и более (делятся на 20 и более СР).
  - б) Муниципалитеты с числом жителей менее 70 000 человек, которые делятся на 3 подгруппы:
    - i) муниципалитеты с числом жителей менее 10 000 человек (один СР);
    - ii) муниципалитеты, насчитывающие до 10 СР;
    - iii) муниципалитеты, насчитывающие 10–19 СР.
  - с) Для каждой из основных групп выборка формируется в два этапа.
14. В группе А на первом этапе в выборку отбираются 10% СР каждого муниципалитета за каждый год. На втором этапе из каждого вошедшего в выборку СР отбираются 10% жилищ/строений.
15. В группе В на первом этапе в выборку отбираются 10% муниципалитетов из каждой подгруппы за каждый год. На втором этапе из каждого вошедшего в выборку муниципалитета отбираются 10% жилищ/строений. Для каждой группы и подгрупп выборка представляет собой простую случайную выборку. Если  $E$  – вероятность выборки в каждом году, то при  $E = 1\%$  за десять лет она составит порядка 10%.
16. ОРС и ОРДХ ежегодно добавляли к выборке порядка 1,5% домохозяйств (ДХ), что в сумме доводит до 2,5% ДХ ежегодную выборку НО и ежегодную выборку ИО. За десятилетний цикл накопленная выборка для обоих обследований достигнет почти 50% ДХ.
17. Предполагается, что при таком размере выборки ВП для самой малой обследуемой группы (возрастная группа в рамках одного СР) ВП не превысит 10%. Ожидается, что такой размер выборки и такая процедура ее формирования обеспечат более равномерное распределение вошедших в выборку единиц наблюдения в рамках СР.
18. Главными преимуществами такой методики формирования выборки является: а) возможность увеличения размера выборки без дополнительных издержек за счет включения в них жилищ, которые вошли в имеющие такую же основу – РЖЗ – выборки для других текущих обследований; б) более равномерное распределение вошедших в выборку единиц наблюдения по территории СР; и с) сведение к минимуму вероятности того, что одно и то же жилище в течение одного года включено в выборки нескольких обследований.
19. Один из ожидаемых недостатков был связан с возможностью того, что отсутствующие в РЖЗ жилища/строения останутся неохваченными. Доля незарегистрированных жилищ и строений в разных муниципалитетах сильно различается и в некоторых из них может достигать 5%. Предполагалось, что по мере непрерывного совершенствования РЖЗ эта проблема через несколько лет исчезнет.
20. Основой выборки для ИО послужил список «административных домохозяйств» (полные семьи, проживающие по одному адресу, согласно ЦРН), который был составлен на основе ЦРН.
21. В выборку ИО для целей переписи вошли 22 000 жилищ из 41 муниципалитета, половина из которых расположена в муниципалитетах, насчитывающих

свыше 70 000 жителей (на них приходится свыше 40% общей численности населения), а половина – в муниципалитетах, число жителей в которых не достигает 70 000 человек. Выборка НО для целей ОРС включает 18 000 жилищ (равномерно распределенных по всем муниципалитетам), которые были добавлены к выборке переписи. На первом этапе перепись проводилась без увязки с ОРДХ.

22. Выборка НО насчитывала 44 000 административных ДХ (около 2% всех ДХ).

### **С. Усовершенствованный центральный регистр населения (УЦРН)**

$P_t$  – ЦРН по состоянию на 31 декабря года  $t$

$D_t$  – смерти в году  $t$ , зарегистрированные в году  $t+1$

$V_t$  – рождения в году  $t$ , зарегистрированные в году  $t+1$

$I_t$  – иммигранты, въехавшие в страну по иммиграционной визе в году  $t$  и зарегистрированные в году  $t+1$ , вне зависимости от продолжительности их постоянного проживания до контрольной даты переписи

$E_t$  – эмигранты определяются как резиденты, покинувшие страну на срок свыше 90 дней в году  $t-1$  или ранее и не возвращавшиеся на срок не менее 90 дней до конца года  $t$

$M_t$  – резиденты, которые лишились права на получение медицинских услуг после того как было установлено, что они больше постоянно не проживают в Израиле

$D_t'$  – резиденты в возрасте старше 110 лет

$УЦРН_t = P_t - D_t + V_t + I_t - E_t - M_t - D_t'$

### **Д. Переписные листы и технология**

23. Переписные листы были примерно такими же, что и в 2008 году. Из них был исключен вопрос об инвалидности. Вопросы были скорректированы с учетом требований расширенной программы сбора данных.

24. Использовались те же технологии автоматизированного личного опроса и опроса по телефону, что и в ходе КП 2008 года.

25. С учетом методологических изменений, связанных с включением в программу переписи НО, в инфраструктуру ГИС были внесены необходимые коррективы. Для облегчения работы переписчикам выдавались вспомогательные бумажные карты.

## **III. Результаты**

26. Были составлены две выборки: выборка жилищ из РЖЗ для НО и независимая выборка «административных ДХ» из УЦРН для проведения ИО.

27. Для расчета параметров охвата были дополнительно использованы данные обследования рабочей силы.

28. В ходе обследования НО ответы были получены от 63 937 человек, по которым рассчитывался показатель неполноты охвата  $P1+$  обследуемых в рамках СР групп. В ОРС приняли участие 40 680 человек, по которым рассчитывался показатель  $P1+$  для групп, обследуемых в рамках муниципалитетов.

29. В ходе обследования ИО ответы предоставили 72 022 человека, по которым рассчитывался показатель избыточности охвата  $\lambda$  обследуемых в рамках СР групп. Расчет  $\lambda$  для групп, обследуемых в рамках муниципалитетов, проводился на основе ответов 58 399 человек.

30. Два показателя  $P1+$  и  $\lambda$ , рассчитанные в отношении включенных в выборку 2011 года муниципалитетов, были сопоставлены с показателями, рассчитанными в ходе КП 2008 года по каждой обследуемой группе в каждом муниципалитете.

31. Результаты такого сопоставления показали, что в 2011 году 95% показателей  $P1+$  были выше, чем в 2008 году. Коэффициент корреляции двух результатов был высоким,  $R=0.84$ . Это должно было указывать либо на значительное повышение качества регистрации данных в УЦРН, либо на наличие погрешности в параметрах  $P1+$ . Высокий уровень  $P1+$  означает, что в СР проживает меньше людей, неверно учтенных в УЦРН.

32. Если в период между 2008 и 2011 годом регистрация адресов улучшилась, это должно было отразиться на параметре избыточного охвата  $\lambda$ . Высокий показатель  $\lambda$  означает, что в УЦРН сократилось число людей, зарегистрированных в одном, но проживающих в другом СР (своего рода зеркальный параметр из иного источника данных).

33. Однако этого не произошло. В 2011 году параметр  $\lambda$  был выше лишь в случае 42% обследуемых групп в этих муниципалитетах. Сопоставление окончательных результатов по муниципалитетам и СР, включены в выборку 2011 года, с текущими данными о населении за 2011 год (на основе КП 2008 года) показало, что минимальная погрешность для общей численности населения муниципалитетов составляет 5%, а максимальная для СР – 27%.

34. Полученные результаты были досконально изучены созданными для этой цели двумя группами: одна из них была создана для анализа процедур переписи, а другая – для анализа методологии.

## **А. Результаты анализа методологии**

35. Был сделан вывод о существовании сильной корреляции между вероятностью правильной регистрации в УЦРН и вероятностью того, что указанная в РЖЗ фамилия является фамилией реально проживающего в жилище человека. Для оценки уровня зависимости источников были объединены следующие файлы:

- a) УЦРН 2011 года;
- b) РЖЗ 2011 года;
- c) результаты обследования ИО за 2008 год.

36. Полученные результаты показали, что 234 671 учетные записи были связаны между собой и что по итогам обследования был установлен правильный адрес.

Таблица 2  
Перепись 2008 года

	<i>Неверная регистрация в УЦРН</i>	<i>Верная регистрация в УЦРН</i>	<i>Всего</i>
Неверная регистрация в РЖЗ	14 445	17 486	31 931
Верная регистрация в РЖЗ	4 778	197 962	202 740
<b>Итого</b>	<b>19 223</b>	<b>215 448</b>	<b>234 671</b>

– Критерий хи-квадрат = 67 451 при  $P=0.00001$ .

– Фи-коэффициент Крамера ( $\phi C$ ) = 0,54, демонстрирует взаимозависимость двух дискретных переменных (принимает значение от 0 до 1<sup>1</sup>(6), где 0 означает отсутствие связи, а 1 – полную идентичность).

– Соотношение шансов = 34,2 (при отсутствии связи должно быть равно  $\approx 1$ ).

37. В процедурах увязки данных недостатков выявлено не было.

38. Продолжительность переписи никак не сказалась на результатах. Параметры за первые четыре месяца были сопоставлены с параметрами за последние четыре месяца переписи, и в них не было обнаружено никаких различий.

39. Между небольшими и крупными муниципалитетами не было обнаружено различий в размахе погрешности.

40. По сравнению с ОРС, имеющем такую же основу выборки, наблюдалась схожая тенденция указывать жилища как «незанятые» или «не используемые в качестве жилья». В то же время, по сравнению с КП 2008 года эта тенденция была более ярко выраженной.

41. Доля участия респондентов в ИО была довольно низкой (74%). Однако это никак не сказалось на качестве результатов наблюдения.

## **В. Результаты анализа процедур**

42. Возможно, что некоторые счетчики недостаточно строго придерживались подробных инструкций руководства по проведению опросов. Предложение тщательно проверить это предположение принято не было, поскольку это было бы сопряжено с большими трудностями и расходами. Было решено исходить из того, что в качестве главного идентификатора жилища счетчики использовали фамилию владельца, а не другие критерии, которые являются не столь четкими, могли вызвать дополнительные вопросы или не в полной мере соответствовали инструкциям. За единицу выборки было взято жилище, а не зарегистрированный в нем жилец.

43. Доля респондентов, отказавшихся отвечать на вопросы СКП и ОРС, была примерно одинаковой и несколько более высокой, чем во время КП 2008 года.

## **IV. Обсуждение**

44. Единым идентификатором жилища в РЖХ служит муниципальный налоговый номер и адрес строения. В РЖЗ отсутствуют номера жилищ, соответствующие номеру на двери жилища. На практике жилища идентифицируются по фа-



мили на входной двери, которая обычно указывает на жильца – муниципального налогоплательщика. Тот факт, что между двумя регистрами (УЦРН и РЖЗ) существует корреляция, повышает вероятность того, что объектом переписи (по которому был сдан полностью заполненный переписной лист) станет ДХ, правильно зарегистрированное в УЦРН.

45. Все это свидетельствует о том, что до тех пор, пока надежная процедура, позволяющая удостовериться, что счетчик правильно идентифицировал включенное в выборку жилище, отсутствует, РЖЗ в качестве основы выборки не отвечает требованиям МД.

46. Главное методологическое требование к применению МДО, соблюдение которого позволит получать данные без ошибок, заключается в том, чтобы две основы выборки были независимыми и не коррелировали между собой. Наблюдавшаяся между УЦРН и РЖЗ корреляция является одной из главных причин наличия статистических погрешностей.

47. Проведенный анализ позволил сделать вывод о наличии ошибок в идентификации конкретных жилищ на основе данных РЖЗ.

48. Следствием выводов, сделанных по результатам предварительного раунда, стало проведение завершившегося в апреле 2013 года второго раунда СКП и всестороннего методологического анализа методологии.

49. Результаты проведенного анализа свидетельствуют о том, что в качестве основы выборки для проведения обследования НО в рамках СКП регистр жилищ не может служить надежным источником данных о распределении населения ни в рамках обследуемых групп, ни в пределах муниципалитетов. Основа выборки для проведения обследования ИО была той же, что и в ходе КП 2008 года, и она позволила получить результаты без статистических погрешностей.

50. Как уже отмечалось, первоочередной задачей было как можно скорейшее завершение сбора данных с тем, чтобы свести к минимуму задействование ресурсов и высвободить привлеченные к переписи кадры. Подготовка оценок за первый год была завершена в конце марта 2013 года, а сбор данных второго раунда (2012 год) – в конце апреля 2013 года. Итоги углубленного анализа были готовы в конце июня 2013 года.

51. Главный вывод, сделанный по итогам анализа, заключается в том, что основа выборки не удовлетворяет требованиям СКП, проводимой с использованием МДО.

52. По общему мнению, основа выборки для обследования НО должна быть аналогична той, которая использовалась в ходе КП 2008 года.

53. Итоги повторной оценки усовершенствований, внесенных в различные административные источники, которые в ближайшем будущем сделали бы возможным проведение регистровой переписи, говорят об отсутствии любых масштабных подвижек, которые позволили бы перейти к полностью регистровой переписи.

54. Сегодня ведется интенсивная работа по проверке возможности проведения СКП с использованием тех же основ выборки, что и в рамках КП 2008 года, хотя и с некоторыми изменениями. Одно из таких изменений заключается в переходе на двухступенчатую процедуру формирования выборки. Это должно улучшить охват территории СР за счет включения в выборку большего числа «территориальных ячеек». На первом этапе будет составляться 60-процентная выборка территориальных ячеек, затем будет составляться список жилищ в каждой ячейке, на базе которого будет формироваться 30-процентная выборка жилищ в каждой

отобранной для выборки ячейке (это изменение продиктовано представлениями о том, что дисперсия в рамках ячеек меньше, чем между ячейками). Предлагается сократить цикл переписей с десяти до пяти лет и создать систему предоставления ответов через Интернет для обоих обследований (НО ИО) с тем, чтобы повысить коэффициент участия с трудом поддающихся учету групп населения (например, городской молодежи мужского пола). Изучается также возможность разработки показателей, которые будут давать априорную информацию о СР, в которых ожидается высокий уровень дисперсии результатов (что свидетельствует о высокой доле неверных записей в УЦРН по таким СР). Выборочный анализ таких СР будет проводиться чаще и на основе более представительных выборок. За период, прошедший после первого раунда СКП в 2011 году, были внедрены более современные технологии: стали применяться глобальная система местопределения (GPS), электронные карты и планшеты вместо бумажных карт и портативных компьютеров, а также системы предоставления ответов через Интернет.

55. В течение последующих двух лет планируется провести две пилотных переписи (сбор данных в 2016 и 2017 годах). Их цель будет заключаться в тестировании новых процедур и технологий и изучении влияния «скользящей методики» на оценки.

56. Обследование с ИО также нуждается в определенных технических коррективах, поскольку поступило предложение составлять выборку физических лиц на основе УЦРН, а не «административных домохозяйств», и оно требует проверки на практике.

57. По итогам этих двух пилотных переписей будет принято окончательное решение о том, является ли проведение СКП целесообразным и позволят ли такие переписи получать адекватные данные о населении, которых от них ожидают. Корректировка данных с учетом «скользящей» методики будет проводиться с помощью встроенных статистических моделей.

58. Консультативный комитет, который руководил подготовкой этих будущих переписей, предположил, что СКП, несмотря на все свои преимущества, не позволит получать качественные данные. По мнению представителей пользователей в комитете, местным общинам и исследователям разобраться в результатах СКП будет непросто.

59. Вопрос об окончательной контрольной дате переписи и пригодности «скользящего» метода в израильских условиях остается открытым, и его решение будет зависеть от итогов двух пилотных переписей.