

UNESCE

**РУКОВОДСТВО
ПО ИСПОЛЬЗОВАНИЮ
ПРОДОЛЬНЫХ ДАННЫХ
ДЛЯ СТАТИСТИКИ
МИГРАЦИИ**



UNITED NATIONS

**Руководство по
использованию
лонгитюдных
(продольных) данных
для статистики
миграции**

ПРИМЕЧАНИЕ

© 2020 Организация Объединенных Наций

Данная работа находится в открытом доступе при соблюдении Публичной лицензии Creative Commons, созданной для межправительственных организаций, которая доступна по адресу <http://creativecommons.org/licenses/by/3.0/igo/>.

Издатели должны удалить эмблему ООН из своего издания и создать новый дизайн обложки. На переводах должна быть следующая оговорка: «Настоящий труд представляет собой неофициальный перевод, за который издатель несет полную ответственность».

Издателям следует направлять файлы своего издания по электронной почте по адресу permissions@un.org. Обозначения, используемые в настоящем издании, и изложение материала не подразумевают выражения какого-либо мнения со стороны Секретариата Организации Объединенных Наций относительно правового статуса какой-либо страны, территории, города или района, или их властей, или относительно делимитации их границ.

Фотокопирование и воспроизведение отрывков допускается с надлежащим указанием авторства. Настоящее издание выпущено на английском языке.

ECE/CES/STAT/2020/6
ИЗДАНИЕ ОРГАНИЗАЦИИ ОБЪЕДИНЕННЫХ НАЦИЙ
№ для продажи E.20.II.E.46
ISBN 78-92-1-117263-8
eISBN 978-92-1-005489-8
ISSN 0069-8458

Авторское право © Организация Объединенных Наций, 2020 год. Все права защищены по всему миру.
Издание Организации Объединенных Наций, выпущенное Европейской экономической комиссией.

ПРЕДИСЛОВИЕ

По мере роста числа международных мигрантов для общественности и лиц, определяющих политику, становится все более важным понимать миграционные потоки и влияние миграции на отдельных лиц, семьи, общества и экономики. Для этого, как подчеркивается в Глобальном договоре о безопасной, упорядоченной и легальной миграции, необходимо рассматривать изменения с течением времени. Лонгитюдный (продольный) подход, при котором информация собирается с одних и тех же физических лиц или домохозяйств с течением времени, особенно полезен для этого, поскольку он может раскрыть географические и социально-экономические результаты опыта миграции.

Источники данных, доступные многим национальным статистическим управлениям, обладают большим потенциалом для производства лонгитюдной статистики миграции, который на сегодняшний день используется недостаточно. Чтобы улучшить ситуацию, Конференция европейских статистиков создала Целевую группу по использованию лонгитюдных (продольных) данных для статистики миграции. Эта целевая группа экспертов из национальных статистических управлений и международных организаций работала с 2018 по 2020 год над разработкой настоящего Руководства. В нем показано, как развивать источники лонгитюдных данных, как решать различные проблемы, связанные с такими проектами, и как распространять ключевые показатели и выводы.

Руководство было одобрено на 68-м пленарном заседании Конференции европейских статистиков в 2020 году.

ЕЭК ООН благодарна всем экспертам, участвовавшим в подготовке данной публикации.

ВЫРАЖЕНИЕ ПРИЗНАТЕЛЬНОСТИ

Руководство было подготовлено Целевой группой ЕЭК ООН по использованию продольных данных для статистики миграции, в которую входили следующие члены:

Скотт МакЛейш (Канада) — *Председатель Целевой группы*
Штефан Марик-Лебек (Австрия)
Патрик Люсин (Бельгия)
Роуз Эвра, Тристан Кэйн, Елена Прокопенко (Канада)
Гюнтер Брюкнер, Ян Эберле (Германия)
Чинция Конти, Энрико Туччи (Италия)
Бакыткул Утеулина (Казахстан)
Эдгар Виельма Орозко (Мексика)
Хан Николаас (Нидерланды)
Лиана Вологирова (Российская Федерация)
Хорхе Вега Валье, Амор Гонсалес Редондо (Испания)
Флоренс Бартошик (Швейцария)
Алпер Аджар, Нериман Джан Эрган, Шерифе Дилек Йылмаз (Турция)
Никки Роджерс, Луиза Блэквелл, Джастин МакНэлли, Кэтрин Уорли (Великобритания)
Джампаоло Ланцери (Евротат)
Сесиль Торо, Кристоф Дюмон (ОЭСР)
Андрес Викат, Адам Томас (ЕЭК ООН)

Главы Руководства были обсуждены и согласованы всей Целевой группой. Некоторые страны взяли на себя основную ответственность за подготовку определенных глав и разделов, как указано ниже:

Глава 2 — Великобритания и Бельгия
Глава 3 — Канада; раздел 3.2 — Великобритания
Глава 4 — Швейцария и Канада

Целевая группа также выражает признательность за вклад следующих экспертов:

Эдвард Нг, Элен Махо (Канада)
Анна Саара Лигети (Венгрия)
Минья Теа Джамария (Норвегия)

СОДЕРЖАНИЕ

ПРИМЕЧАНИЕ	II
ПРЕДИСЛОВИЕ	III
ВЫРАЖЕНИЕ ПРИЗНАТЕЛЬНОСТИ	IV
ГЛАВА 1 ВВЕДЕНИЕ	1
ГЛАВА 2 ОБЗОР ИСТОЧНИКОВ ПРОДОЛЬНЫХ ДАННЫХ ДЛЯ СТАТИСТИКИ МИГРАЦИИ	3
2.1. Определение	3
2.2. Типы данных	3
2.2.1. Типология	3
2.2.2. Лонгитюдные панельные обследования	5
2.2.3. Современная практика	5
2.2.4. Регистры населения, связывание записей и интеграция других данных	8
2.3. Дополняемость и сопоставимость лонгитюдных и перекрестных измерений	14
2.3.1. Ретроспективный контент в перекрестных обследованиях	14
2.3.2. Синтетические когорты	14
2.3.3. Использование сильных сторон перекрестных и лонгитюдных данных	15
2.3.4. Маломасштабные или лонгитюдные тематические исследования	15
2.4. Резюме	15
ГЛАВА 3 КАК РАЗРАБОТАТЬ НАБОР ПРОДОЛЬНЫХ ДАННЫХ ДЛЯ СТАТИСТИКИ МИГРАЦИИ С ИСПОЛЬЗОВАНИЕМ ИНТЕГРИРОВАННЫХ ДАННЫХ	17
3.1. Введение	17
3.2. Система ошибок для продольных данных из интегрированных источников данных	18
3.3. Этап 1: Статистический дизайн	24
3.3.1. Определение статистического проекта	26
3.3.2. Структурные ограничения, которые могут повлиять на статистический дизайн	27
3.3.3. Определение вариантов дизайна и оптимизация статистической погрешности	28
3.4. Этап 2: Оценка и предварительная обработка исходных файлов	35
3.4.1. Оценка качества исходных файлов	35
3.4.2. Предварительная обработка исходных файлов	39
3.4.3. Отчет о качестве данных исходных файлов	42
3.5. Этап 3: Интеграция данных для лонгитюдных данных	42
3.5.1. Методы интеграции данных	42
3.5.2. Документирование ошибок интеграции	46
3.6. Этап 4: Присвоение лонгитюдных индивидуальных идентификаторов	46
3.6.1. Интеграция данных с общими уникальными идентификаторами	47
3.6.2. Интеграция данных без общих уникальных идентификаторов	49
3.6.3. Семейные идентификаторы	50
3.7. Этап 5: Создание окончательной базы данных	50
3.7.1. Протоколы конфиденциальности	50
3.7.2. Структура базы данных	51
3.7.3. Отбор записей	53
3.7.4. Унифицированный контент и стандартные названия переменных	53
3.8. Этап 6: Распространение результатов	54
3.8.1. Определение правил конфиденциальности для доступа и распространения	54
3.8.2. Окончательная оценка качества базы данных	55
3.8.3. Подготовка словаря данных и технического отчета / руководства пользователя	56
3.9. Резюме	57
ГЛАВА 4 РАСПРОСТРАНЕНИЕ РЕГУЛЯРНОЙ СТАТИСТИКИ МИГРАЦИИ ИЗ ИСТОЧНИКОВ ПРОДОЛЬНЫХ ДАННЫХ	59
4.1. Ключевые лонгитюдные показатели	59
4.1.1. Миграционные модели	62
4.1.2. Социально-экономические результаты	81

4.1.3. Семейная миграция	96
4.2. Передовая практика распространения лонгитюдной статистики	99
4.2.1. Разные пользователи статистики	99
4.2.2. Распространение лонгитюдной статистики миграции	99
ГЛАВА 5 ВЫВОДЫ И РЕКОМЕНДАЦИИ	106
5.1. Выводы	106
5.2. Рекомендации	107
5.3. Направления будущей работы	107
СПИСОК ЛИТЕРАТУРЫ	109

ГЛАВА 1

ВВЕДЕНИЕ

1. По мере роста числа международных мигрантов для общественности и лиц, определяющих политику, становится все более важным понимать миграционные потоки и влияние миграции на отдельных лиц, семьи, общества и экономики. Во многих случаях ключевые вопросы касаются процесса обустройства мигрантов — как долго мигранты остаются в принимающей стране, как они интегрируются с принимающими обществами и как со временем меняются их социально-экономические результаты.
 2. В конечном счете, изучение миграции — это изучение изменений. Это начинается с изменений места жительства, но может расширяться и включать изменения правового или резидентского статуса и изменения социально-экономических результатов. Миграция и обустройство — это процессы, а не состояния, и результаты могут быть краткосрочными или долгосрочными. Кроме того, миграция — это возобновляемое событие, и результаты могут проявляться в различных моделях с течением времени. Лонгитюдный подход можно использовать для определения того, является ли человек мигрантом, и, следовательно, он был бы полезен для статистики прибытий и выбытий мигрантов, чистой миграции, а также для измерения продолжительности пребывания, особенно когда она прерывается множественными пересечениями границ. Из-за этого преходящего характера эти темы хорошо подходят для изучения с помощью лонгитюдного подхода. Необходимость временной основы для статистики миграции была подчеркнута в Глобальном договоре о безопасной, упорядоченной и легальной миграции, который призывает к данным, которые «позволяют эффективно отслеживать и оценивать выполнение обязательств с течением времени».
 3. Традиционные методы сбора лонгитюдных данных (например, когортные исследования и панельные обследования) стало сложнее осуществлять из-за высоких затрат и отсева. Следовательно, страны все чаще обращаются к альтернативным источникам данных для получения лонгитюдных результатов.
 4. Производство статистики во многих странах эволюционировало благодаря увеличению доступности и пригодности административных данных. Все чаще фонды административных данных улучшаются с точки зрения полноты, частоты и качества собираемой информации. В связи со все более широким использованием административных данных и интеграции данных для производства статистики миграции все больше стран могут создавать лонгитюдные наборы данных без чрезмерных затрат. Страны с ограниченными возможностями интеграции административных данных могут использовать альтернативы, такие как сравнение синтетических когорт.
 5. В то время как в прошлом лонгитюдные исследования миграции часто были самостоятельными или проводились на разовой основе, национальные статистические управления (НСУ) начинают все чаще включать их в регулярное производство статистики миграции. На глобальном уровне ряд руководств и рекомендаций Организации Объединенных Наций касается статистики международной миграции и ее конкретных аспектов. Однако в настоящее время не существует международных руководящих принципов о том, как разрабатывать источники лонгитюдных данных из интегрированных данных, как решать различные проблемы, связанные с такими проектами, и как распространять ключевые показатели и выводы.
 6. Использование лонгитюдного подхода для измерения интеграции мигрантов неоднократно обсуждалось на Совместных рабочих совещаниях ЕЭК ООН-Евростат по статистике миграции (например, в 2012, 2014 и 2017 годах). В октябре 2017 года Рабочее совещание рекомендовало продолжить дальнейшую методологическую работу по этой теме, рассмотреть национальную практику и разработать рекомендации, которые будут способствовать международной сопоставимости лонгитюдных данных.
 7. В феврале 2018 года Бюро Конференции европейских статистиков (КЭС) создало Целевую группу ЕЭК ООН по использованию лонгитюдных данных для статистики миграции. Настоящее Руководство содержит результаты работы этой Целевой группы.
-
1. Руководство по измерению международной миграции с помощью переписей населения (Организация Объединенных Наций, 2017 г.), Международные рекомендации по статистике беженцев (EGRIS, 2018 г.), Руководство по системам регистрации актов гражданского состояния и статистики естественного движения населения: управление, функционирование и техническое обслуживание (Организация Объединенных Наций, 2018 г.) и Руководящие принципы законодательной основы для регистрации актов гражданского состояния, статистики естественного движения населения и управления удостоверениями личности (Организация Объединенных Наций, 2019 г.).

8. Настоящее Руководство основывается на недавней публикации ЕЭК ООН «Измерение изменений в социально-экономических характеристиках мигрантов» (ЕЭК ООН, 2015), в которой были проиллюстрированы преимущества использования продольных данных и рекомендовано странам разрабатывать методологии связывания данных для получения лонгитюдных (продольных) наборов данных. Последующее «Руководство по интеграции данных для измерения миграции» (ЕЭК ООН, 2019b) представило несколько примеров, где интеграция различных наборов данных привела к компиляции лонгитюдных данных. Настоящее Руководство предоставляет более конкретную информацию в обеих этих областях — о построении лонгитюдных (продольных) наборов данных и о показателях, которые должны производиться из этих наборов данных.
9. Руководство состоит из трех основных содержательных глав:
- *Глава 2: Обзор источников продольных данных для статистики миграции*
 - *Глава 3: Как разработать набор продольных данных для статистики миграции с использованием интегрированных данных*
 - *Глава 4: Распространение регулярной статистики миграции из источников продольных данных*
10. Главы представляют собой самостоятельные части, и не требуют прочтения в сочетании с другими частями публикации. Важно использовать те части, которые наиболее актуальны для работы.
11. Глава 2 предоставляет обзор источников продольных данных для статистики миграции. В ней содержатся рекомендации по различным типам данных для статистики миграции и приводятся конкретные примеры из различных стран. Специальный раздел содержит рекомендации, специфичные для использования регистров населения для статистики миграции, включая конкретные проблемы и способы их решения.
12. Глава 3 представляет собой руководство по созданию наборов продольных данных для статистики миграции с использованием интегрированных данных. Со ссылкой на работу, проделанную предыдущими целевыми группами, она предоставляет методику следования от статистического дизайна до распространения, включая общие проблемы и конкретные примеры передовой практики из различных стран. Рассматриваемые этапы включают:
- *Статистический дизайн*
 - *Оценка и предварительная обработка исходных файлов*
 - *Интеграция данных для продольных данных*
 - *Присвоение продольных индивидуальных идентификаторов*
 - *Создание окончательной базы данных*
 - *Распространение результатов*
13. Глава 4 включает предлагаемый набор продольных показателей миграции вместе с передовой практикой распространения лонгитюдных (продольных) результатов. Каждый представленный показатель включает краткое описание проблем и передовой практики с иллюстративными примерами из различных стран.
14. Одной из сквозных тем данного Руководства является то, что разработка и использование продольных данных для статистики миграции является сложной и трудной задачей. Однако отдельные главы показывают, как эти ограничения можно преодолеть. Глава 4 подчеркивает, как, несмотря на ограничения результирующей базы данных, все же можно производить ценные показатели. Для решения возникающих и растущих потребностей в лучшем понимании моделей миграции и обустройства международных мигрантов важно, чтобы НСУ понимали доступные источники данных, рассматривали, как можно смягчить или решить ограничения, и предлагали подходы к распространению результатов, которые говорят об этих моделях.
15. Благодаря более широкому использованию интегрированных данных и статистических регистров многие национальные статистические управления обладают огромным объемом неиспользованного потенциала для лонгитюдной статистики миграции. Глава 3 предоставляет рецепт для стран по разработке новых источников продольных данных. Глава 4 предоставляет примеры того, что можно сделать с результатами.
16. В конечном счете, изучение миграции является естественным применением лонгитюдного анализа. Более широкое использование интеграции данных и статистических регистров открывает новые

возможности для разработки и распространения продольных данных для статистики миграции. Хотя есть проблемы, преимущества использования этих новых возможностей гораздо больше. Изучать миграцию — значит изучать изменения, а для изучения изменений продольные данные необходимы.

17. Статистические системы сильно различаются по возможностям построения лонгитюдных наборов данных. Поэтому предоставление достаточных средств и статистического потенциала имеет решающее значение. Настоящее Руководство предоставляет ориентиры для стран, участвующих в Конференции европейских статистиков, с признанием того, что не все страны смогут применять все аспекты рекомендаций.

ГЛАВА 2

ОБЗОР ИСТОЧНИКОВ ПРОДОЛЬНЫХ ДАННЫХ ДЛЯ СТАТИСТИКИ МИГРАЦИИ

2.1 ОПРЕДЕЛЕНИЕ

18. Продольные данные относятся к информации, которая собирается с одних и тех же единиц анализа, таких как физические лица или домохозяйства, с течением времени. Лонгитюдный анализ может уникально и точно описывать индивидуальные траектории во времени. Продольные данные позволяют нам изучать и понимать жизненные события и переходы на протяжении жизненного пути и между поколениями. Это особенно полезно для изучения международной миграции, поскольку обустройство в новой стране — это долгосрочный процесс. Продольные данные могут раскрыть географические и социально-экономические результаты опыта миграции. Продольные данные, собранные в течение длительного периода, позволяют нам понять не только опыт мигрантов, но и опыт их детей.
19. В этой главе описываются различные типы продольных данных для изучения международной миграции. Также рассматриваются перекрестные данные с точки зрения того, как их можно использовать для получения лонгитюдной перспективы миграции и опыта мигрантских групп. Также рассматриваются дополняемость и сопоставимость между лонгитюдными и перекрестными измерениями, а также обсуждаются дизайны, сочетающие оба подхода.
20. Примерами могут служить панельные дизайны, встроенные в перекрестные обследования. Другим примером являются ретроспективные вопросы для получения лонгитюдных свидетельств в рамках дизайна перекрестного обследования. Синтетические когорты в последовательных переписях или опросах в некотором роде имитируют лонгитюдные исследования, но не отслеживают одни и те же случаи с течением времени.
21. Наконец, поскольку многие страны полагаются на регистры населения для расчета показателей миграции, особенности этих регистров, которые делают их хорошо или менее подходящими для оценки лонгитюдных аспектов миграции, также изучаются достаточно подробно.

2.2 ТИПЫ ДАННЫХ

2.2.1 Типология

22. Приведенная ниже диаграмма предоставляет концептуальную типологию для лонгитюдной информации по статистике международной миграции.
23. Как видно из Рисунка 1, данные можно классифицировать по тому, связывают ли они информацию об одних и тех же единицах с течением времени (и, следовательно, являются продольными данными), или по тому, предоставляют ли они некоторую лонгитюдную перспективу. «Интеграция микро-данных» относится к данным, которые связаны на уровне записей с течением времени. «Макроуровневые сравнения» относятся к лонгитюдным перспективам, полученным посредством сравнений временных рядов или с использованием синтетических когорт. В отличие от интеграции микроданных, макроуровневые сравнения не требуют, чтобы одни и те же физические лица выбирались для каждого раунда сбора данных.
24. Методологические подходы располагаются вдоль континуума с различным балансом между использованием данных обследований (окрашены желтым на диаграмме) и административных данных (окрашены синим). Методы, использующие как данные обследований, так и административные данные, окрашены зеленым; они включают случаи интегрированных данных обследований и административных данных, а также гибридов данных обследований/ административных данных. В случае данных обследований национальные статистические управления, как правило, могут контролировать не только то, какие данные собираются, но и время сбора данных (в рамках имеющихся ресурсов). В случае административных источников НСУ имеют меньший контроль над сбором данных и должны максимально согласовывать временные ссылки для достижения желаемых исследовательских или аналитических целей.

РИСУНОК 1: ТИПОЛОГИЯ ПРОДОЛЬНЫХ ДАННЫХ И ПЕРСПЕКТИВЫ ДЛЯ СТАТИСТИКИ МИГРАЦИИ



Источник: Управление национальной статистики Великобритании.

25. Далее в этой главе рассматриваются два аспекта времени, которые требуют тщательного обдумывания при лонгитюдном использовании административных данных. Первый — это операционное использование времени в исследовательском дизайне, что подразумевает, что наблюдения делаются в момент фактического проведения измерительных мероприятий. Второй — это экзогенное проявление времени в административном наборе данных, что подразумевает, что регистрации осуществляются в соответствии с административной практикой, регулированием, инфраструктурой и качеством администрации. Различия между эндогенным использованием и экзогенным проявлением времени необходимо понимать, поскольку они являются источником потенциальной систематической ошибки в данных. Проблема снятия с регистрации является явным примером этого. Это случается, когда эмиграция не фиксируется в административных данных, и многие источники, в частности регистры населения, имеют эту проблему.
26. Новые проспективные панельные исследования предполагают отбор выборки (физических лиц или домохозяйств, но также могут включать, например, предприятия или события). Выборка обследуется, а затем наблюдается с течением времени через регулярные промежутки времени. Продольные данные из панельных обследований также могут быть в форме новых ретроспективных данных. В этом типе данных респондентов спрашивают о событиях, произошедших в прошлом. Проспективный дизайн с ретроспективным базовым исследованием сочетает в себе два типа: ретроспективное базовое исследование в начале, за которым следует проспективное наблюдение с течением времени.
27. Национальные статистические управления все чаще используют административные источники данных для статистического анализа. Анализ результатов мигрантов может быть полностью основан на административных записях, посредством связывания записей для создания лонгитюдных наборов данных и сравнений временных рядов административных данных.

28. Комбинированный подход может принимать ряд форм. Он может брать информацию из обследования в начале и обновлять ее административными записями. В Таблице 1 ниже описано создание продольных данных путем связывания записей переписи и административных записей в Лонгитюдных исследованиях Великобритании. Дальнейшей альтернативой является сочетание перекрестных и продольных дизайнов, как это видно в Обследовании рабочей силы ЕС. Ретроспективные вопросы в рамках перекрестных обследований также дают лонгитюдные сведения.
29. При отсутствии лонгитюдной информации, полученной из повторных измерений одних и тех же физических лиц, лонгитюдная перспектива может быть получена из синтетических когорт. Когорты могут быть определены как «группы людей, марширующих вместе сквозь время»; они, например, могут быть определены с привязкой к их рождению или периоду их прибытия в страну.

2.2.2 Лонгитюдные панельные обследования

30. Проспективное лонгитюдное панельное обследование — это исследовательский дизайн, который намеренно включает наблюдения за одной и той же выборкой или перекрывающимися выборками в разные моменты времени. Хотя они, как правило, предназначены для количественного анализа, они могут включать сбор качественных данных. Период наблюдения может быть любой продолжительности, от данных, собранных в течение нескольких дней, до исследований, охватывающих десятилетия. В то время как лонгитюдные когортные исследования, как правило, сосредотачиваются на возрастной когорте, родившейся в определенный месяц или год, панельные исследования обычно берут свою выборку из полного возрастного диапазона. В примере из Канады в Таблице 1 есть когорта мигрантов, определенная с точки зрения периода их прибытия в Канаду, включающая всех мигрантов в возрасте старше 15 лет.
31. Существуют особые методологические проблемы в использовании лонгитюдных панельных обследований для анализа опыта мигрантов. Общей проблемой для всех сборов лонгитюдных данных является отсев выборки. Это происходит, когда члены выборки либо не отвечают, либо их невозможно найти. Мигранты, как правило, более мобильны географически, чем немигранты, и с ними обычно труднее связаться для последующего наблюдения. Отсев может создавать систематическую ошибку в обследованиях, если характеристики тех, кто выбыл из последующего наблюдения, отличаются от характеристик оставшихся участников обследования.
32. Лонгитюдные панельные обследования, которые берут свою выборку из генеральной совокупности, могут упускать мигрантов, если они составляют небольшое меньшинство. Те, кто включен в выборку, будут подвержены выборочной вариативности. Один из способов преодолеть эту проблему — разработать лонгитюдное панельное обследование, специально ориентированное на мигрантов, как это сделано в Канаде.

2.2.3 Современная практика

33. Ниже приведена диаграмма, категоризирующая современную международную практику сбора продольных данных о мигрантах.
34. Конкретные обследования и административные данные из каждой категории обобщены в Таблице 1.

ТАБЛИЦА 1. ПРИМЕРЫ СОВРЕМЕННОЙ МЕЖДУНАРОДНОЙ ПРАКТИКИ ИСПОЛЬЗОВАНИЯ ПРОДОЛЬНЫХ ДАННЫХ ДЛЯ СТАТИСТИКИ МИГРАЦИИ

Название обследования/ административного источника данных	Описание, включая типы используемых данных, охват и размер выборки	Темы исследований, методы и примеры анализа
<i>Продольные панельные обследования</i>		
Лонгитюдное исследование иммигрантов, прибывших в Канаду (LSIC)	Специально разработан для изучения интеграции новых иммигрантов Включает всех мигрантов в возрасте старше 15 лет, прибывших в Канаду в период с 10.10.2000 по 30.09.2001 Данные были собраны через 6 месяцев, 2 года и 4 года после прибытия мигрантов в Канаду	Двухэтапная стратифицированная выборка обеспечила отражение сезонных особенностей иммиграции в выборке обследования Результаты получены на национальном и региональном уровнях, а более подробные данные доступны для отдельных провинций и крупных городов (где сосредоточены мигранты).

	<p>лет) из школ, в которых обучается не менее 5 иностранных учащихся</p> <p>Охвачено 1437 школ с 31 700 иностранными учащимися. В исследование включено такое же количество итальянских одноклассников, что и в контрольной группе</p> <p>Данные впервые были собраны в 2015 году</p> <p>Также были опрошены учителя и директора школ, что позволило провести многоуровневый анализ на индивидуальном и школьном уровнях</p> <p>Данные были собраны с помощью электронных анкет, которые заполнялись самостоятельно</p> <p>Данные опроса 2015 года ежегодно сопоставляются с данными регистра учащихся и регистра населения</p>	<p>благополучии в школе, дружбе в школе и вне ее, семье, условиях жизни, чувстве принадлежности и гражданственности</p> <p>Регион школы на уровне NUTS2 (Номенклатура территориальных единиц для статистики уровня 2) Поддерживает анализ характеристик с течением времени, сравнение ожиданий на 2015 год и фактических результатов</p>
<p>Лонгитюдное исследование для Англии и Уэльса (LS)</p>	<p>LS объединяет данные переписи и административные данные</p> <p>В него входят более 500 000 человек, обычно проживающих в Англии и Уэльсе в любой момент времени, что составляет 1,1% населения Англии и Уэльса</p> <p>Он связывает переписи 1971, 1981, 1991, 2001 и 2011 годов с информацией о жизненно важных событиях (рождения, смерти, живорождения и мертворождения у выборочных матерей, смерти младенцев, рожденных у выборочных матерей, регистрация случаев рака, иммиграция и эмиграция, повторный въезд после эмиграции и призыв в вооруженные силы) с 1971 года</p> <p>Отбор образцов осуществляется по 4 датам рождения в календаре</p> <p>Включает как мигрантов, так и немигрантов, включая мигрантов, которые прибыли в Великобританию и впоследствии уехали, а также возвращающихся мигрантов, параллельные исследования проводятся в Шотландии и Северной Ирландии</p>	<p>Высококачественная привязка записей, включающая автоматическое и канцелярское сопоставление данных</p> <p>Более подробный анализ местности поддерживается только в тех местах, где мигранты сконцентрированы географически.</p> <p>Диапазон охватываемых переписью тем актуален для различных исследований в области миграции – временной охват поддерживает исследования динамики социально-демографических изменений и изменений состояния здоровья с течением времени, места и между поколениями, для мигрантов и их детей</p> <p>Исследуемые темы включают время, прошедшее с момента иммиграции, и язык, на котором они говорят, влияние международной миграции на жилье в Лондоне, структурная ассимиляция мигрантов, этнические характеристики международных иммигрантов и их потомков, оценка рождаемости мигрантов, смертности мигрантов и их потомков</p>
<p>Обследование в поперечном сечении с использованием панельных элементов</p>		
<p>Микроперепись населения Германии (МС)</p>	<p>МС в Германии – это комбинированное обследование отдельных лиц и домохозяйств, расширенная версия европейской ОРС</p> <p>Оно проводится ежегодно с 1957 года и включает репрезентативную выборку из 1% немецких домохозяйств, включающую примерно 380 000 домохозяйств и 820 000 членов домохозяйств, включая мигрантов и немигрантов, проживающих в Германии</p> <p>С 2017 года мигрантов можно отличить по вопросу о причине их миграции (это, например, позволяет идентифицировать гуманитарных мигрантов)</p> <p>Опрос проводится с использованием выборки по регионам</p> <p>Двухэтапная региональная кластерная выборка включает в себя 102 региона обследования (этап 1) и население в разбивке по полу, возрасту, семейному положению и национальности (этап 2).</p> <p>Каждое домохозяйство остается в рамках выборки в течение 4 последовательных раундов обследования - таким образом,</p>	<p>Большинство вопросов являются обязательными, в том числе о составе семьи и домохозяйства, занятости, доходах, образовании и профессиональной подготовке, а также необязательными. Вопросы для ответов включают вопросы о состоянии здоровья, медицинском страховании, жилищных условиях и пенсионном обеспечении.</p> <p>Большой размер выборки позволяет анализировать такие подгруппы населения, как мигранты</p> <p>Зональная выборка производится в рамках регионального анализа, за участниками не ведется слежка, если они покидают зону выборки</p> <p>Смещение в охвате мигрантов маловероятно до тех пор, пока вероятность включения в выборку микропереписи не будет систематически варьироваться в зависимости от региона (предполагается, что это не так)</p> <p>Весовые коэффициенты обследования – это соотношение соответствующих</p>

	<p>ежегодно меняется 25% от общей выборки, и в течение любых двух лет подряд наблюдается перекрытие в 75%</p> <p>Структура выборки основана на постоянно обновляемом реестре адресов, “пропорциональном пространству”, т.е. адреса выбираются случайным образом в пределах регионов обследования</p> <p>Мигранты и беженцы включаются в список, если они проживают в частных домохозяйствах - беженцы, которые обязаны проживать в учреждениях на время рассмотрения их ходатайства о предоставлении убежища, являются частью «институционализованного населения» и в прошлом были недопредставлены</p>	<p>категорий населения, участвующего в обследовании, и внешнего населения</p>
<p>Обследование рабочей силы Великобритании (LFS)</p>	<p>Британский опрос LFS представляет собой перекрестный опрос с элементами панельного анализа</p> <p>Опрос отслеживает изменения в экономическом статусе участников в период между выборками</p> <p>Выборка охватывает примерно 0,2% населения, выбранного из справочника почтовых индексов. Вопрос идентифицирует мигрантов</p> <p>Панели меняются ежеквартально, так что испытуемые остаются в выборке в течение 15 месяцев</p> <p>Выборка мигрантов высока, поскольку мигранты могут быть очень мобильными – к концу 5-й волны число мигрантов, не прошедших обследование, будет высоким, что повлияет на результаты исследования, если число убывших будет отличаться от числа оставшихся</p>	<p>Лонгитюдный анализ проводится в течение 2 кварталов (сравнение текущего квартала с предыдущим) и 5 кварталов (сравнение текущего квартала с тем же кварталом предыдущего года).</p> <p>ИС ОРС поддерживает некоторые результаты, относящиеся к миграции, в том числе: оценки занятости, безработицы, экономической неактивности и другие статистические данные, связанные с занятостью, для Великобритании, включая оценки активности на рынке труда в разбивке по национальностям и странам рождения</p> <p>Ежегодное обследование населения Великобритании (APS), которое состоит из данных ОРС, дополненных данными по выборке в Англии, Уэльсе и Шотландии (для обеспечения достаточного охвата небольших территорий), позволяет получить ежегодные оценки численности населения Великобритании в разбивке по странам рождения и национальности</p>

2.2.4 Регистры населения, связывание записей и интеграция других данных

35. Помимо лонгитюдных панельных обследований, для лучшего понимания лонгитюдной миграции могут использоваться административные источники данных. В частности, для этого хорошо подходят регистры населения, так как многие страны уже используют их в качестве основного источника для получения моментных показателей миграции.
36. Регистры населения используются несколькими способами для изучения лонгитюдной миграции. На макроуровне временные ряды моментных показателей миграции (например, количество и тип мигрантов) обычно являются частью основного набора статистики миграции, которую национальные статистические управления (НСУ) производят ежегодно.
37. Интеграция микроданных, включающая в себя увязку и последующее наблюдение за отдельными лицами, является более сложной задачей. Увязка индивидуальных записей с течением времени, будь то между обновлениями регистра народонаселения или между регистром народонаселения и другими административными данными, часто представляет проблемы, даже при наличии личного идентификационного номера. Эти ситуации более подробно рассматриваются в главе 3.
38. Кроме того, как объяснялось ранее в этой главе, могут существовать различия между эндогенным использованием времени при планировании исследования, когда проводятся наблюдения, и экзогенным использованием времени в регистре населения, сроках и темпах регистрации. Последнее определяется несколькими факторами, и почти все они находятся вне контроля национальных статистических управлений.

39. Внешние (макроуровневые) временные факторы связаны с организацией и управлением административным источником и изменениями, которые могут в нем происходить с течением времени. Они влияют на источник в целом. Типичные примеры: изменения в законодательстве и определениях, а также изменения в порядке ведения административного регистра. Внешние факторы могут влиять как на сравнения на макроуровне (разрыв в временных рядах), так и на интеграцию микроданных (отслеживание индивидуальных записей).
40. Внутренние (микроуровневые) временные факторы относятся к проблемам, которые влияют в большей степени на отдельные записи, а не на регистр в целом. Это часто структурные ограничения, связанные с тем, как управляются регистрации, что в некоторых случаях затрудняет отслеживание отдельных лиц с течением времени. Проблемой является взаимосвязь между базовым временным ориентиром регистраций и временным ориентиром производимой статистики. Примером являются эмигранты, которые не снимаются с учета. Хотя эти факторы усложняют интеграцию микроданных, они также могут ухудшать сравнения на макроуровне, особенно если они происходят с некоторой частотой или приводят к неслучайному смещению.
41. Ряд этих вопросов, влияющих на временные измерения, как внешних, так и внутренних, рассматривается ниже более подробно.

2.2.4.1 Внешние временные факторы

42. Во многих странах муниципалитеты или местные администрации являются основными ответственными органами за ведение регистров населения. В некоторых странах также есть центральный регистр населения, но степень централизации, технические особенности и юридические вопросы, касающиеся доступа, сильно различаются между странами. Существует множество аспектов централизации местных регистров населения, которые прямо или косвенно влияют на то, как и в какой степени регистры населения могут быть использованы для производства качественной статистики.
43. На самом базовом оперативном уровне крайне важно наличие согласованных правил и руководящих указаний о том, как регистрировать людей. В большинстве стран существует законодательство, регулирующее этот вопрос, однако это не гарантирует, что регистрация будет осуществляться везде одинаково. Степень соблюдения администрациями этих правил, которая может варьироваться в зависимости от страны или в пределах одной страны, определяет качество регистров и, следовательно, статистических данных, основанных на них. Языковые барьеры, иностранные названия и культурные предубеждения в интерпретации понятий (например, определение домохозяйства) создают дополнительные сложности.
44. Крупные администрации могут быть способны справиться с наплывом мигрантов, в то время как небольшие города и муниципалитеты часто не имеют средств для быстрой адаптации. Например, значительное увеличение числа беженцев в Европе в 2015 году иллюстрирует, как качество регистрации может меняться с течением времени, потенциально влияя на показатели миграции в дальнейшем, поскольку некоторые администрации были перегружены огромным количеством мигрантов, которых им пришлось регистрировать (Eberle, 2018).
45. Помимо операционных вопросов, на пригодность регистра для производства демографической статистики влияет и то, как с институциональной точки зрения организована регистрация населения. Как уже упоминалось, местные власти почти всегда несут ответственность за ведение регистров населения, но то, как организовано участие властей более высокого уровня, в разных странах устроено по-разному.
46. Во многих странах министерство внутренних дел отвечает за координацию и централизацию местных регистров (Poulain et al, 2006), а также другие департаменты, независимо от того, входят они в состав министерства или нет, которые могут взять на себя ответственность за определенные аспекты работы. Во многих странах существуют отдельные процедуры или ведутся специальные регистры для таких субпопуляций, как иностранцы или лица, ищущие убежища. Различия в процедурах, управлении и хранении между администрациями усложняют интеграцию информации, делая хорошее взаимодействие между органами власти крайне важным.
47. Кроме того, последствия может иметь и институциональное положение самого НСУ. То, насколько «гладко» НСУ получает доступ к данным, может зависеть от институциональных отношений между

НСУ и органами, управляющими регистрами населения. Также в некоторых странах на НСУ возложена юридическая обязанность рассчитывать численность населения, служащую исходными данными для ряда финансовых параметров (в частности, для финансирования муниципалитетов). С одной стороны, это облегчает доступ, но, с другой стороны, это может создать нагрузку на НСУ, вынуждая его расставлять приоритеты в выполнении своих нестатистических обязанностей. Даже конституционное устройство государства может играть роль. Не каждая страна организована иерархически, и заставить неиерархические субъекты соблюдать стандарты предоставления (или сбора) данных может быть крайне сложно.

48. Даже между странами с централизованным регистром населения могут существовать различия в доступе к переменным, необходимым для построения лонгитюдных показателей, таким как персональные идентификационные номера, имена и точные даты. Наконец, периодичность доступных данных является решающей. Для лонгитюдных показателей надежное предоставление данных из регистров населения в НСУ как минимум на ежегодной основе является минимальным требованием. НСУ, возможно, придется самим проявлять инициативу, чтобы обеспечить заключение и выполнение соглашений об обмене данными с агентствами, отвечающими за источники.
49. Другой важный фактор — объем обработки, которую НСУ необходимо применить к данным после их получения. Это включает в себя такую работу, как проведение контроля качества, дополнительное слияние записей и централизация самого регистра.
50. В некоторых странах регистр очень богат информацией, полностью централизован и отличается высоким качеством. Однако в других странах НСУ должны находить другие средства для создания стабильного регистра. Это может означать выполнение относительно простой работы, такой как контроль качества или очистка переменных, или более сложные мероприятия, такие как интеграция дополнительной информации из других баз данных. В некоторых странах просто не существует «централизующего» органа, кроме НСУ, которому, следовательно, приходится брать на себя ответственность за интеграцию всех регистров населения.
51. Как правило, статистикам почти всегда необходимо работать с «сырыми» данными. С концептуальной точки зрения эта работа подразумевает преобразование регистра из административной базы данных в статистический регистр. Как объяснялось выше, объем этой фазы может значительно различаться в разных странах, но результат всегда должен быть довольно схожим: стабильная консолидированная база данных, отвечающая минимальному числу базовых предпосылок. Одним из примеров этого является необходимость наличия последовательного идентификатора записи с течением времени, который подробно рассматривается в разделе 3.6.

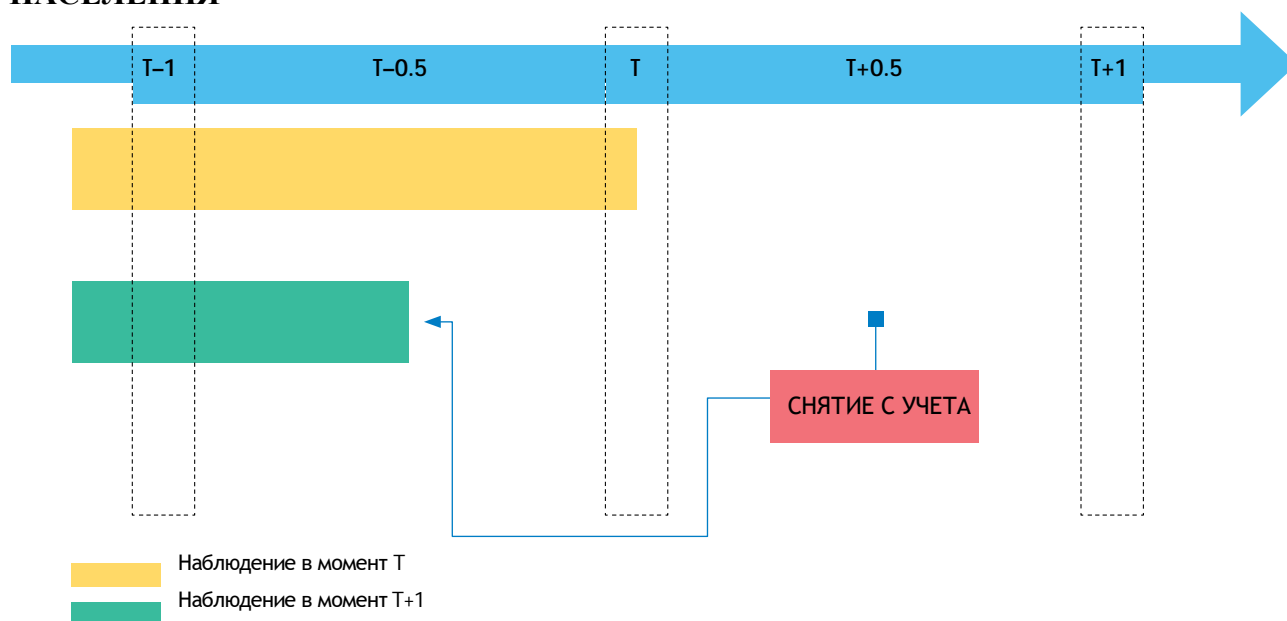
2.2.4.2 Внутренние временные факторы

52. Внутренние факторы – это проблемы, которые со временем затрудняют отслеживание отдельных случаев или записей. Многие из них связаны не только с тем, «как» рассматриваются дела (например, ошибочная регистрация), но и с тем, «когда» они рассматриваются и к какому моменту относятся.
53. Регистрация может относиться к самому событию или ко времени регистрации. На практике между ними могут существовать большие различия. Но даже время события как таковое может относиться к разным моментам, связанным с одним событием.
54. В первую очередь, в случае иммиграции часто существует разница между моментом въезда лица в страну и датой его или ее регистрации в регистре населения. В большинстве стран по этому поводу существуют строгие правила, и регистрация является обязательной в течение определенного периода времени, часто максимум несколько недель или месяцев. В случае измерения долгосрочной миграции эти несколько недель или месяцев не имеют большого значения. Однако в случае измерения миграции на ежегодной основе и превращения этих ежегодных измерений в ряд это вполне может оказать влияние.
55. Наиболее явным примером этого снова стало значительное увеличение числа беженцев в Европе в 2015 году. Число беженцев достигло пика в конце лета и начале осени 2015 года. Однако многие из этих беженцев были зарегистрированы лишь спустя недели или даже месяцы после въезда в Европу, причем те, кто прибыл позже в году, могли быть зарегистрированы в 2016 году. Как следствие, пик регистраций наступил через несколько недель или месяцев после пика самого явления, и в некоторых странах 2016 год в исторической перспективе окажется годом, когда миграция стремительно росла, в то время как на самом деле она уже прошла свой пик, если не снижалась быстрыми темпами.

56. Более того, степень, в которой это имеет место, зависит не только от того, как проводились эти регистрации, например, является ли ретроспективная регистрация административно возможной или разрешенной.
57. Степень, в которой это имеет место, также зависит от того, как проводились эти регистрации и возможна или разрешена ли ретроспективная регистрация. Также решающее значение имеет степень, в которой статистики имеют доступ к этим датам. Например, в Бельгии для лиц, ищущих убежища, регистрируются как дата прибытия, так и дата подачи заявления, но Статистическое управление Бельгии имеет доступ только к последней.
58. Даже просто определение того, кто является мигрантом, может влиять на то, как измеряется момент начала миграции. Некоторые мигранты, например, изначально въезжают в качестве студентов, краткосрочных рабочих или даже туристов. Некоторые делают это несколько раз, прежде чем окончательно решиться поселиться. Тогда возникает вопрос, когда начинать считать их мигрантами — с момента их первой регистрации или когда их правовой статус изменился. Этот вопрос далее обсуждается в главе 4 при определении того, как измерять показатели, связанные со временем.
59. Более того, при определении миграции почти всегда присутствует явное или неявное указание на продолжительность пребывания. Иногда в определении явно упоминается минимальная продолжительность пребывания, или даже предполагается намерение оставаться на минимальный период. В других определениях, однако, критерий продолжительности не применяется, и каждый иностранец, который регистрируется, считается иммигрантом.
60. Лонгитюдное измерение миграции и связанная с ним оценка времени могут играть роль в определении миграции как таковой. Однако определение миграции может влиять на то, как необходимо регистрировать время и как использовать его в анализе. Если, например, короткие периоды пребывания в стране необходимо считать миграциями, эти короткие периоды также должны надежно регистрироваться. С другой стороны, если используется 12-месячный критерий, должна быть возможность отслеживать лиц в течение как минимум года, независимо от того, прерывались ли эти 12 месяцев короткими периодами пребывания за границей. Для этого должна быть возможность следить за одними и теми же лицами с течением времени.
61. Наконец, люди не всегда тщательно соблюдают административные правила, такие как сроки регистрации, даже если они юридически обязательны. В случае первичной регистрации это может быть менее проблематично, потому что у новых иммигрантов мотивация к соблюдению обычно выше. Однако в случае регистрации после периода отсутствия в стране эта мотивация может быть ниже, поскольку административные или юридические последствия менее значительны. Это касается не только иностранцев, но и возвращающихся граждан, возможно, даже в большей степени, поскольку стимулы для немедленной регистрации после возвращения могут быть еще меньше. Как следствие, регистрация может произойти через значительное время после возвращения и только потому, что начинают проявляться первые административные последствия отсутствия регистрации. Время, проведенное за границей, будет тогда искажено.
62. Однако все вышеперечисленное является лишь незначительными проблемами по сравнению с частой ситуацией, когда лицо, эмигрирующее из страны, просто уезжает без дальнейшего уведомления. Это, очевидно, имеет важное влияние на показатели миграции. Это занижает показатели эмиграции и завышает показатели чистой миграции, и, следовательно, с этим следует бороться прямо или косвенно.
63. Самый прямой способ сделать это — «просто» выявить эти случаи и удалить их из регистра населения. Существуют различные способы осуществления подобной идентификации, и практика может различаться в зависимости от того, вмешивается ли администрация, НСУ или оба учреждения.
64. Например, в Бельгии, Италии, Нидерландах и Испании практика, директивы или законодательство позволяют муниципальным администрациям снимать с учета физлицо, как только появляются доказательства того, что он или она больше не проживает в стране. Обычно для запуска данного процесса служат административные сигналы, такие как регистрация новых арендаторов в бывшем жилье эмигранта или невыполнение определенных административных обязанностей. Как правило, муниципальные администрации не обладают ни знаниями, ни средствами для интеграции различных источников данных, поэтому обычно они будут действовать на основе информации, находящейся в самом регистре населения или в других муниципальных регистрах.

65. Однако для большинства НСУ интеграция информации из различных источников является просто основной деятельностью. Как следствие, привязка регистра населения к налоговым регистрам или информации о пограничном контроле может значительно оптимизировать поиск случаев эмиграции, которые остались незамеченными. Насколько НСУ может или юридически вправе делиться информацией с администрациями, зависит от страны. Поскольку конфиденциальность является одним из основных принципов государственной статистики, в некоторых странах может действовать законодательство, запрещающее это.
66. На рисунке 3 представлен пример типичного случая последствий эмиграции, которые не учитываются в рамках простого лонгитюдного исследования с тремя точками измерения T-1, T и T+1. Для большей ясности они рассматриваются как ежегодные данные о численности населения, но возникшие проблемы могут возникнуть и при работе с другими административными данными.
67. В данном случае речь идет о человеке, который въехал в страну незадолго до T-1. В момент времени T это лицо все еще фигурирует в реестре или административных данных, поэтому продолжительность его пребывания на данный момент соответствует желтому прямоугольнику. Однако между T и T+1 становится ясно, что этого человека больше нет в стране и что он покинул страну где-то между T-1 и T. Таким образом, истинный статус и продолжительность пребывания этого человека в момент времени T соответствовали не желтому прямоугольнику, а зеленому.
68. Последствия этой ошибки для наблюдения в момент T очевидны: численность населения будет переоценена, а показатель эмиграции между T-1 и T будет недооценен. Однако проблема также возникает в момент T+1, поскольку лицо больше не будет добавляться к подсчету населения без соответствующей эмиграции, чтобы это сбалансировать.
69. Пример на рисунке 3 также объясняет разницу между схемами лонгитюдных исследований и административными данными, касающимися управления и концептуализации времени в этих схемах исследований. При проведении лонгитюдного обследования описанная выше ситуация и ее неблагоприятные последствия в момент времени T просто не имели бы места, поскольку человек, как правило, больше не имел бы права на последующее наблюдение из-за отсутствия контакта с ним.
70. Обследования являются первичными источниками статистических данных. Они разрабатываются исследователями исключительно для статистического использования, и, как таковые, время управляется внутренне как часть их дизайна. Обычно это подразумевает, что наблюдения совпадают с точками измерения исследования с течением времени, как, например, в случае с волнами в панельных обследованиях.

РИСУНОК 3: НАБЛЮДЕНИЕ ЗА ПРИСУТВИЕМ ЛИЦА В РЕГИСТРЕ НАСЕЛЕНИЯ



Источник: Статистическое управление Бельгии.

71. Административные данные служат другим целям, кроме статистики. Следовательно, проявление времени в административной базе данных будет совпадать с ее административной целью, и это будет главным образом зависеть, среди прочего, от регулирования, практики, инфраструктуры и качества администрации. Это, конечно, факторы, экзогенные для дизайна исследования, и они подразумевают, что почти всегда наблюдения больше не будут полностью совпадать с точками измерения дизайна исследования, как в случае на Рисунке 3.
72. Следствием этого является то, что в отличие от лонгитюдных обследований, при использовании административных данных для лонгитюдного исследования релевантны два измерения времени: (1) эндогенное использование времени в дизайне исследования (когда необходимо производить показатели); и (2) экзогенное проявление времени в административной базе данных (когда регистрируются миграции). Более того, как объяснялось на протяжении всего текста, смещение часто является результатом расхождений между этими двумя. Это, конечно, недостаток использования административных данных для лонгитюдного исследования, и с ним необходимо бороться или, по крайней мере, смягчать его воздействие.
73. Для сравнений на макроуровне способами решения этой проблемы являются либо моделирование, либо коррекция. Все эти меры не являются взаимоисключающими, и многие НСУ используют несколько из них для взаимодополнения.
74. Самый простой способ смоделировать влияние ошибки, представленной на Рисунке 3, – это оценить количество аналогичных случаев за период с $T-1$ по T . Это оценочное число затем может быть вычтено из числа населения в момент времени T и добавлено к числу эмигрантов в период между $T-1$ и T . Другой способ смоделировать эту проблему – просто рассмотреть этот случай как эмиграцию в момент снятия с регистрации и, следовательно, добавить его к показателю эмиграции в период между T и $T+1$. Это также своего рода моделирование, поскольку основная гипотеза заключается в том, что то, что восстанавливается из более ранних периодов, компенсирует то, что было упущено в течение исследуемого периода. Эффективность обоих способов моделирования зависит от стабильности лежащих в их основе явлений.
75. Самый простой способ исправить ситуацию, показанную на Рисунке 3, или, по крайней мере, теоретически – это ретроспективно скорректировать в момент времени $T+1$ оценку численности населения в момент времени T и показатель эмиграции за период с $T-1$ по T . Это гарантирует, что показатели будут согласованными на протяжении всего периода наблюдения. Само собой разумеется, что с практической точки зрения это не простое решение. Второй способ "исправления" – просто ничего не делать и игнорировать ситуацию. Затем этот случай будет учтен при статистической коррективке в момент времени $T+1$.
76. В случае интеграции микроданных решение этих проблем, связанных со временем, является гораздо более сложной задачей. Хотя обнаружение и исправление ошибок, подобных представленным на рисунке 3, является простым делом, устранение их последствий является сложной задачей. Технически, исправление случая на Рисунке 3 приводит к появлению в момент времени $T+1$ новой версии базы данных по сравнению с моментом времени T . Как следствие, все зависимости, которые были действительны в момент времени T (например, связи с другими базами данных), становятся недействительными.
77. Управление версиями базы данных может рассматриваться как первоочередное решение этих проблем. Однако проблема заключается в том, что в принципе каждое новое обновление источника данных может привести к появлению новой версии исследовательской базы данных. Некоторые НСУ получают обновления из регистра населения еженедельно или даже ежедневно, что делает это решение несколько непрактичным, если его тщательно внедрять.
78. В зависимости от степени проблем, возникающих в период между обновлениями, может оказаться приемлемым, например, выпускать ежегодную версию исследовательской базы данных. Информация для отдельной записи или лица, относящаяся к определенному моменту времени, может отличаться от последующих версий базы данных, но изменения, по крайней мере, поддаются идентификации, могут быть проанализированы и должны быть хотя бы частично задокументированы.

79. Другой мерой, которую можно принять, является наложение временного ориентира статистики на неявный временной ориентир источника. Это означало бы, что для примера на Рисунке 3 лицо снимается с учета только тогда, когда регистр населения или НСУ обнаруживает ошибку (в момент $T+1$). Концептуально это означает, что время события устанавливается равным времени регистрации. Это, конечно, вносит смещение, но оно значительно облегчает управление лонгитюдной базой данных.
80. Наконец, для оптимизации управления версиями базы данных могут быть применены более совершенные методы управления данными. Например, декомпозиция базы данных может быть очень полезной. Некоторые характеристики могут меняться в жизни человека довольно часто (например, адрес проживания), другие - редко или никогда (гражданство, пол, дата рождения). Поскольку каждая часть информации хранится отдельно в таблицах меньшего размера, скорость обновления характеристики, которая изменяется больше всего, не определяет, сколько раз обновляется вся база данных. Хотя изменение приводит к созданию новой версии соответствующей подтаблицы, это не означает создания новой версии всей базы данных.
81. В дополнение к этому, создание двухвременных баз данных (Johnston & Weiss, 2010; Johnston, 2014) – это практика, которая помогает контролировать их временное измерение. Довольно простой способ ввести базовую форму битемпоральности – это добавить, в дополнение к переменной даты, относящейся к информации в записи как таковой, вторую переменную даты, относящуюся к действительности записи.
82. В двухвременной базе данных случай, показанный на Рисунке 3, появился бы в базе данных по крайней мере два раза. Один раз с записью А, отражающей (ошибочно) время ожидания до времени T , и второй раз с записью В, отражающей истинную ситуацию (время ожидания до времени $T-0,5$).
83. Вместо того, чтобы предоставлять информацию о фактическом присутствии или отсутствии данного лица, вторая переменная времени или даты предоставила бы информацию о действительности этих двух записей. Первая запись (А) действительна, по крайней мере, до времени T (а на самом деле, до времени $T+1$, поскольку между T и $T+1$ нет измерений), вторая запись действительна с момента $T+1$ и далее. На рисунке ниже представлена выдержка из того, как могла бы выглядеть таблица при применении базовой двухвременности (битемпоральности) к случаю, показанному на Рисунке 1.

ТАБЛИЦА 2: ПРИМЕР БАЗОВОЙ ДВУХВРЕМЕННОЙ ТАБЛИЦЫ, ОТРАЖАЮЩЕЙ СЛУЧАЙ ИЗ РИСУНКА 3

ID	ПРИСУТСТВИЕ	DT_STRT	DT_STOP	DT_STRT_VLDT	DT_STOP_VLDT
A	1	T-1.5	T+1	T-1	T+1
A	1	T-1.5	T-0.5	T+1	31дек9999

84. Преимущество введения битемпоральности заключается в том, что оно позволяет воспроизвести любую более раннюю версию базы данных путем выбора по второй переменной даты, отражающей действительность записей. В принципе, это делает больше ненужным хранение слегка отличающихся версий одной и той же базы данных. Это также упрощает управление связанными со временем зависимостями между характеристиками, многие из которых имеют свой собственный цикл обновления.
85. В Бельгии существует более развернутый пример (полу-) битемпорального способа управления поступлением данных из регистра населения и их использования для создания концептуальных рамок населения.

2.3 ДОПОЛНЯЕМОСТЬ И СОПОСТАВИМОСТЬ ЛОНГИТЮДНЫХ И ПЕРЕКРЕСТНЫХ ИЗМЕРЕНИЙ

2.3.1 Ретроспективный контент в перекрестных обследованиях

86. Перекрестные источники данных, содержащие социально-демографическую и социально-экономическую информацию, как правило, более доступны, чем лонгитюдные источники. Они могут быть использованы для получения лонгитюдных представлений и перспектив в тех случаях, когда действительно лонгитюдная информация с повторными измерениями для одних и тех же респондентов или случаев недоступна. Это достигается либо за счет повторения одних и тех же вопросов в разных циклах поперечного измерения, либо за счет ретроспективных вопросов в рамках одного цикла.
87. Некоторые перекрестные обследования задают респондентам ретроспективные вопросы о том, как и когда характеристика менялась с течением времени. Например, когда респондент впервые или в последний раз прибыл в страну, работа, занимаемая до миграции, или время, которое потребовалось мигрантам, чтобы получить первую работу после миграции. Эту информацию можно проанализировать, чтобы получить информацию о динамических характеристиках процессов адаптации мигрантов.
88. Ретроспективные, связанные со временем переменные позволяют нам анализировать современную информацию в источнике моментных данных в контексте исторических характеристик миграции. Информация о годе прибытия может отличаться от года последнего прибытия, хотя обе могут быть использованы для предоставления информации о том, как социально-экономические характеристики, такие как занятость, различаются у мигрантов с разной продолжительностью пребывания в принимающей стране. Данные о стране рождения вместе с информацией о стране рождения родителей позволили бы проводить отдельные сравнения между первым, вторым, третьим или более высокими поколениями мигрантов. При интерпретации информации о стране рождения необходима осторожность, учитывая изменения международных границ с течением времени.
89. Когорты могут быть определены в соответствии с любой общей характеристикой, и анализ как моментных, так и лонгитюдных данных по группам мигрантов должен учитывать как возраст с момента рождения, так и продолжительность с момента прибытия. Оба оказывают изменяющееся во времени влияние на социально-экономические результаты и взаимодействуют вместе, определяя траектории мигрантов.
90. Используя переменные, которые идентифицируют мигрантов (по их статусу иностранного происхождения или статусу их родителей), наряду с социально-экономическими условиями, наблюдаемыми в различных источниках, можно проанализировать тенденции, которые показали бы изменение распространенности той или иной характеристики с течением времени. Этот показатель отражает чистое изменение этой характеристики для контингента мигрантов, представленного в каждом перекрестном источнике. Одно из ограничений этого подхода связано с тем фактом, что численность мигрантов, как правило, динамична, и их состав, вероятно, будет отличаться при каждом перекрестном снимке. В отсутствие связанных микроданных нелегко отделить социально-экономические изменения от изменений в составе населения.

2.3.2 Синтетические когорты

91. Этот тип анализа идентифицирует данную когорту в разных перекрестных источниках и сравнивает социально-экономические условия этой группы с течением времени. Это не реальные когорты, так как они, вероятно, состоят из разных лиц в разных источниках. Существует риск того, что различия в составе между сравниваемыми когортами внесут смещение в сравнение. Такой подход был использован Германией при проведении Немецкой микропереписи.

2.3.3 Использование сильных сторон перекрестных и продольных данных

92. Возможно, удастся «позаимствовать силу» у перекрестных и продольных источников. Некоторые конкретные применения термина «устойчивость к заимствованиям» включают эмпирическую байесовскую оценку (см., например, Burke et al., 2016) и оценку небольших площадей (см. Chambers et al., 2009 о прочности к заимствованиям в масштабах пространства). Джексон и др. (2017) стремятся определить заимствование силы математически. Использование метода обобщенной оценки с сохранением структуры (GSPREE) со временем «зарекомендовало себя», поддерживая подготовку

оценок на основе переписи за пределами переписных лет за счет сочетания административных данных и данных обследований с данными переписи (см. Корреа-Онель и др., 2016).

-- --

2 Доступно по адресу https://ec.europa.eu/eurostat/cros/content/maintenance-rich-frame-using-data-platform-example-statistics-belgium-0_en

93. Предлагается более простое использование этого термина, которое предполагает объединение информации из разрозненных источников при условии, что объединяемые данные имеют взаимосвязанные и релевантные характеристики. При использовании данных с продольной привязкой может оказаться полезным получать экологические данные, такие как характеристики района, отрасли или профессии, из других соответствующих источников с достаточным объемом выборки, таких как перепись населения, и применять их к продольному обследованию или набору административных данных.

2.3.4 Маломасштабные или лонгитюдные тематические исследования

94. Национальные статистические управления могут пожелать рассмотреть маломасштабные или лонгитюдные тематические исследования для поддержки использования лонгитюдно связанных административных записей. Это можно рассматривать как способ обогатить или проверить достоверность связанных административных данных, например, по возвратной миграции, статусу занятости мигрантов или чтобы дать больше информации о качестве связанных данных. Часто использование административных источников данных для решения конкретных исследовательских вопросов будет включать некоторую форму компромисса, если существует разрыв между информацией, которую вы ищете, и информацией, которая была собрана в административных данных. В этом сценарии анализируемые данные могут предоставлять индикаторы лежащих в основе процессов или результатов или использоваться для проверки достоверности связанных административных данных через параллельное маломасштабное исследование. Это позволило бы сравнить целевые концепции и те, которые были получены на основании административных источников данных. Глава 3 описывает этот потенциальный разрыв более подробно.

2.3 РЕЗЮМЕ

95. Продольные данные позволяют нам изучать и понимать жизненные события и переходы во времени и, что особенно важно, между поколениями. Это может быть особенно полезно для изучения международной миграции и расселения мигрантов, поскольку лонгитюдные данные могут пролить свет на географические и социально-экономические последствия миграции. В этой главе представлен обзор различных источников лонгитюдных данных и приведены примеры для конкретных стран, включая использование лонгитюдных панельных обследований, административных данных, включая регистры населения, а также сочетание обследований и административных данных.

96. Не все НСУ будут иметь доступ к административным данным или финансированию для проведения лонгитюдных исследований о мигрантах, поэтому, возможно, потребуется более прагматичный подход. Например, перекрестные опросы могут дать лонгитюдную информацию и перспективы с течением времени, а к существующим опросам можно добавить ретроспективные вопросы. В отсутствие лонгитюдных данных также можно рассмотреть возможность составления синтетических когорт с течением времени, все из которых довольно подробно описаны в этой главе. Для оценки и дополнения существующих источников некоторые НСУ могут также иметь возможность использовать большие данные.

97. При наличии продольных данных учитываются проблемы и ограничения, связанные с этими данными, а также тот факт, что они будут различаться в зависимости от источника данных. Традиционные методы, такие как лонгитюдные панельные обследования, имеют преимущества в плане контроля, предлагаемого статистическим управлениям, такие как сроки, охват и измерение. Однако, поскольку мигранты, как правило, более мобильны в географическом отношении, чем немигранты, и с ними, как правило, труднее связаться для последующего наблюдения, эти источники данных, как правило, имеют меньший объем выборки, страдают от истощения и могут не обеспечивать регулярность сбора данных. Все это может препятствовать их использованию для изучения краткосрочных или долгосрочных результатов для мигрантов.

98. Существуют также ограничения, связанные с использованием административных данных, таких как регистры населения, или с процессом интеграции административных данных для создания лонгитюдного набора данных. Интеграция микроданных, включающая слияние и отслеживание лиц, является сложной задачей. Связывание индивидуальных записей за период времени, будь то между обновлениями регистра населения или между регистром населения и другими административными данными, часто представляет трудности, даже когда существует персональный идентификационный номер. Эти проблемы подробно описаны в этой главе, но ключевым ограничивающим фактором является то, что эти данные не собирались для статистических целей. Следовательно, теряется контроль над временем, охватом населения и соответствием концепциям и определениям миграции. Но административные источники обычно могут производить большие размеры выборки, чем традиционные лонгитюдные обследования, с более регулярным последующим наблюдением и стремящимся к нулю оттоком из-за отсутствия ответов. В результате они, в целом, хорошо подходят для лонгитюдного изучения малых популяций мигрантов.
99. Эти проблемы и ограничения поднимаются в Главе 3, которая описывает передовую практику о том, как разработать лонгитюдный набор данных для статистики миграции и стремится устранить ограничения различных источников данных.

ГЛАВА 3

КАК РАЗРАБОТАТЬ НАБОР ПРОДОЛЬНЫХ ДАННЫХ ДЛЯ СТАТИСТИКИ МИГРАЦИИ С ИСПОЛЬЗОВАНИЕМ ИНТЕГРИРОВАННЫХ ДАННЫХ

3.1 ВВЕДЕНИЕ

100. В этой главе представлены рекомендации о том, как разработать продольный набор данных для статистики миграции на основе интегрированных данных, с учетом Универсальной модели статистического бизнес-процесса (GSBPM) (UNECE, 2019). В главе основное внимание уделяется элементам, специфичным для интеграции данных, продольных данных и статистики миграции. Руководство ЕЭК ООН по интеграции данных для измерения (UNECE, 2019) следует использовать в качестве дополнения — особенно для ссылок на интеграцию данных.
101. Каждый раздел этой главы будет охватывать этап процесса, предоставляя объяснение того, как предпринять каждый шаг, с использованием конкретных примеров из различных национальных статистических управлений. Будут определены конкретные проблемы и предложены решения.
102. Этапы разработки лонгитюдного набора данных для статистики миграции на основе интегрированных данных включают:
1. Статистический дизайн
 2. Оценка и предварительная обработка исходных файлов
 3. Интеграция данных для продольных данных
 4. Назначение лонгитюдных индивидуальных идентификаторов
 5. Создание окончательной базы данных
 6. Распространение результатов
103. В этой главе представлены рекомендации по каждому из этих этапов. Их необходимо рассматривать как в контексте разработки новой базы данных, так и в контексте обновления существующей (например, для добавления новых результатов через интеграцию данных).
104. Порядок этапов важен. Однако, в зависимости от результатов последующих этапов, может потребоваться итерация между этими этапами. Например, оценка исходных файлов (этап 2) может потребовать пересмотра статистического дизайна (этап 1). Также возможно, что проблемы качества данных, связанные с процессом интеграции, могут быть выявлены только во время окончательной оценки базы данных (этап 5). Это может потребовать нового процесса интеграции данных (этап 3) и повторения последующих этапов.
105. Хотя все этапы должны быть рассмотрены, принятие мер не требуется во всех случаях. Например, обновление существующего проекта может не требовать пересмотра этапа статистического дизайна, а в случае производства продольных данных из системы регистров этап интеграции данных может быть прямолинейным.
106. В этой главе представлены не отдельные тематические исследования, а примеры из разных стран для конкретных элементов. Если пользователям интересно, как были разработаны конкретные источники данных от начала до конца, им рекомендуется ознакомиться с доступной документацией по существующим источникам данных. Например, в Канаде база данных лонгитюдной иммиграции (IMDB) представлена в техническом отчете.

РИСУНОК 4: ЭТАПЫ РАЗРАБОТКИ ЛОНГИТЮДНОГО НАБОРА ДАННЫХ



Источник: Статистическое управление Канады.

3.2 СИСТЕМА ОШИБОК ДЛЯ ПРОДОЛЬНЫХ ДАННЫХ ИЗ ИНТЕГРИРОВАННЫХ ИСТОЧНИКОВ ДАННЫХ

107. Качество данных является повторяющимся предметом рассмотрения на всех этапах. Исходные файлы должны быть выбраны так, чтобы наилучшим образом соответствовать целям статистического дизайна, и качество потенциальных исходных файлов имеет критически важное значение для этого решения. Качество исходных файлов далее оценивается на втором этапе и, где возможно, корректируется с использованием различных методов. Сама интеграция данных требует оценки ошибок интеграции, а назначение лонгитюдных индивидуальных идентификаторов требует некоторого согласования результатов интеграции данных. При создании окончательной базы данных используются методы для гармонизации концепций с течением времени и подготовки базы данных к анализу. Наконец, все соображения по качеству данных должны быть документированы, и соответствующее руководство должно быть предоставлено пользователям перед распространением результатов.

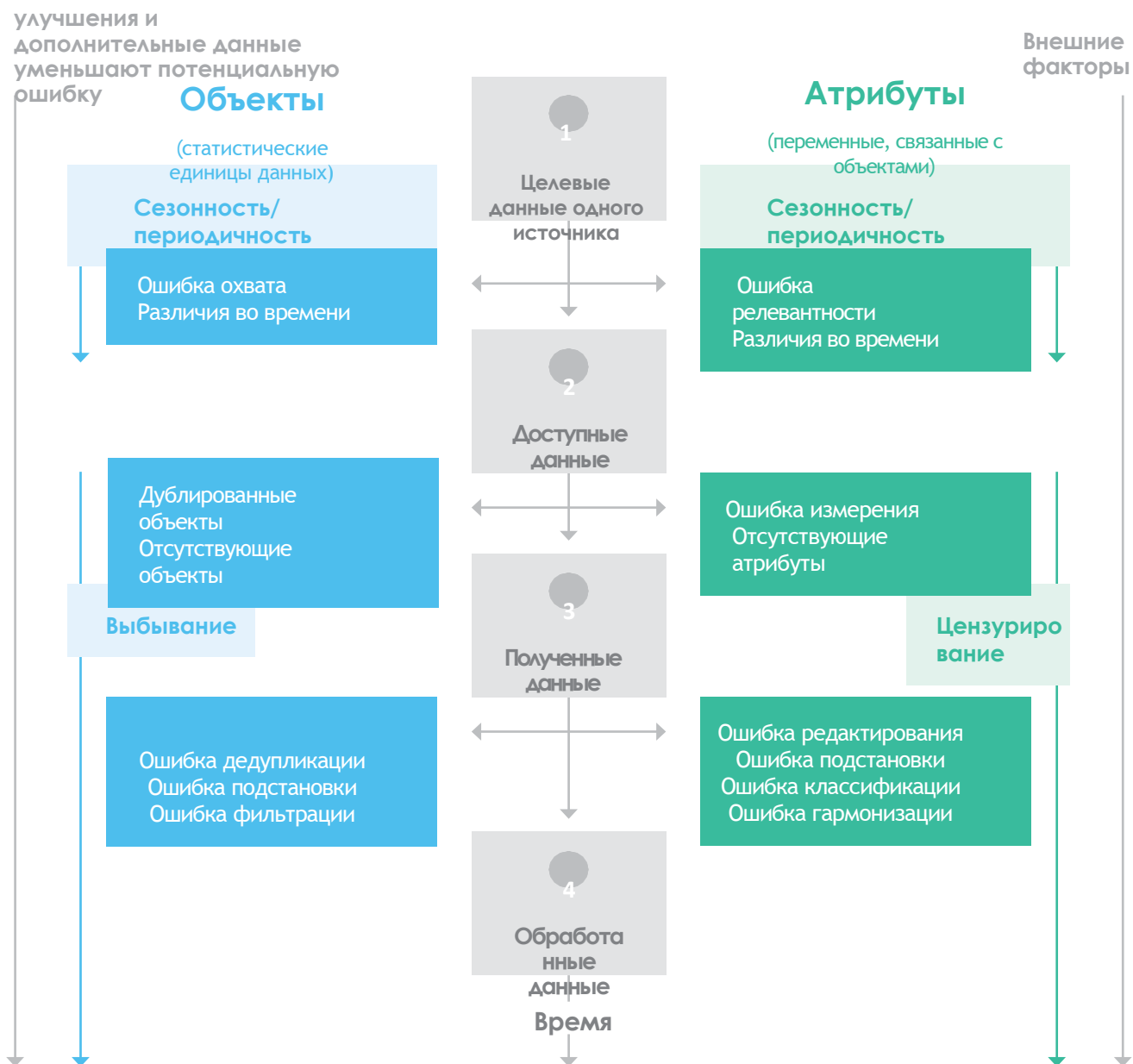
108. Полезно опираться на Рамку погрешности сплошного обследования, которая была разработана для поддержки статистического дизайна (Groves, 2004), (Zhang, 2012). Это позволяет обеспечить полезную таксономию различных потенциальных источников ошибок, относящихся в основном к административным данным. Не все типы ошибок измеримы. Но о них нужно помнить при проектировании лонгитюдно-интегрированного набора данных. Ниже приведено расширенное описание рамки, предложенной Чжаном (2012) для более полного обсуждения источников ошибок, связанных с связыванием административных источников для производства лонгитюдных наборов данных.

109. Рамка, разработанная в Великобритании (Blackwell & Rogers, 2019), предлагает поэтапный подход к пониманию качества данных, при котором сначала рассматриваются отдельные наборы данных, а затем оценивается ошибка при создании наборов данных, созданных путем интеграции нескольких источников. В соответствии со структурой, предложенной Чжаном (Zhang, 2012) и доработанной Статистическим управлением Новой Зеландии (Statistics New Zealand, 2016), рассматриваются ошибки, связанные как с объектами данных (записями), так и с их атрибутами (переменными). Однако в административных данных эти ошибки задаются исходными наборами данных; атрибуты

разрабатываются не так, как в опросе, и то, что уже собрано, используется повторно. В основе структуры лежат наборы данных, которые не были разработаны, они такие, какие они есть.

110. Рисунок 5 и Рисунок 6 описывают рамки погрешности для одиночных и множественных источников. Таблица 3, Таблица 4 и Таблица 5 определяют ошибки на каждом этапе. Рамка для одиночного источника идентифицирует четыре различных этапа пути данных — целевые данные, доступные данные, данные в доступе и обработанные данные. Ошибки представлены как концептуальная разница между данными на каждом из этих этапов. Целевые данные являются концептуальными — идеальные данные для сбора, поэтому ошибки между целевыми данными и доступными данными являются концептуальными.

РИСУНОК 5 РАМКА ПОГРЕШНОСТИ ДЛЯ ОДНОГО ИСТОЧНИКА

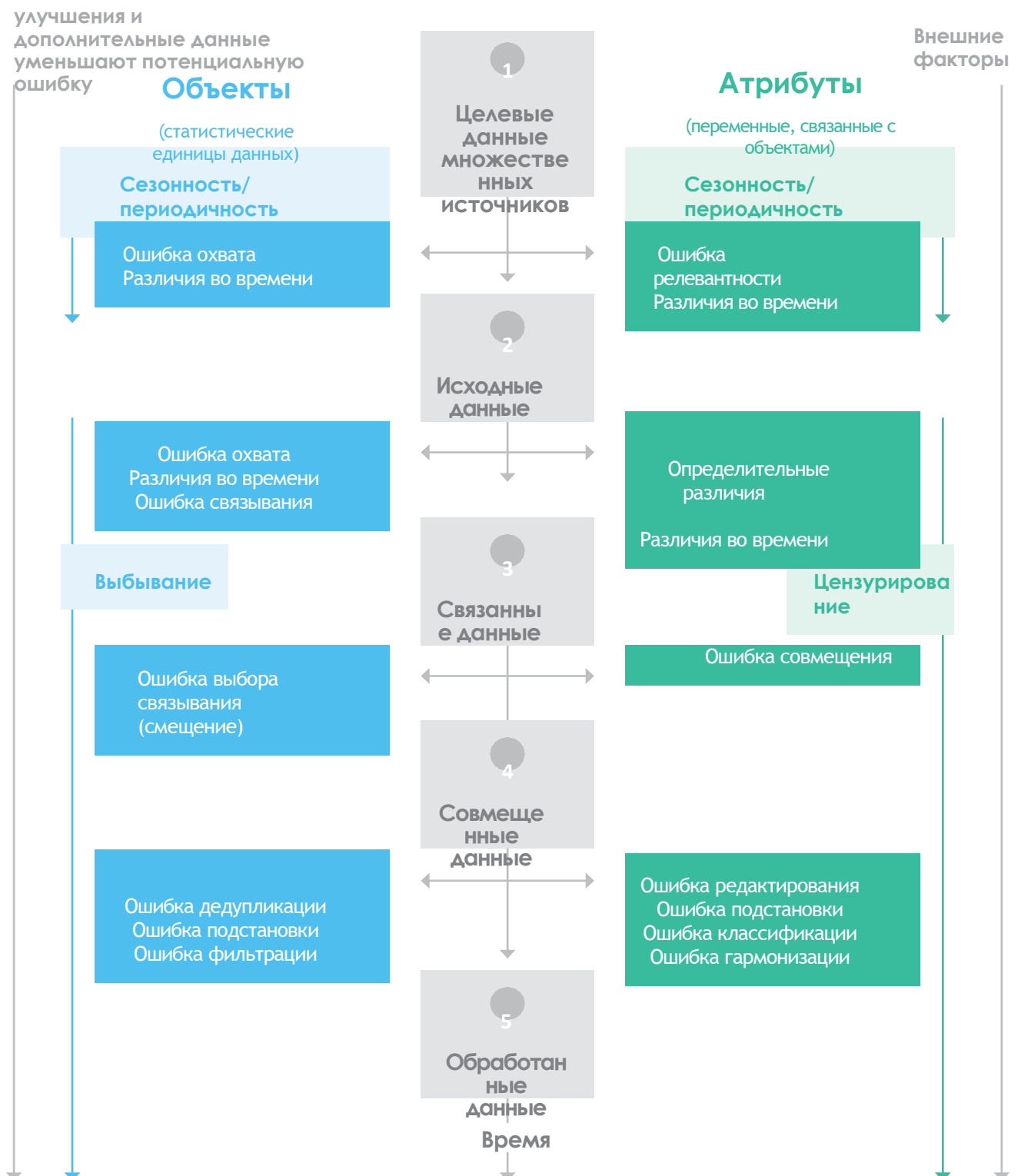


Источник: Управление национальной статистики Великобритании.

Примечания: Время представлено вертикально. Целевые данные представляют собой идеальные данные для сбора. Объекты (статистические единицы данных) относятся к строкам в источнике данных и тому, что эти строки данных представляют. Атрибуты (переменные, связанные с объектами) относятся к столбцам в источнике данных и тому, что эти столбцы представляют.

111. Ошибки разделены между объектами и атрибутами: ошибки, происходящие с объектами, относятся к сущности, для которой собраны данные, будь то люди, события или предприятия. Ошибки, относящиеся к атрибутам, относятся к тем параметрам, которые измеряются для объектов. Ошибки как для объектов, так и для атрибутов могут влиять друг на друга (представлено двойной стрелкой между ними). Не все типы ошибок будут применимы к каждому источнику данных, и могут быть выявлены другие источники ошибок, особенно для обработанных данных. В контексте данной публикации объекты, как правило, относятся к отдельным лицам, определяемым Файлами когорты (или когортами, определяемыми Файлами результатов), в то время как атрибуты могут относиться к результатам, определяемым Файлами результатов.

РИСУНОК 6: РАМКА ПОГРЕШНОСТИ ДЛЯ МНОЖЕСТВЕННЫХ ИСТОЧНИКОВ



Источник: Управление национальной статистики Великобритании

Примечания: Время представлено вертикально. Целевые связанные данные представляют идеальные связанные данные для производства. Объекты (статистические объекты данных) относятся к строкам в источнике данных и тому, что эти строки данных представляют. Атрибуты (переменные, связанные с объектами) относятся к столбцам в источнике данных и тому, что эти столбцы представляют.

112. Ошибка распространяется по всей рамке, хотя и не обязательно накапливается на каждом этапе. Эта рамка использовалась, чтобы помочь исследователям в Управлении национальной статистики Великобритании понять продольно связанные административные данные, включая данные проверок на выходе из Министерства внутренних дел. Цель состоит в том, чтобы оценить их использование для статистики международной миграции. Производство данных проверок на выходе начинается с поставки в Министерство внутренних дел данных о пассажирах, предоставляемых коммерческими перевозчиками в портах въезда/ выезда Великобритании (United Kingdom Home Office, 2019). Министерство внутренних дел обрабатывает эти данные и следит за их качеством, связывая их с другими операционными данными для визовых иностранцев.
113. Это приложение показало, что в связанных административных данных ошибка не обязательно является кумулятивной. Это связано с тем, что эксперты в Министерстве внутренних дел стремятся устранить ошибку в данных по мере их прохождения через процессы. Хотя ошибка может накапливаться с течением времени, также существует возможность того, что компенсирующие ошибки могут осложнить положение вещей. Компенсирующие ошибки (например, подстановка данных для пропущенных значений) могут быть не видны в моментных распределениях данных, но могут усугубляться лонгитудно. Это объясняется в контексте рамки множественных источников (Рисунок 6).
114. Рамка также включает лонгитудную ошибку, создаваемую сбором данных с течением времени — сезонность/периодичность, выбывание и цензурирование (Таблица 4).

ТАБЛИЦА 3: ОШИБКИ ЕДИНОГО ИСТОЧНИКА

Объекты ¹	Атрибуты ²
<p>Ошибка рамки: <u>Ошибка охвата</u> Оценка объектов, которых нет в целевых данных, или невозможность доступа к объектам, которые есть в целевых данных. <u>Временные различия</u> Объекты в целевых данных, которые недоступны из-за несоответствия во временном окне для получения наблюдений.</p>	<p>Ошибка релевантности: <u>Ошибка достоверности</u> Разница между идеальным измерением искомым атрибутов объекта и оперативным показателем, используемым для его сбора. <u>Временные различия</u> Концептуальное несоответствие во времени измерения атрибутов между целевыми данными и доступными данными.</p>
<p>Ошибка выбора: <u>Дублирующиеся объекты</u> Объекты, которые представлены в доступных данных более одного раза. <u>Отсутствующие объекты</u> Объекты, которые теоретически доступны, но отсутствуют в доступных данных.</p>	<p>Ошибка измерения: <u>Ошибка измерения</u> Ошибки, возникающие из-за неточно записанных атрибутов. <u>Отсутствующие атрибуты</u> Атрибуты, отсутствующие в доступных данных (могут отсутствовать для определенных объектов или для всех объектов).</p>
<p>Ошибка обработки: <u>Ошибка дедупликации</u> Ошибки, возникающие при дедупликации объектов в доступном наборе данных. Сюда может относиться как дедупликация объектов, которые на самом деле отличаются (ложноположительная ошибка), так и невозможность дедупликации объектов, которые</p>	<p>Ошибка обработки: <u>Ошибка редактирования</u> Ошибки, возникающие при редактировании значения атрибута. Сюда может относиться редактирование в результате проверки достоверности или обеспечения качества. <u>Ошибка подстановки</u> Ошибки, возникающие при вычислении</p>

являются одинаковыми (ложноотрицательная ошибка). <u>Ошибка подстановки</u> Ошибки, возникающие при вычислении отсутствующих объектов. <u>Ошибка фильтрации</u> Ошибки, возникающие при выборе или отмене выбора доступных объектов для создания целевого набора.	отсутствующих значений атрибута. <u>Ошибка классификации</u> Ошибки, возникающие при классификации значений по группам или при выводе новых атрибутов. <u>Ошибка согласования</u> Ошибки, возникающие при согласовании значений атрибутов с идеальной или целевой концепцией
---	--

Источник: Управление национальной статистики Великобритании.

1 Это относится к единицам данных, и может быть событиями, транзакциями, лицами, домохозяйствами, фирмами или другими записями в административном наборе данных.

2 Это относится к мерам или переменным, которые были собраны и которые относятся к объектам/единицам данных.

ТАБЛИЦА 4: ЛОНГИТЮДНАЯ ПОГРЕШНОСТЬ – ПРИМЕНЯЕТСЯ КАК К СОБЫТИЯМ ОДИНОЧНОГО ИСТОЧНИКА, ТАК И К МНОГОИСТОЧНИКОВЫМ ЛОНГИТЮДНЫМ НАБОРАМ ДАННЫХ

Объекты¹	Атрибуты²
Выбытие Потеря объектов исследования или единиц измерения с течением времени. Происходит естественным путем, в результате смерти (или ненаблюдаемой миграции). Также это происходит из-за непринятия последующих мер, отказа от участия, в случае получения данных обследования, или из-за отсутствия информации или сбоя связи в административных источниках.	Цензурирование Когда значение измерения или наблюдения известно лишь частично. Правая цензура – это когда объект исследования выпадает из данных до окончания периода наблюдения или не испытывает интересующего события в течение периода наблюдения. Левая цензура применяется, когда интересующее событие уже произошло до того, как началось окно наблюдения.
Ошибка периодичности/ сезонности Объекты не наблюдаются, поскольку сбор данных осуществляется недостаточно часто (периодичность) и недостаточно адекватно для учета сезонности в данных (сезонность).	Ошибка периодичности/ сезонности Измерения атрибутов с течением времени происходят недостаточно часто (периодичность) и недостаточно адекватно для отражения сезонности в данных (сезонность).

Источник: Управление национальной статистики Великобритании.

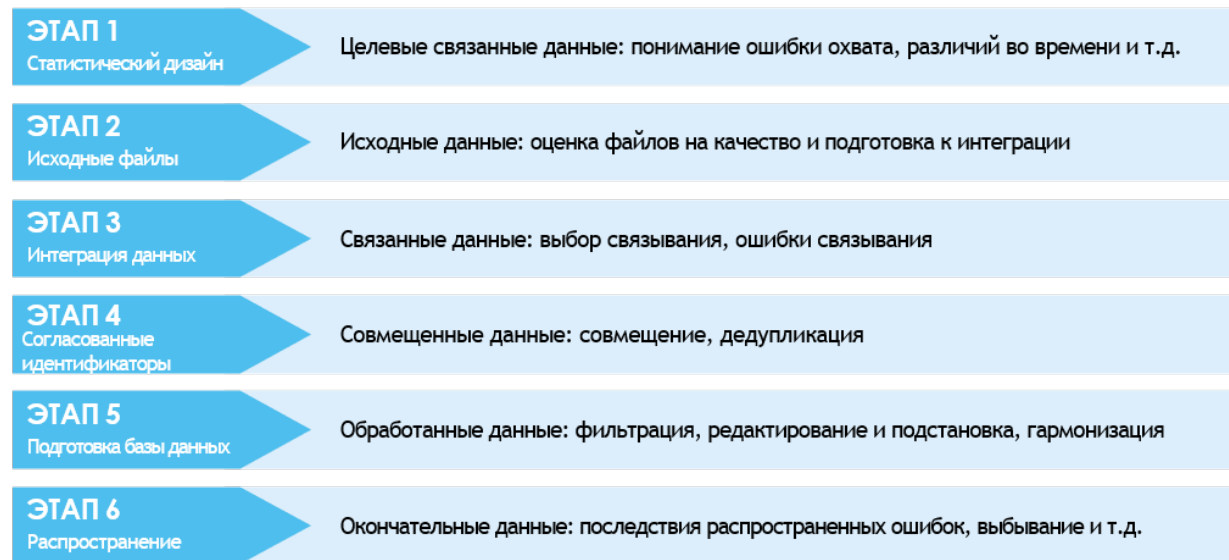
1 Это относится к единицам данных, и может быть событиями, транзакциями, лицами, домохозяйствами, фирмами или другими записями в административном наборе данных.

2 Это относится к мерам или переменным, которые были собраны и которые относятся к объектам/единицам данных.

115. Рамка множественных источников предназначена для интеграции нескольких источников данных. Она состоит из пяти этапов — целевые связанные данные, исходные данные, связанные данные, совмещенные данные и обработанные данные. В контексте данной публикации эти этапы соответствуют фазам, рассматриваемым в этой главе, как показано на Рисунке 7.

116. Ошибки представлены как концептуальная разница между данными на каждом из этих этапов. Многие ошибки в рамке множественных источников концептуально схожи с рамкой единого источника. Основное различие заключается в том, что ошибки измеряются между исходными наборами данных и целевыми связанными данными (а не целевыми данными), а также ошибки между исходными наборами данных, которые необходимо связать. Эти концептуально схожие ошибки имеют одинаковые названия для рамок единого источника и множественных источников. Например, различия во времени, ошибка охвата, ошибка релевантности, ошибка подстановки, ошибка отбора и ошибка обработки.

РИСУНОК 7: ЭТАПЫ ИНТЕГРАЦИИ МНОЖЕСТВЕННЫХ ИСТОЧНИКОВ ДАННЫХ



Источник: Статистическое управление Канады.

117. Целевые связанные данные отличаются от целевых данных для каждого отдельного источника. Целевые связанные данные, вероятно, будут специфичны для группы объектов, которые должны быть измерены посредством связывания нескольких источников. Для заключительного этапа (обработанные данные) обработка могла уже произойти в рамках единого источника или в рамках множественных источников. Например, подстановка данных уже могла произойти на уровне единого источника, или она может быть сделана после связывания.

ТАБЛИЦА 5: ОШИБКИ МНОЖЕСТВЕННЫХ ИСТОЧНИКОВ

Объекты ¹	Атрибуты ²
<p>Ошибка рамки: <u>Ошибка охвата</u> Наблюдение за объектами, которых нет в целевых связанных данных, или невозможность доступа к объектам, которые есть в целевых связанных данных.</p> <p><u>Временные различия</u> Объекты не наблюдаются из-за концептуальных расхождений во времени захвата между целевыми связанными данными и исходными данными.</p>	<p>Ошибка релевантности: <u>Ошибка достоверности</u> Различия между идеальным измерением искомых атрибутов объекта и оперативными показателями, используемыми для их сбора в каждом исходном наборе данных.</p> <p><u>Временные различия</u> Концептуальное несоответствие во времени измерения атрибутов между целевыми данными и доступными данными.</p>
<p>Ошибка охвата: <u>Ошибка охвата</u> Объекты не связаны из-за различий в охвате объектов между источниками данных.</p> <p><u>Разница во времени</u> Разница между наблюдаемыми объектами в исходных наборах данных из-за того, что данные были получены в разное время.</p> <p><u>Ошибка привязки</u> Ошибки, возникающие при неправильном соединении объектов (ложноположительная ошибка) и невозможности связать объекты, которые должны были быть связаны (ложноотрицательная ошибка).</p>	<p>Ошибка отображения: <u>Различия в определениях</u> Различия в способах оперативного измерения атрибутов в каждом из исходных наборов данных.</p> <p><u>Временные различия</u> Различия между значениями атрибутов для связанного объекта в разных исходных наборах данных, вызванные тем, что данные были получены в разное время.</p>
<p>Ошибка идентификации:</p>	<p>Ошибка сопоставимости:</p>

<p><u>Ошибка выбора связи (смещение)</u> Ошибки, возникающие при выборе связанных объектов (или отмене выбора несвязанных объектов) из-за смещений в связи или из-за ошибки в разрешении конфликтующих связей.</p>	<p><u>Ошибка выравнивания</u> Ошибки, возникающие из-за выравнивания конфликтующих значений атрибутов в разных источниках.</p>
<p>Ошибка обработки: <u>Ошибка подстановки</u> Ошибки, возникающие при вычислении отсутствующих объектов. <u>Ошибка фильтрации</u> Ошибки, возникающие при выборе или отмене выбора доступных объектов для создания целевого набора.</p>	<p>Ошибка обработки: <u>Ошибка редактирования</u> Ошибки, возникающие при редактировании значения атрибута. Сюда может относиться редактирование в результате проверки достоверности или обеспечения качества. <u>Ошибка подстановки</u> Ошибки, возникающие при вычислении отсутствующих значений атрибута. <u>Ошибка классификации</u> Ошибки, возникающие при классификации значений по группам или при выводе новых атрибутов. <u>Ошибка согласования</u> Ошибки, возникающие при согласовании значений атрибутов с идеальной или целевой концепцией</p>
<p>Объекты¹</p> <p>Выбытие Потеря объектов исследования или единиц измерения с течением времени. Происходит естественным путем, в результате смерти (или ненаблюдаемой миграции). Также это происходит из-за непринятия последующих мер, отказа от участия, в случае получения данных обследования, или из-за отсутствия информации или сбоя связи в административных источниках.</p>	<p>Атрибуты²</p> <p>Цензурирование Когда значение измерения или наблюдения известно лишь частично. Правая цензура – это когда объект исследования выпадает из данных до окончания периода наблюдения или не испытывает интересующего события в течение периода наблюдения. Левая цензура применяется, когда интересующее событие уже произошло до того, как началось окно наблюдения.</p>
<p>Ошибка периодичности/ сезонности Объекты не наблюдаются, поскольку сбор данных осуществляется недостаточно часто (периодичность) и недостаточно адекватно для учета сезонности в данных (сезонность).</p>	<p>Ошибка периодичности/ сезонности Измерения атрибутов с течением времени происходят недостаточно часто (периодичность) и недостаточно адекватно для отражения сезонности в данных (сезонность).</p>

Источник: Управление национальной статистики Великобритании.

1 Это относится к единицам данных, и может быть событиями, транзакциями, лицами, домохозяйствами, фирмами или другими записями в административном наборе данных.

2 Это относится к мерам или переменным, которые были собраны и которые относятся к объектам/единицам данных.

118. Цель применения системы учета погрешности административных данных состоит в том, чтобы обеспечить оптимизацию качества данных на всех этапах статистического проекта. На начальном этапе необходимо выявить и изучить источники ошибок для принятия статистических проектных решений в процессе последующего связывания данных. Однако также важно заново оценивать различные источники ошибок на всех последующих этапах. Перечни в Таблицах 4 и 5 предлагают ряд показателей качества. Национальному статистическому управлению (НСУ) может быть нецелесообразно измерять каждый показатель, но необходимо определить ключевые показатели для включения в отчет о качестве (см. раздел 3.4.3).

119. Существует взаимосвязь между различными источниками ошибок. Например, может возникать компромисс между ошибками связывания, охвата и подстановки данных. Записи с данными низкого качества, возможно, из-за ошибок измерения, также может быть сложнее связать. Это может быть обусловлено качеством идентификаторов, используемых для связывания. Один из вариантов — разработать усовершенствованные методы ведения записей, чтобы минимизировать ложные

отрицательные сопоставления (тем самым допуская больше ложных положительных) и максимизировать охват связанного набора данных. Но существует вероятность, что атрибуты, относящиеся к этим объектам, также низкого качества и приведут либо к пропускам в полях атрибутов, либо к необходимости подстановки данных. Подстановка часто нежелательна в продольных данных, поскольку может приводить к возникновению ложных результатов. Предотвращение потери данных и подстановки может быть первостепенной задачей.

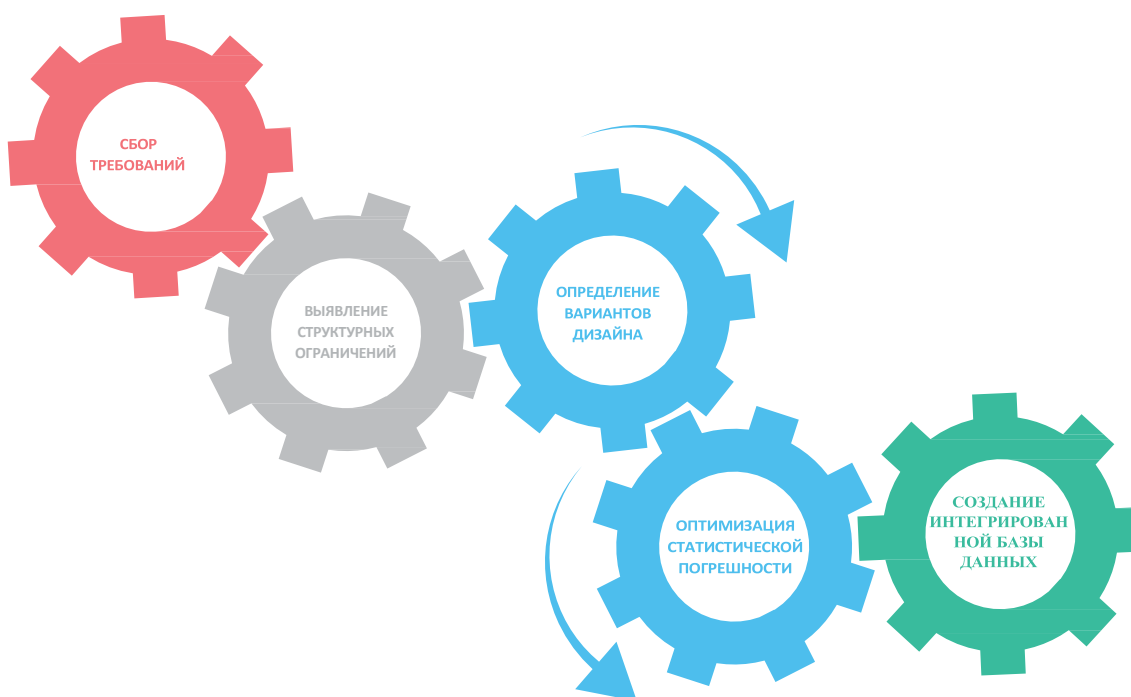
120. Также существует взаимодействие между ошибками в объектах (или когортах) и в атрибутах (или исходах), а также между наборами данных из одного и нескольких источников. Отсутствующие или ошибочно записанные данные атрибутов могут влиять на возможность дедублирования или связывания записей на этапе работы с одним источником. Это, в свою очередь, влияет на возможность связывания объектов и, следовательно, создает ошибки в наборе данных из нескольких источников.

3.3 ЭТАП 1: СТАТИСТИЧЕСКИЙ ДИЗАЙН⁴

121. Это первый шаг в процессе разработки любого источника данных. Ключевыми результатами данного этапа являются определение идеальных целей и задач, которые должны быть решены с помощью лонгитюдного набора данных, включая необходимые концепции и переменные, и, в конечном счете, определение, какие исходные файлы будут интегрированы для наилучшего достижения поставленных целей. Как описано в разделах 3.1 и 3.2, важно учитывать различные ошибки, которые могут оказывать влияние на протяжении всего проекта.

⁴ Этот раздел основан на ЕЭК ООН (2019), который является рекомендуемым источником для использования в дополнение к настоящему Руководству.

РИСУНОК 8: ОБЗОР ПРОЦЕССА СТАТИСТИЧЕСКОГО ДИЗАЙНА



Источник: Управление национальной статистики Великобритании.

122. Этап проектирования включает:

- *Определение целей статистического проекта*

- *Выявление любых структурных ограничений, которые могут повлиять на статистический дизайн*
 - *Определение вариантов дизайна, включая выбор исходных файлов, с учетом статистических ошибок*
123. Цель заключается в выборе проекта, включая исходные файлы, который обеспечивает оптимальный баланс между различными источниками ошибок, присущими созданию интегрированных источников для достижения целей проекта.
124. При определении того, какие исходные файлы необходимы для достижения поставленных целей, можно провести различие между файлами Когорт и файлами Результаты — особенно в контексте интеграции данных для лонгитюдных данных.
125. Файлы Когорт — это исходные файлы, используемые для идентификации интересующей(их) популяции(й) и ее характеристик, в то время как Файлы Результаты — это лонгитюдные файлы, предоставляющие данные о результатах с течением времени. В некоторых случаях файлы Результаты также могут служить файлами Когорт.
126. Например, Лонгитюдная база данных по иммигрантам (Канада) связывает административные файлы по иммиграции (файлы Когорт) с лонгитюдными налоговыми файлами (файлы Результаты). Налоговые файлы также могут быть использованы для идентификации когорты, либо самостоятельно, либо в сочетании с иммиграционными файлами. Например, когорта может состоять из иммигрантов, принятых в определенный год (когорта, определенная только файлами Когорт), лиц, заявивших о расходах на обучение в определенный год (когорта, определенная только файлами Результаты), или иммигрантов, имевших доход от занятости в первый год пребывания в Канаде (когорта, определенная файлами Когорт и Результаты).
127. В конечном счете, результатом этого этапа является отбор исходных файлов, которые наилучшим образом соответствуют заявленным целям проекта, исходя из охватываемых ими популяций, их временного охвата и включенных показателей. Выборочный метод и любые другие потенциально дисквалифицирующие проблемы качества данных (например, файлы Результаты основаны на выборках, слишком малых для анализа интересующих популяций) также следует учитывать, но более глубокий обзор качества данных будет проведен на этапе 2. Любые недостатки, связанные с выбранными исходными файлами, должны быть хорошо документированы.

3.3.1 Определение статистического проекта

3.3.1.1 Постановка целей

128. В первую очередь крайне важно установить цели статистического проекта. Это должно делаться в тесном сотрудничестве с соответствующими пользователями и заинтересованными сторонами. Цели должны быть четко определены и хорошо документированы, поскольку они будут направлять все остальные этапы проекта.
129. Для этого необходимо ответить на несколько вопросов о задачах проекта, а также о концепциях и переменных, представляющих интерес:
- Какие вопросы должны быть решены с помощью данных?
 - Почему для решения этих вопросов необходим лонгитюдный подход?
 - Является ли целью понимание процессов миграции и опыта мигрантов? Или мы стремимся производить надежные показатели?
 - Какова интересующая популяция? Можем ли мы учесть эту популяцию в полном объеме? Нужна ли нам контрольная популяция для сравнения?
 - Каковы переменные или концепции, представляющие интерес?
 - Как определяется «мигрант»? Можно ли провести идентификацию в исходных наборах данных?
 - Состоит ли наша основная задача в изучении миграции как компонента изменения численности населения? Если да, то в равной ли нас интересуют иммиграция и эмиграция?
 - Представляют ли интерес сезонные модели?

- Как насчет внутренней миграции? Интересна ли внутренняя миграция международных мигрантов? Стремимся ли мы узнать, как это отличается от модели внутренней миграции немигрантов?
- Все ли международные мигранты представляют одинаковый интерес? Или первостепенное значение имеют намерения мигрантов? Например, если мигранты приезжают или уезжают для учебы, работы, по семейным обстоятельствам, для обеспечения безопасности или по другим причинам?
- Применяются ли различные миграционные политики к различным типам мигрантов?
- Нужно ли нам определять, имеют ли мигранты временные, краткосрочные или постоянные намерения в принимающей стране?
- Как далеко за пределы события миграции мы хотим заглянуть? Как долго нам нужно наблюдать постмиграционные результаты? Это может включать социальные и экономические результаты, включая показатели социальной интеграции.
- Представляют ли интерес траектории жизненного пути? Модели создания партнерских отношений, формирования семьи, смерти и причины смерти.
- Интересны ли межпоколенческие воздействия миграции? Стремимся ли мы отслеживать детей, рожденных до и/или после миграции?
- Насколько детализированные данные требуются, например, по характеристикам и географии? Является ли целью получение национальных, региональных или местных результатов?

130. Эти вопросы имеют последствия для необходимых показателей, охватываемых популяций, а также частоты и интервалов наблюдаемых исходов.

131. В идеале, окно наблюдения за самой миграцией должно предшествовать событию миграции, но это был бы амбициозный проект, требующий международного обмена данными. Если наблюдение начинается в точке иммиграции, период последующего наблюдения должен быть достаточно продолжительным, чтобы охватить интересующие результаты.

132. Переменные, которые могут представлять интерес, могут быть обширными. Они могут включать национальность, гражданство, страну рождения, возраст, пол, этническую принадлежность, семейное положение, год прибытия или место жительства.

133. Один критически важный момент, который следует учитывать, — будет ли лонгитюдный источник данных воспроизводить те же оценки, что и источники, используемые в интеграции данных. Поскольку интегрированные лонгитюдные данные требуют соединений между источниками, дубликаты (т.е. несколько идентификаторов для одной и той же единицы) или разделения (т.е. несколько единиц для одного и того же идентификатора) могут быть обнаружены в процессе интеграции, что в противном случае повлияло бы на сами оценки исходных файлов. Рекомендуется рассмотреть возможность отказа от этого требования из-за его сложности. Вместо этого следует рассматривать лонгитюдный источник данных в свете общих статистических целей и использовать оценки исходных файлов для оценки качества данных.

3.3.1.2 Понимание требований к статистическому качеству

134. Крайне важно понимать требования к статистическому качеству. Является ли целью создание доказательной базы для политики? Для планирования? Для национальной статистики? Будет ли это основой для распределения ресурсов? Если да, то на каком географическом уровне? Имеют ли заинтересованные стороны национальные или местные интересы? Есть ли международная заинтересованная сторона?

135. Требование к точности имеет последствия для управления ошибками. Например, проектирование связи записей включает компромисс между ошибками ложных отрицательных и ложных положительных результатов. Существует ли критический порог точности, измеряемый в терминах статистической достоверности, который должны поддерживать данные? Требуется ли проект мер статистической неопределенности? Различаются ли требования к точности в наборе данных? Например, пользователи могут допускать определенную степень пропусков для одних переменных, но не для других.

3.3.1.3 Согласованность и связность

136. Поддержание согласованных показателей с течением времени может иметь критическое значение для лонгитюдного исследования. Далее, необходимо определить, необходимо ли иметь результаты, сопоставимые с другими источниками данных. Например, абсолютно ли необходимо, чтобы лонгитюдный источник данных давал перекрестные оценки, согласующиеся с другими источниками данных? Какие переменные должны быть согласованы во времени? Потребуется ли корректировка, например, для учета потерь при последующем наблюдении или отсева из данных? Будет ли база данных использовать стандартные концепции и классификации?

3.3.1.4 Своевременность

137. Еще одним соображением является требование о своевременном представлении доказательств. Может ли проект выдержать задержки из-за необходимых последующих действий, например, для выполнения требований к определению? «Долгосрочный» мигрант может быть определен как прибывающий и остающийся мигрант или выезжающий и отсутствующий в течение как минимум 12 месяцев. Это подразумевает задержку в выявлении мигрантов на 1 год. Если это неприемлемо, необходимо рассмотреть возможность оценки или условного обозначения этого показателя.

3.3.1.5 Доступность

138. На структуру статистики могут влиять намерения в отношении доступа к данным. Если цель состоит в том, чтобы предоставить исследователям весь набор данных, требования к подготовке и обработке данных могут быть более обременительными, чем если бы цель состояла в том, чтобы хранить микроданные внутри компании и предоставлять результаты внешним заинтересованным сторонам. Степень ясности и прозрачности, необходимая третьей стороне для использования и понимания данных, может потребовать обширных метаданных, подтверждающей документации, руководства или даже службы поддержки пользователей.

3.3.1.6 Релевантность

139. Административные данные имеют свой собственный функциональный план, основанный на конкретных целях. В интегрированных наборах данных они перепрофилируются для удовлетворения другого набора требований. Это неизбежно требует компромисса. В качестве постоянного процесса могут потребоваться консультации с поставщиками административных данных и заинтересованными сторонами в статистических проектах, чтобы гарантировать, что решения по разработке данных основаны на статистических целях, а решения по разработке статистических данных - на изменениях в сборе административных данных. При этом следует также учитывать, что у разных пользователей могут быть разные потребности.

140. Если переменные, имеющие критическое значение для пользователей, недоступны ни в одном из исходных наборов данных, являются ли прокси достаточно хорошими? Насколько хорошо они должны соответствовать желаемой информации?

3.3.1.7 Интерпретируемость

141. Наконец, важно учитывать, будут ли результаты хорошо поняты пользователями. Необходимо установить надлежащую документацию и метаданные, чтобы объяснить, как была разработана интегрированная база данных, а также как использовать и понимать результаты.

3.3.2 Структурные ограничения, которые могут повлиять на статистический дизайн

142. Обзор соответствующих источников позволит выявить наборы данных, доступные для интеграции. Потенциальные файлы Когорт и Исходов следует оценивать с точки зрения того, насколько хорошо они соответствуют целям, определенным в разделе 3.3.1.

143. Прежде чем рассматривать содержание, охват, своевременность, качество и другие статистические аспекты, могут существовать более структурные ограничения, которые могут повлиять на выбор файлов или возможность осуществления интеграции данных. Вопросы, которые следует рассмотреть при оценке потенциальных исходных файлов, включают:

- Как будет защищена конфиденциальность данных — на этапах разработки данных и для распространения?
- Каковы затраты, связанные с доступом и использованием данных?
- Существуют ли какие-либо правовые барьеры или барьеры доступа, такие как необходимость информированного согласия респондентов? Существует ли правовая или институциональная рамка, регулирующая доступ или использование этих данных? Существуют ли ограничения на интеграцию данных?
- Существуют ли какие-либо этические проблемы? Требуется ли вам этическое одобрение для использования данных?
- Возможно ли получение или обмен записями с другими национальными статистическими управлениями или другими иностранными организациями?
- Требуются ли особые навыки работы с программным обеспечением для манипулирования данными или управления ими?
- Какая инфраструктура требуется для хранения и обработки этих данных или структуры данных? Безопасна ли инфраструктура? Достаточно ли места? Как долго может поддерживаться хранение?
- Следует ли интегрировать каждый источник для создания связанного набора данных, или было бы полезно сохранить некоторые данные в качестве средства сравнения, возможно, для калибровки или корректировки?
- Является ли интеграция или связывание этих данных устойчивой с течением времени?
- Какие технические ресурсы требуются для поддержки роста масштаба и сложности данных с течением времени?
- Существуют ли какие-либо известные риски для непрерывности источников? Какой контроль существует над будущим производством? Каковы риски того, что критические элементы данных будут изменены или прекращены? Каков план экстренных мер, если это произойдет?

144. Разрывы в определениях и классификациях между источниками и с течением времени обсуждаются ниже в контексте источников статистической ошибки.

3.3.3 Определение вариантов дизайна и оптимизация статистической погрешности

145. Этап проектирования требует серии решений, на которые влияют требования проекта, доступные данные и любые структурные ограничения. Они также требуют оптимального баланса различных источников потенциальной ошибки внутри интегрированного набора данных, принимая во внимание цели проекта. Рекомендуется обратиться к рамке во введении этой главы для получения дальнейших указаний.
146. Более тщательная оценка качества данных выбранных исходных файлов будет проведена на этапе 2. Однако на этом этапе важно определить оптимальные потенциальные файлы Когорт и Результатов с точки зрения статистической погрешности для достижения заявленных целей проекта при любых структурных ограничениях. Этот раздел будет включать оценку охвата популяции (и выборочного метода), измерения, временного охвата и частоты сбора (например, в реальном времени или периодически).
147. Потенциальные исходные файлы могут быть исключены из рассмотрения, если они не позволяют адекватно измерить либо интересующие популяции, либо показатели. Это может быть связано с малым размером выборки (после интеграции), отсутствующими или некачественными показателями или неадекватными прокси. С лонгитюдной точки зрения, временной элемент также имеет критическое значение — файлы могут не адекватно покрывать интересующий период или иметь недостаточно частые наблюдения.

Например, в Венгрии в 2018 году был запущен статистический проект по оценке доли маятниковых мигрантов в общей численности постоянного населения. В качестве первого шага были разработаны определения маятниковой миграции и список переменных, минимально необходимых для проведения исследования. В ходе проекта в 3 разных источниках данных была выявлена группа из 8 различных подгрупп, вовлеченных в процессы международной миграции. Однако после оценки исходных данных только 2 подгруппы были признаны подходящими для проведения исследования. Остальные подгруппы были исключены из-за структурных ограничений (например, некоторые данные отсутствовали в лонгитюдной структуре, в то время как в других случаях отсутствовали персональные идентификаторы или другие существенные переменные). В конце этапа разработки проекта цели были пересмотрены в соответствии с результатами процесса оценки.

148. Необходимо рассмотреть и взвесить следующие вопросы по отношению к заявленным целям, такие как:

- Какой временной период охвачен?
- Какие популяции идентифицированы и охвачены?
- Будет ли выборка достаточно большой после интеграции для удовлетворения требований проекта?
- Как часто измеряются исходы?
- Какие показатели должны быть включены?

3.3.3.1 Различия в охвате

149. При обсуждении охвата удобно различать два разных типа: временной охват и охват популяции.

150. Временной охват представляет собой период, охватываемый различными источниками данных. Этот референтный период не всегда может быть идеальным для заявленных целей проекта. Например, если исходы или когорты доступны только за периоды в далеком прошлом, они могут не быть релевантными для текущей ситуации. Также возможно, что качество сбора данных за определенные периоды может различаться. Это также следует принимать во внимание.

151. Важно учитывать, наилучшим ли образом возможные файлы Результатов и Когорт решают поставленные задачи с этой точки зрения. Также следует отметить, что возможные исходные файлы могут не идеально соответствовать целям, но все же могут давать полезные результаты. Например, файлы Когорт могут включать только недавних мигрантов — это может не дать результатов для всех мигрантов, но они все же могут быть полезны, если проект дает приоритетный анализ недавних мигрантов.

152. В межстрановом разрезе существует широкий разброс во временном охвате среди различных источников данных. Часто это связано с тем, когда собирались административные данные или данные обследований. Временной охват этих источников данных в конечном счете влияет на то, как они могут быть использованы для решения различных статистических проектов.

В Бельгии централизованное управление Центральным регистром населения (ЦРН) ведется с 1985 года. Регистр содержит информацию за записи до 1985 года. Однако надежность этой информации, которая хранилась в муниципальных регистрах, ниже.

Обследование рабочей силы (ОРС) в Бельгии существует с 1983 года, но может быть связано с ЦРН с 2003 года. Обследование по доходам и условиям жизни (ОДУЖ) существует с 2004 года и может быть связано с 2005 года. Кроме того, переписи 2001 и 2011 годов также могут быть связаны с ЦРН.

Административные записи Канады по иммиграции, беженцам и гражданству (IRCC) охватывают разные периоды в зависимости от того, как собирались административные данные.

- Иммиграционные записи с 1952-1979 гг. (с минимальным содержанием)
- Иммиграционные записи с 1980 г. по настоящее время (с детальным содержанием)
- Записи о временных резидентах (включая заявителей на убежище) с 1980 г. по настоящее время
- Записи о гражданстве с 2005 г. по настоящее время
- Записи об услугах по обустройству с 2013 г. по настоящее время

Эти файлы используются как файлы Когорт для Лонгитюдной базы данных по иммигрантам (IMDB), которая объединяет связанные административные иммиграционные и налоговые файлы. База данных включает все доступные иммиграционные записи и записи о временных резидентах и объединяет их с налоговыми файлами Результатов с 1982 года по настоящее время.

Лонгитюдное обследование иммигрантов в Канаду (LSIC) охватывает иммигрантов, прибывших в 2000-01 гг., с тремя волнами сбора данных через шесть месяцев (2001-02), два года (2002-03) и четыре года (2004-05) после прибытия в Канаду.

В Германии период охвата Центрального реестра иностранцев начинается с 2007 года. Центральный реестр иностранцев содержит информацию обо всех иностранцах (включая заявителей на убежище или беженцев), в настоящее время зарегистрированных в Германии (активные) или ранее зарегистрированных в Германии (неактивные).

Ситуация в Италии более сложная. Данные доступны с 1955 года. Однако использование электронной базы данных было внедрено в 1970 году, а первая форма сбора данных, которая кодирует все иностранные государства, была введена в 1995 году. Даже если информация о видах на жительство существует с 1995 года (и стала доступна итальянскому НСУ в 2010 году), их можно использовать для лонгитюдного

анализа. Наконец, первая версия обследования интеграции второго поколения была проведена в 2015 году и будет проведена повторно в 2021 году.

В Испании Регистр населения (Padrón) начинается в мае 1996 года и управляется непрерывно с тех пор. Padrón также можно рассматривать как лонгитюдную базу данных, с 265 млн зарегистрированных перемещений с мая 1996 года (хотя текущее население составляет 46,7 млн человек). Это позволяет анализировать всю историю каждого жителя в Испании за последние 22 года. Особенностью системы является то, что, если человек с определенным идентификатором покидает страну и позже возвращается, присвоенный ему идентификатор будет другим.

В Швейцарии Швейцарская лонгитюдная демографическая база данных (SLDD) содержит информацию, относящуюся ко всем швейцарским гражданам, проживающим в Швейцарии с 2010 по 2017 год, а также иностранцам, проживающим в Швейцарии с 1998 по 2017 год с регулярным видом на жительство.

153. Охват популяции (и выборочный метод) соответствует популяции, охватываемой этими источниками. Иногда это означает, что включены только определенные субпопуляции или что проводилась выборка. Очень важно знать степень охвата каждого перспективного источника данных. Например: включает ли он всех мигрантов или только тех, кто мигрировал законными способами? Охват популяции каждым возможным исходным файлом следует проверять на соответствие заявленным целям проекта. Если определенные субпопуляции исключены, их нельзя изучить с помощью этой базы данных. Если размер выборки слишком мал, это может препятствовать проведению определенных типов анализа.
154. В контексте интеграции данных выборочный метод еще более критичен для рассмотрения. Интеграция микроданных двух источников данных даст результаты только по перекрывающимся записям. Когда две переписи интегрируются с полным охватом одной и той же популяции, в лучшем случае можно ожидать получения интегрированных результатов для всей охваченной популяции. Это может быть «в лучшем случае», потому что могут возникать ошибки интеграции, и, как правило, увеличение ложных отрицательных результатов предпочтительнее увеличения ложных положительных (это потому, что ложные положительные могут генерировать ложные или бессмысленные исходы). Однако, когда один из этих источников становится выборкой, в лучшем случае можно ожидать получения интегрированных результатов только для источника выборки.
155. Если объединить две выборки, то результаты могут быть получены только в той выборке, которая перекрывается между этими двумя выборками. В случае простых случайных выборок наилучшим сценарием было бы получение результатов для $A\% \times B\%$ населения, где $A\%$ – это доля выборки из первого источника, а $B\%$ – это доля выборки из второго источника. Например, увязка обследования с 20-процентной долей выборки и обследования с 5-процентной долей выборки даст результаты максимум для 1 процента населения. Если один источник содержит выборку из 1000 человек, представляющую аналитический интерес, и связан с другим источником с 20-процентной долей случайной выборки, то интегрированные результаты будут доступны максимум для 200 человек.
156. Ограничения по охвату усугубляют эти ограничения. Если временной охват отличается в разных источниках, возможно, будут включены только те лица, которые охвачены обоими периодами времени. Если один файл представляет собой перепись мигрантов, а другой - перепись записей о госпитализации, то без ошибки интеграции результаты будут доступны только для тех мигрантов, у которых была запись о госпитализации. Если один файл представляет собой перепись мигрантов, а другой - разовое выборочное обследование, то в лучшем случае результаты будут доступны для тех мигрантов, которые ответили на это выборочное обследование.
157. Пользователям необходимо будет рассмотреть, в какой степени они готовы работать с этими ограничениями или устранять их впоследствии. Если выбран источник с ограничениями по выборке или охвату, следующий этап рассмотрит это более подробно, а также возможность устранения этих ограничений (например, с использованием весов выборки). Однако, если ожидаемая интегрированная выборка явно слишком мала для какого-либо значимого анализа, это должно привести к корректировке выбора исходных файлов и статистического проектирования в целом.
158. Подобно временному охвату, охват популяции часто определяется тем, как собираются административные данные или как проводятся обследования. Как таковая, может существовать большая вариативность в популяции, охватываемой источниками данных разных стран или даже между разными источниками в одной и той же стране.

В Бельгии каждое лицо, законно проживающее в Бельгии (более 3 месяцев), зарегистрировано в ЦРН. Соискатели убежища также включены, хотя они регистрируются не муниципалитетами, а централизованно Иммиграционной службой.

Поскольку все переменные, касающиеся миграции в других источниках (переписи, ОРС и ОДУЖ), происходят из ЦРН, те же группы могут быть идентифицированы в этих источниках.

В Канаде также существуют различия в популяции, охватываемой различными доступными источниками. Лонгитюдная база данных по иммигрантам (IMDB) объединяет иммиграционные и налоговые административные записи. Хотя выборка не проводится, исходы доступны только для связанных налогоплательщиков. Это означает, что определенные популяции, которые с меньшей вероятностью подают налоговые декларации (например, дети), недостаточно представлены в файлах Исходов. Чтобы решить эту проблему, IMDB недавно выпустила новые файлы исходов для детей, связывая их с экономическими результатами их родителей и облегчая анализ межпоколенческой экономической мобильности.

Административные записи IRCC содержат исчерпывающую информацию об иммигрантах, прибывших в Канаду с 1952 года (включая данные о гражданстве для тех, кто стал гражданином Канады с 2005 года). Они также содержат временных резидентов, которым были выданы временные виды на жительство, и заявителей на убежище с 1980 года. Однако они не включают тех, кому не требуется разрешение на проживание в Канаде (например, посетители, некоторые иждивенцы временных резидентов и т.д.).

Лонгитюдное обследование иммигрантов в Канаду (LSIC) имело целевую популяцию, состоящую из иммигрантов, которые соответствовали всем следующим критериям:

- Прибыли в Канаду между 1 октября 2000 года и 30 сентября 2001 года
- Были в возрасте 15 лет или старше на момент прибытия
- Прибыли из-за границы, должны были подать заявление через канадскую миссию за границей.

Лица, которые подали заявление и прибыли изнутри Канады, были исключены из обследования. Эти люди могли находиться в Канаде в течение значительного времени до официального «прибытия» и, следовательно, вероятно, демонстрировали иные характеристики интеграции по сравнению с недавно прибывшими в Канаду. Беженцы, запрашивающие убежище изнутри Канады, также были исключены из объема обследования.

Интересующей популяцией LSIC были иммигранты в целевой популяции, которые все еще проживали в Канаде на момент интервью. Кроме того, для целей обследования проводилась выборка — на это дополнительно повлиял уровень отсутствия ответов. Из примерно 165 000 иммигрантов, входящих в объем этого обследования, около 12 000 ответили на первую волну, 9 300 — на волну 2 и 7 700 — на волну 3.

В Германии Центральный реестр иностранцев включает только иностранцев, будь то первое/второе поколение, соискатели убежища, беженцы и т.д.

В Италии регистр населения включает всех людей, которые меняют муниципалитет своего обычного проживания. Этот административный регистр исключает иностранцев в нелегальном положении. В случае видов на жительство охватываемая популяция — это граждане, не принадлежащие к странам ЕС, и законно проживающие в Италии. Для обследования интеграции второго поколения выборочная рамка была сформирована из 31 700 иностранных студентов в отобранных средних школах.

В Казахстане Статистический регистр населения (СРН) содержит записи о резидентах Республики Казахстан, а также о гражданах Республики Казахстан, временно проживающих за границей. Большинство иностранных мигрантов в Казахстан не поселяются постоянно (т.е. не получают вид на жительство), а допускаются по трудовому договору и регистрируются временно.

База данных «Перепись 2009 года» содержит информацию о резидентах или временно проживающих в Казахстане на момент проведения переписи.

Существует два подходящих источника для статистики миграции. Информационные системы Министерства внутренних дел содержат данные о внутренней и внешней миграции. Информационная система «Беркут» содержит записи о лицах, пересекающих границу Республики Казахстан для въезда/выезда из страны на постоянное жительство и получения гражданства страны.

В Испании Регистр населения охватывает все население, проживающее в Испании, независимо от их правового статуса и национальности (включает граждан и иностранцев).

В Швейцарии Швейцарская лонгитюдная демографическая база данных (SLDD) содержит информацию о швейцарских гражданах, которые проживали в Швейцарии с 2010 по 2017 год, и иностранцах (с видами на жительство), которые проживали в Швейцарии между 1998 и 2017 годами. Это включает соискателей убежища и натурализованных иммигрантов, но не людей в нелегальном положении.

В Турции Адресная система регистрации населения (ABPRS) является уникальной системой, из которой распространяется ежегодная статистика населения. В этом объеме ежегодная информация о запасах и потоках международной миграции по гражданству и стране рождения производится на основе ABPRS и других дополнительных административных источников.

Охват лиц в ежегодных данных ABPRS может варьироваться в зависимости от конкретных условий, таких как поздняя регистрация смертей, поздняя регистрация детей до пяти лет, изменение идентификационного номера после получения гражданства (т.е. потеря следа натурализованного лица) и недостаточное/ избыточное покрытие предыдущего/ нынешнего года ABPRS.

Хотя реестры иностранцев работают аналогично регистрам населения, они охватывают только иностранцев, имеющих законное место жительства в стране. В этом отношении на первый план выходят проблемы недостаточного покрытия (для тех, кто отсутствует в административных реестрах, находясь в стране) и избыточного покрытия (для тех, кто есть в административных реестрах, но отсутствует в стране). Поэтому регистрация и снятие с регистрации иностранцев имеют важное значение для определения международных мигрантов. Хотя это уже не такая большая проблема, как в прошлом, декларация имеет важное значение для иностранцев, которые прожили в стране определенное время или которые намерены остаться по крайней мере на определенный срок и которые покинут страну.

3.3.3.2 Периодичность Результатов

159. Помимо временного охвата, также представляет интерес то, как часто собираются или обновляются данные. Если результаты измеряются только периодически, они могут не учитывать более частые события (или интенсивность событий), которые лучше фиксировать в более сжатые периоды времени или даже в режиме реального времени. Примером может служить прибытие большого количества иностранных студентов в начале учебного года. Более подробные соображения, влияющие на измерение времени, рассматриваются в Главе 2.

160. Возможные файлы Результатов следует рассматривать с точки зрения того, как часто измеряются события по сравнению с заявленными целями проекта. Также следует отметить, что возможные исходные файлы могут не идеально соответствовать целям, но все же могут давать полезные результаты. Например, файлы Результатов могут иметь только ежегодные результаты. Это может не дать представления о том, через сколько дней или недель мигранты находят работу, но их все же можно использовать, чтобы понять, как мигранты получали работу по годам.

161. Также важно отметить разницу между сбором данных и данными, доступными для статистического проекта, поскольку между этими двумя процессами могут быть задержки.

В Бельгии Статистическое управление Бельгии получает двойной приток данных из своего регистра (CPR): раз в год, в марте, сводную информацию о полном реестре в том виде, в каком он существует на данный момент, и каждые две недели - все данные о регистрации за предыдущие две недели. Годовой приток используется для расчета всей демографической статистики, поскольку постоянный приток служит для обновления рамок выборки для социальных обследований.

Таким образом, можно постоянно получать результаты сечения. Тем не менее, ошибочные регистрации частично исправляются и удаляются. В результате могут возникнуть расхождения между двумя разными сечениями для одного и того же момента, выполненными в разные моменты времени. Работа со случаями, присутствующими в одном поперечном сечении, но полностью отсутствующими в более поздней версии, оказалась сложной задачей.

В Канаде существуют различия в том, как информация обновляется в различных доступных источниках. В административных записях IRCC лонгитюдные результаты, связанные с видами на жительство, можно наблюдать в реальном времени (например, точные даты выдачи разрешений). Большинство этих файлов передаются в Статистическое управление Канады на ежемесячной основе. Однако для Лонгитюдной базы данных по иммигрантам (IMDB) исходы из налоговых файлов наблюдаются ежегодно и в основном состоят из информации, предоставляемой лицами при подаче налоговых деклараций в первые несколько месяцев после окончания отчетного года (например, лица, подающие налоговые декларации за 2017 год, обычно делают это в первые несколько месяцев 2018 года).

Результаты для Лонгитюдного обследования иммигрантов в Канаду (LSIC) наблюдались в конкретные моменты времени: через шесть месяцев (2001-02), два года (2002-03) и четыре года (2004-05) после прибытия в Канаду.

В Германии Центральный реестр иностранцев содержит информацию обо всех иностранцах (включая соискателей убежища или беженцев), в настоящее время зарегистрированных в Германии (активные) или ранее зарегистрированных в Германии (неактивные). Центральный реестр иностранцев постоянно обновляется, и статистические данные относятся к 31 декабря каждого года.

В Италии Anvis (Виртуальная статистическая база данных регистров) – это статистическая система, полностью управляемая Istat, которая на основе информации, поступающей из регистров населения, позволяет проводить мониторинг на индивидуальной основе и предоставлять результаты на ежегодной основе. Информация о видах на жительство обновляется ежегодно.

Данные итальянского обследования интеграции второго поколения, проведенного в 2015 году, связываются ежегодно с административными данными, поступающими из регистра студентов.

В Казахстане данные регулярно передаются в Статистический регистр населения (СРН) административными источниками и базой данных переписи 2009 года. Это делается как систематически, в случае возникновения события, так и через регулярные промежутки времени.

В Испании муниципалитеты постоянно обновляют информацию о регистрации и снятии с регистрации в Padrón. Кроме того, Реестр актов гражданского состояния также обновляет информацию, касающуюся рождений и смертей. Эти обновления отправляются каждый месяц в INE, который отвечает за обновление своей базы данных. Таким образом, можно получать точные данные в любое время года с обновленной информацией.

Дата регистрации в Padrón обычно близка к фактической дате въезда в страну, которая неизвестна из-за преимуществ, которые влечет за собой регистрация. С другой стороны, дата снятия с учета в большинстве случаев не совпадает с датой выезда гражданина из страны, поскольку он покидает страну без какого-либо заявления.

В Швейцарии Лонгитюдная демографическая база данных (SLDD) — это база данных на основе связанных регистров, созданная в два разных этапа. Сначала строится демографическая база путем связывания Центрального реестра иностранных граждан (ZAR, впоследствии ZEMIS) и ежегодной статистики населения и домохозяйств (STATPOP) Федерального статистического управления (FSO).

На втором этапе база данных также организуется для включения другой статистики. До сих пор Структурное обследование (с 2010 по 2014 год) и данные реестра социального страхования были связаны с SLDD, предоставляя полезную информацию о социально-экономических характеристиках жителей Швейцарии.

Наблюдаемые лонгитюдные исходы SLDD обновляются ежегодно. База данных организована таким образом, который позволяет ее регулярно расширять год за годом. Такое обновление увеличит период наблюдения и, следовательно, способность базы данных предоставлять полезную информацию о миграции и интеграции.

3.3.3.3 Различия в измерениях

162. Важно оценить, соответствуют ли измерения — либо популяций, либо результатов — концепциям, заявленным в целях проекта.
163. Здесь может быть некоторое наложение с оценкой периодичности файлов как исходов в реальном времени. Например, трудоустройство может наблюдаться как ежегодная сводная переменная. Однако могут быть наложения с охватом популяции, поскольку определение интересующих групп населения может отличаться от того, что наблюдается в источниках данных.
164. На следующем этапе некоторые различия могут быть устранены с помощью предварительной обработки, особенно для переменных, необходимых для интеграции данных. С другой стороны, различия в аналитических переменных могут быть устранены на этапе 5, когда контент гармонизируется.
165. Особого внимания заслуживает то, как различия между измерениями и концепциями могут относиться к мигрантам. Некоторые из этих различий могут быть обусловлены:
 - Правовым статусом мигрантов: некоторые источники собирают информацию обо всех мигрантах (независимо от их правового статуса в новой стране проживания), в то время как другие включают только тех, кто имеет определенный статус проживания (например, имеющих действительный вид на жительство).
 - Минимальной продолжительностью пребывания для классификации в качестве мигранта: различия в концепции мигранта, которая используется, в зависимости от того, учитывается ли минимальное пребывание в стране (например, 12 месяцев).
166. При изучении опыта мигрантов могут быть важны различия в географических концепциях. Например, изменения границ с течением времени или различия в концепциях между разными источниками (например, район проживания либо район предполагаемого проживания).
167. Различия между концепциями и фактическими измерениями исследуются в разделе 4.1 о лонгитюдных показателях, где интересующие показатели противопоставляются ограничениям того, как данные могут быть собраны.

Например, в Канаде все переменные географии в иммиграционных и налоговых файлах Лонгитюдной базы данных по иммигрантам (IMDB) обычно кодируются на основе географии предыдущей переписи. География переписи обновляется каждые 5 лет. Кроме того, концепция географии назначения (найденная в иммиграционных файлах) отличается от географии проживания. Хотя эти две объединяются для понимания местного «удержания» недавних иммигрантов, возможно, что некоторые лица никогда не проживают в своей географии назначения.

Ежегодные налоговые отчеты на IMDB позволяют получить подробную информацию об экономических результатах иммигрантов в Канаде. Однако, поскольку эти файлы представляют собой ежегодные сводки, некоторые детали не полностью отражены. Например, хотя в ежегодных документах указывается, сколько иммигрантов имели доход от трудовой деятельности в данном году (и в каком размере), они не могут указать уровень занятости за определенный момент времени или базовую неделю. Аналогичным образом, существуют переменные, позволяющие отдельным лицам претендовать на оплату обучения после окончания средней школы (показатель участия в обучении после окончания средней школы), но подробная информация об уровне или основной области обучения или результатах образования недоступна.

Налоговые файлы также могут предоставить некоторые из тех же демографических переменных, что и иммиграционные файлы (такие как дата рождения — или возраст). Могут возникать несоответствия между двумя источниками информации для одного и того же лица. Эти различия необходимо будет устранить либо на этапе предварительной обработки для облегчения интеграции данных (3.4), либо для гармонизации концепций для распространения (3.7).

Центральный реестр иностранцев Германии содержит информацию обо всех иностранцах (включая соискателей убежища или беженцев), в настоящее время зарегистрированных в Германии (активные) или ранее зарегистрированных в Германии (неактивные).

Одна из самых важных задач — разработать обходной путь при работе с изменениями гражданства с течением времени или с изменениями национальных административных границ.

В Испании все иностранцы перечислены в Регистре населения (Padrón) вместе с населением, рожденным в Испании. Хотя невозможно определить количество беженцев или иностранцев в нелегальном положении, считается, что Padrón практически включает всех иностранных резидентов из-за множества преимуществ регистрации.

В регистре населения Италии информация о нелегальных мигрантах исключена, и правило 12 месяцев считается верным, поскольку при сборе данных о миграции не применяются временные критерии.

Источник данных о видах на жительство содержит информацию только о гражданах не из ЕС. В случае обследования интеграции второго поколения данные, касающиеся видов на жительство, были интегрированы с данными о приобретении гражданства, Регистром населения и Регистром социального обеспечения для мониторинга процесса включения граждан не из ЕС. Эта интеграция была проведена для отслеживания студентов в выборке после обследования и для сравнения ожиданий и мнений студентов, выраженных в обследовании, с их реальным поведением. Источники данных подвержены изменениям в административных процедурах и различным географическим границам. Эти изменения могут повлиять на сопоставимость информации с течением времени.

В Соединенном Королевстве Лонгитюдное исследование УНС для Англии и Уэльса (LS) учитывает изменения географических границ с течением времени. Изначально это вытекало из запроса пользователя посмотреть на потоки внутренней миграции за 2011 год на основе старых границ переписи 1991 года для местных органов власти.

Самая маленькая географическая область для переписи и некоторых оценок населения — это Выходная область. Это строительный блок для создания более высоких уровней географии, таких как Районы местных органов власти (LAD). Эксперты по географии в УНС создали соответствия между Выходными областями (и Местными органами власти) между разными годами переписи.

«Было проведено прямое распределение районных территорий местных органов власти 1991 года по отдельным переписным листам 2011 и 2001 годов. Каждая запись о каждом участнике переписи 2011 года была нанесена на карту на основе привязки к сетке, связанной с этой записью, и привязана к цифровой границе LAD 1991 года, в пределах которой она находилась. С помощью этого метода все точки можно было отнести к области. Таким образом, 585 800 записей участников LS за 2011 год были переданы одному из 403 местных органов власти, существовавших в 1991 году. Было принято решение о прямом распределении средств таким образом. Это отличается от метода, предпочитаемого географической политикой Государственной статистической службы Великобритании (GSS), согласно которому данные были бы распределены по производственному району, а затем наилучшим образом соответствовали бы данным местных органов власти за 1991 год. Было принято решение о прямом распределении с использованием привязки к таблице, поскольку политика GSS ориентирована на производителей табличной статистики, в то время как основной целью лонгитюдного исследования является наблюдение за отдельными лицами с течением времени.»

3.4. ЭТАП 2: ОЦЕНКА И ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ИСХОДНЫХ ФАЙЛОВ

168. На Этапе 1 были определены возможные исходные файлы и оценена их способность наилучшим образом соответствовать целям проекта на основе охватываемых ими групп населения, их временного охвата и включаемых показателей. На данном этапе следует проконсультироваться с поставщиками данных и проинформировать их о планируемой интеграции их данных в лонгитюдные данные, поскольку они могут знать о потенциальных проблемах с качеством данных или ожидаемых изменениях в содержании. Поставщики данных являются ключевыми партнерами, поскольку они лучше всех знают данные и со временем могут сообщать об изменениях в данных. Кроме того, они обладают знаниями о том, как и в каком контексте собирались данные. Для административных данных, которые не собирались для статистических целей, это особенно важно, поскольку это может привести к определенным ограничениям качества данных. По завершении этого этапа будут определены наилучшие доступные исходные файлы для проекта.
169. На Этапе 2 качество этих исходных файлов будет оценено более тщательно. На основе этой оценки может потребоваться обработка (например, подстановка пропущенных значений) для устранения любых выявленных проблем с качеством. Качество данных исходных файлов следует оценивать и документировать до и после любой обработки. Рекомендуется обратиться к рамкам, изложенным в разделе 3.2, для получения дальнейших указаний относительно различных источников ошибок.
170. Если на этом этапе будут обнаружены какие-либо существенные ограничения качества данных, может возникнуть необходимость вернуться к Этапу 1 и рассмотреть новые источники данных.
171. Наконец, файлы будут предварительно обработаны для подготовки их к интеграции на Этапе 3. На этом этапе будет оценено качество полей, используемых для облегчения интеграции данных. Для лонгитюдной базы данных эти этапы необходимо выполнять каждый раз при обновлении базы данных, поскольку изменения могут происходить регулярно.

3.4.1 Оценка качества исходных файлов

172. На этом этапе следует оценить качество данных каждого исходного файла. Оценка качества должна учитывать ошибки охвата и измерения, а также любые ошибки, которые могут повлиять на этап интеграции. В приведенном ниже списке представлены примеры ошибок, которые можно исследовать, вместе с предложениями по их документированию и возможными методами устранения.
173. Для целей настоящего руководства исходные файлы по-прежнему подразделяются на два типа (в контексте продольных данных из интегрированных исходных файлов):
- Файлы когорт (в рамках классификации ошибок они называются «объектами»)
 - Файлы результатов (в рамках классификации ошибок они называются «атрибутами»)
174. Файлы когорт – это исходные файлы, используемые для идентификации интересующей(их) совокупности(ей) и их характеристик, тогда как файлы результатов – это лонгитюдные файлы, предоставляющие результаты с течением времени. В некоторых случаях файлы результатов также могут служить файлами когорт.

3.4.1.1 Проблемы, влияющие на охват

175. Существуют различные типы проблем, которые могут повлиять на охват набора лонгитюдных данных. К ним могут относиться ошибки в охвате, ошибки выборки и другие источники недостающих записей, такие как неполучение единичных данных.

3.4.1.1.1 Ошибка охвата

176. Здесь рассматривается степень, в которой источник данных, при отсутствии пропущенных записей и выборки, был бы репрезентативен для целевой совокупности. Фактически, обычно существует два вида ошибок охвата:

- Недостаточный охват: когда единицы наблюдения должны быть в сфере охвата источника данных, но не охвачены им
- Избыточный охват: когда единицы наблюдения должны быть вне сферы охвата источника данных, но включены в него (включая дубликаты единиц, входящих в сферу охвата)

177. Эта оценка должна опираться на обзор, проведенный на Этапе 1, который требовал оценки ограничений охвата, связанных с выбранными исходными файлами, и того, как они могут повлиять на цели, определенные в статистическом дизайне.

178. Для файлов когорт ошибки охвата отражают разницу между интересующей совокупностью и охватом файла когорты.

-- --

5. Данный раздел основан на ЕЭК ООН (2019), который рекомендуется использовать в дополнение к настоящему Руководству.

Например, в Лонгитюдной базе данных по иммигрантам (Канада) в качестве файла когорты используются административные записи об иммиграции. Мигранты без административных записей (т.е. те, у кого нет видов на жительство или заявлений на убежище) обязательно будут исключены файлом когорты – это включает посетителей и некоторые другие категории путешественников. Следовательно, эти файлы будут недостаточно охватывать общее число лиц, въезжающих в Канаду.

С другой стороны, эти файлы могут включать несколько записей для одного и того же лица (с разными уникальными идентификаторами). Это приведет к избыточному охвату в файле когорты.

179. Для файлов результатов ошибки охвата следует рассматривать аналогичным образом. Напомним, что файлы когорт будут использоваться для идентификации интересующей совокупности. Важно будет учитывать охват файлов результатов в контексте как генеральной совокупности, так и интересующей совокупности.

180. Оценка избыточного охвата в файлах результатов может ограничиваться идентификацией дубликатов. Процесс интеграции ограничит итоговый набор данных файлом когорты и, таким образом, устранил другие единицы, не входящие в сферу охвата.

Например, Канадская лонгитюдная база данных по иммигрантам связывает файлы по иммиграции (Когорта) с ежегодными налоговыми файлами (Исходы). Охват налоговых файлов имеет особенности, которые влияют как на генеральную совокупность (например, дети реже подают налоговые декларации и, следовательно, недостаточно охвачены), так и на конкретную интересующую совокупность (например, временные иностранные работники реже подают налоговые декларации и, следовательно, недостаточно охвачены).

В Испании Реестр населения (Padrón) непрерывно функционирует с 1996 года. Он отражает место, где человек был зарегистрирован в определенный момент времени, и содержит некоторые конкретные переменные: пол, возраст/дата рождения, место рождения, гражданство, имя и фамилия, уровень образования или официальный идентификационный номер.

Padrón не только хранит местонахождение человека в конкретный момент времени, но и, как лонгитюдная запись, содержит все места, где человек был зарегистрирован с 1996 года.

Padrón является идеальным источником для получения лонгитюдной информации о мигрантах. Однако у него есть ограничение: его начальная точка – 1996 год, поэтому информация за предыдущие годы отсутствует. Чтобы решить эту проблему, информация из Padrón будет связана с Переписью 2001 года (которая была сплошной), и, таким образом, информация станет доступной для всех соответствующих лиц.

В Венгрии хорошо известно, что административные источники недооценивают реальные масштабы потоков эмиграции, поскольку мигранты часто не сообщают властям о своих миграционных событиях. Чтобы выявить подгруппы мигрантов, потоки которых систематически недооцениваются в административных источниках, административные данные сравнивались с выборочными данными Микрпереписи. Это сравнение выявило основные слабые стороны административных источников, особенно когда речь идет о различных формах множественной миграции.

3.4.1.1.2 Ошибка выборки

181. В некоторых проектах интеграции данных по крайней мере один из исходных файлов может представлять только выборку из охваченной совокупности. Чаще всего это является результатом того, что один из исходных файлов представляет собой выборочное обследование или, что то же

самое, административное подмножество данных целевой совокупности. Существует несколько различных ситуаций, которые могут повлиять на лонгитюдный источник данных, основанный на интегрированных данных.

182. Распространенным примером выборочных файлов когорты без выборки в файлах результатов является случай, когда выборочное обследование связывается с административными файлами для создания лонгитюдного набора данных. Выборочное обследование определяет когорту (наряду с определенными интересующими характеристиками), а административные файлы предоставляют лонгитюдные результаты.
183. В этой ситуации файлы когорты должны иметь вспомогательную информацию, связанную с выборкой (включая веса), а также руководство по использованию файлов для статистического анализа. Эта вспомогательная информация должна оставаться действительной для учета ошибки выборки, связанной с результирующей лонгитюдной базой данных.
184. Примером выборочных файлов исходов без выборки в файлах когорты является ситуация, когда лонгитюдное выборочное обследование проводится с использованием административных данных в качестве основы. Административные данные определяют когорту (наряду с определенными интересующими характеристиками), а лонгитюдное панельное выборочное обследование предоставляет лонгитюдные результаты.
185. В этой ситуации файлы исходов должны иметь вспомогательную информацию, связанную с выборкой (включая веса), а также руководство по использованию файлов для статистического анализа. Эта вспомогательная информация должна оставаться действительной для учета ошибки выборки, связанной с результирующей лонгитюдной базой данных.

Например, Лонгитюдное обследование иммигрантов в Канаде (LSIC) использовало административные записи об иммиграции в качестве основы для выборки.

186. Сценарий выборочных файлов когорты с выборочными файлами когорты или любая другая комбинация выборки может быть менее распространенным, особенно поскольку результатом интеграции выборок является меньшая выборка (выборка из выборки). Это может быть особенно проблематично при изучении мигрантов (уже являющихся частью генеральной совокупности). Однако в случаях, когда интегрируемые выборки достаточно велики, это все же может дать полезный статистический набор данных.
187. Для учета более сложных ситуаций с выборкой необходимо использовать продвинутые статистические методы, которые считаются выходящими за рамки данного Руководства. Однако важно, чтобы, как и в случае первых двух типов выборки, которые могут повлиять на лонгитюдные источники данных, ошибки выборки измерялись и устранялись.

3.4.1.1.3 Пропущенные записи

188. Единицы могут отсутствовать в исходных файлах когорт или результатов. Одной из распространенных причин этого может быть отсутствие ответа от единиц.
189. Отсутствие записей в файлах когорты повлияет на любой последующий анализ, поскольку затронутые единицы не могут быть сопоставлены ни с какими исходами.
190. Отсутствие записей в файлах результатов, с другой стороны, может иметь различные последствия. Эти пропущенные записи могут быть результатом выбывания с течением времени, когда имеется полная информация о результатах за более ранние периоды. Кроме того, это могут быть эпизодически пропущенные записи или даже записи, которые отсутствуют изначально, но становятся полными за более поздние периоды.
191. Миграционные модели могут смешиваться с моделями пропущенных исходов. Например, отсутствие исходов связано с пропущенными записями или с тем, что человек не находился в стране в тот период времени? В некоторых случаях миграционное событие можно измерить, но в других случаях оно остается неизвестным.

192. В зависимости от характера пропущенных записей на этапе анализа могут применяться различные методы.

193. Некоторые вопросы для рассмотрения при оценке проблем охвата:

- Существуют ли отчеты, в которых документированы ограничения, влияющие на охват файлов когорты или результатов?
- Какие группы мигрантов или контрольные группы охватываются файлами когорты?
- Какие группы мигрантов или контрольные группы не охватываются файлами когорты?
- Есть ли дубликаты в файлах когорты?
- Если имеется несколько файлов когорты, есть ли дубликаты между файлами?
- Какие группы охватываются файлами результатов?
- Какие группы не охватываются файлами результатов?
- Существуют ли причины, по которым файлы результатов могут не охватывать определенные группы мигрантов?
- Есть ли дубликаты в файлах результатов?
- Применяется ли выборка к файлам когорты или результатов? Если да, то каков объем выборки с точки зрения числа и доли выборки (для генеральной совокупности и интересующей совокупности мигрантов)? Какие методы выборки использовались?
- Отсутствуют ли записи в файлах когорты или результатов? Какой процент отсутствует? Существуют ли группы, для которых вероятность пропуска выше или ниже?
- Если в файлах результатов отсутствуют записи, отсутствуют ли они эпизодически, только в начале или только в конце?
- Затрагивают ли пропущенные записи в файлах когорты или результатов мигрантов непропорционально сильно?

194. Отчет, документирующий ограничения охвата, может включать различные показатели качества данных, такие как:

1. Различия в распределении между целевой и охваченной совокупностями (на основе источников данных для сравнения⁶)

a. *Возраст и пол*

b. *Миграционный статус*

c. *Другие интересующие поля*

2. Расчетный валовой и чистый недостаточный охват

3. Количество и процент дубликатов

4. Объем выборки (число и процент) по миграционному статусу в файлах когорты или результатов

5. Количество и процент пропущенных записей в файлах когорты

6. Количество и процент пропущенных записей в файлах результатов по периодам

3.4.1.2 Проблемы, влияющие на измерение

195. Также необходимо оценить пропущенные значения и ошибки измерения в исходных файлах. Обе эти проблемы могут затрагивать как файлы когорты, так и файлы результатов.

196. Хотя пропущенные значения можно довольно легко идентифицировать, ошибки измерения, когда измеренное значение отличается от истинного, может быть трудно обнаружить. В некоторых случаях значения могут быть недействительными (например, значения, которые не существуют или являются неправдоподобными). Например, дата рождения после текущей даты или, в случае набора данных, охватывающего ныне живущих людей, дата рождения сотни лет назад.

197. В частности, важно учитывать качество полей, которые будут использоваться для облегчения процесса интеграции (например, поля, пригодные для связи, для сопоставления записей, такие как дата рождения).

198. Некоторые вопросы для рассмотрения при оценке проблем измерения:

- Существуют ли отчеты, в которых документированы ограничения, влияющие на измерение для файлов когорты или результатов?
- Как собирались интересующие переменные и в каком контексте (например, конкретная административная цель может влиять на собираемые значения)?

- Имеются ли пропущенные значения в файлах когорты? Если да, то затрагивают ли эти пропущенные значения непропорционально сильно определенные группы мигрантов? Какие переменные затронуты и в какой степени?
- Имеются ли пропущенные значения в файлах результатов? Отсутствуют ли они постоянно у одних и тех же людей? Затрагивают ли пропущенные значения непропорционально сильно определенные группы мигрантов? Какие переменные затронуты и в какой степени? Есть ли последствия для аналитических результатов?
- Имеются ли недействительные значения в файлах когорты или результатов? Если да, то затрагивают ли эти недействительные значения непропорционально сильно определенные группы мигрантов? Какие переменные затронуты и в какой степени?
- Есть ли переменные в файлах когорты или результатов, которые указывают на противоречия (например, люди, родившиеся после года их миграции)? Есть ли дополнительные переменные, которые можно использовать для определения, какое значение с большей вероятностью является правильным в случае противоречия?

-- --

2. «Данные для сравнения» относятся к сопоставимым данным, с которыми можно сравнивать результаты.

- Каковы некоторые риски других ошибок измерения? Затрагивают ли они непропорционально сильно группы мигрантов?
- Как показатели соотносятся с другими источниками данных?
- Затронуты ли поля привязки в файлах когорт или результатов из-за отсутствия или недопустимых значений? В какой степени проблемы, связанные с объединением файлов, совпадают?

199. Отчет, документирующий ограничения измерения, может включать различные показатели качества данных, такие как:

1. Частота и процент пропущенных значений по переменным, периодам и миграционному статусу
 - a. Следует уделить особое внимание внутрииндивидуальным паттернам, так как определенные люди могут с большей вероятностью иметь множественные пропущенные или недействительные значения
2. Частота и процент недействительных значений по переменным, периодам и миграционному статусу
 - a. Следует уделить особое внимание внутрииндивидуальным паттернам, так как определенные люди могут с большей вероятностью иметь множественные пропущенные или недействительные значения
3. Различия в распределении переменных между сопоставимыми совокупностями (на основе источников данных для сравнения – т.е. альтернативных источников с сопоставимыми показателями)
 - a. Возраст и пол
 - b. Миграционный статус
 - c. Другие интересующие поля
4. Различия в значениях переменных между связанными файлами с сопоставимыми показателями
5. Несоответствия между переменными из одних и тех же исходных файлов
 - a. Вспомогательные переменные или дополнительная информация, которая может поддержать одно значение над другим
 - b. Предпочтение на основе наиболее часто встречающегося значения

3.4.2 Предварительная обработка исходных файлов

200. Теперь, когда качество данных исходных файлов оценено, пришло время подготовить их к интеграции данных и, при необходимости, устранить любые известные проблемы с качеством данных.

201. Конкретные методы, связанные с предварительной обработкой исходных файлов, не входят в настоящее Руководство. Некоторые элементы рассматриваются в руководстве ЕЭК ООН по интеграции данных для измерения миграции (ЕЭК ООН, 2019b), другие хорошо описаны в общей статистической литературе.

3.4.2.1.1 Подготовка файлов к интеграции данных

202. Предварительная обработка исходных файлов когорты и исходов начинается с:

- a. Стандартизации форматов и длин переменных для сопоставления (например, единообразный формат даты рождения)

b. Стандартизации формата имен и обработки специальных символов (например, верхний регистр, удаление акцентов)

203. Эти шаги критически важны для интеграции между файлами когорты и исходов. Здесь есть некоторые элементы, представляющие особый интерес для статистики миграции. К ним относятся обработка написания имен и обеспечение единообразия порядка следования имени и фамилии между файлами. Распространенные проблемы, связанные с именами, которые могут непропорционально сильно затрагивать мигрантов, могут включать:

- Изменение фамилии после замужества (особенно если миграция связана с браком)
- Принятие имени, распространенного в принимающей стране, вместо оригинального имени
- Частичный выбор имени или фамилии в определенных источниках данных (например, возможно, из-за ограничения места в государственной форме)
- Акценты, которые могут фиксироваться не всегда последовательно во всех источниках данных
- Порядок следования имени и фамилии не всегда фиксируется единообразно (например, если фамилия обычно ставится перед именем)
- Различные варианты написания, обусловленные непоследовательной транслитерацией имен между источниками (т.е. передачи имен из одной системы письма в другую)

204. Поля с датами (например, дата рождения) также нуждаются в стандартизации. Это включает обеспечение правильного порядка года, месяца и дня, а также устранение ситуаций, когда элементы отсутствуют или кажутся ошибочными. Одна из распространенных проблем, связанных с административными данными, – аномально большое количество записей с датами рождения, приходящимися на 1 января. Это может быть значением по умолчанию вместо истинно пропущенных значений. Поскольку мигранты могут быть не знакомы с местным календарем, это может привести к неопределенности в отношении точной даты рождения и к большому количеству записей только с действительным годом рождения.

205. Когда отбор выборки основан на календарной дате рождения, как в Лонгитюдном исследовании Управления национальной статистики Великобритании (UK ONS), использование «стандартных» дат рождения для некоторых мигрантов, родившихся за границей, в частности из Южной Азии, приводит к смещению выборки.

206. Кроме того, если для интеграции используется контактная информация, ее актуальность может повлиять на результаты. Поскольку мигранты, по определению, мобильны, их контактная информация, возможно, более подвержена изменениям с течением времени.

207. Помимо стандартизации и очистки исходных файлов, другой рекомендуемой практикой является использование информации, зафиксированной в различные моменты времени, для облегчения процесса связывания. Например, если файлы исходов связываются детерминистически, они могут предоставлять варианты имен и контактной информации с течением времени. Учет этого улучшит качество интеграции.

208. Хотя интеграция многих файлов может привести к различным ошибкам связывания, таким как ложноположительные или ложноотрицательные связи, использование лонгитюдных файлов связывания снижает риск ошибки интеграции. Если один источник способен надежно отслеживать изменения в полях связывания (например, имя(ена), адреса и т.д.), его использование в процессе связывания снизит риск ложноотрицательных результатов (пропущенных связей) из-за изменяющихся характеристик, указанных выше.

В Канаде Лонгитюдная база данных по иммигрантам (IMDB) использует преимущества репозитория информации для связывания, собранной за время наблюдения. Социальная среда связывания данных (SDLE) используется для облегчения сопоставления записей. До этого IMDB полагалась на лонгитюдный файл связывания, состоящий из налоговой информации, связанной через номер социального страхования (SIN). Этот источник отслеживал изменения в полях связывания с течением времени.

3.4.2.1.2 Внутреннее связывание записей до этапа интеграции для идентификации дубликатов

209. Дубликаты в файлах данных о когортах или результатах могут быть идентифицированы с помощью внутренней или внутрифайловой привязки записей. Это может быть полезным упражнением даже для источников с уникальными идентификаторами записей, таких как статистические регистры. В

частности, поскольку мигранты могут повторно въезжать в страну несколько раз, они подвергаются значительному риску получения более чем одного уникального идентификатора с течением времени. Осуществление этого процесса облегчило бы этап 4 по поддержанию единообразия идентификаторов с течением времени.

В случае Канады физические лица могут фигурировать в нескольких административных записях Иммиграционного, беженского и гражданского ведомства Канады (IRCC). Например, человек может иметь несколько разрешений на временное проживание (например, сначала иностранный студент, затем временный иностранный работник), прежде чем стать постоянным резидентом, а затем может получить гражданство. Будучи интегрированными, эти файлы позволяют проводить комплексный анализ опыта, связанного с иммиграцией, в Канаде. Данные сначала интегрируются с использованием общих индивидуальных идентификаторов. Эти идентификаторы выдаются в административном порядке и предназначены быть единственными и уникальными для иммигрантов и временных резидентов в Канаде. Однако при разработке Лонгитюдной базы данных по иммигрантам (IMDB) Статистическое управление Канады выявило дубликаты (случаи, когда у одного человека было несколько идентификаторов личности IRCC) с помощью внутреннего связывания записей. Хотя они редки для иммигрантов (рабочее предположение состоит в том, что они не возникают), они более распространены среди временных резидентов. Статистическое управление Канады устраняет эти дубликаты при замене идентификатора личности IRCC на новый анонимизированный идентификатор IMDB. Этот процесс более подробно объясняется в разделе 3.6.

3.4.2.1.3. Устранение других проблем качества данных, связанных с исходными файлами

210. Это относится к статистическим методам, хорошо описанным в существующей литературе для устранения проблем с качеством данных. Они могут включать:

- Взвешивание для учета выборки или пропущенных записей
- Коррекция или удаление недействительных значений
- Подстановка пропущенных значений

211. Однако следует обратить внимание на то, как эти методы могут быть проблематичными в лонгитюдном интегрированном источнике данных.

212. В случае условной подстановки пропущенных значений, как описано в разделе 3.2, условное вычисление по файлам поперечного сечения может быть корректным, но объединение этих результатов в продольном направлении приведет к отклонениям для каждого условного значения. Например, если лонгитюдная база данных будет получена путем объединения двух циклов из источника данных, и в обоих циклах будут вычислены случайные 5% записей. Это может привести к ошибке в расчете до 9,75%. Больше количество циклов приведет к увеличению этой ошибки. В файлы данных можно было бы добавить условные обозначения, чтобы помочь аналитикам снизить влияние этого риска.

213. Подстановка все же может использоваться, но риск создания ложных исходов следует тщательно учитывать. В случае условной подстановки характеристик в файле когорты риск может быть меньше, например.

214. Для ограничений, связанных с выборкой в файле когорты, следует учитывать сопутствующие веса выборки. Дополнительные шаги для учета ошибки интеграции потребуют более тщательного рассмотрения.

215. Одним из способов устранения ограничений охвата является объединение нескольких источников данных. Например, отдельные файлы могут охватывать только подгруппы населения или только определенные периоды миграции. Объединение с другими источниками, охватывающими дополнительные группы населения, может устранить это ограничение. Как всегда, следует уделять внимание любому процессу интеграции и неизбежному риску ошибок при интеграции.

В Испании Реестр населения (Padrón) является идеальным источником для получения лонгитюдной информации о мигрантах. Он отражает место, где человек был зарегистрирован в конкретный момент времени, и содержит некоторые конкретные переменные: пол, возраст/дата рождения, место рождения, гражданство, имя и фамилия, уровень образования или официальный идентификационный номер. Padrón

не только хранит местонахождение человека в конкретный момент времени, но и, как лонгитюдная запись, содержит все места, где человек был зарегистрирован с 1996 года.

Одним из ограничений Padrón является то, что он содержит информацию лишь по нескольким переменным. С другой стороны, стоит упомянуть, как проводились последние две Переписи населения в Испании:

- - Перепись 2001 года была сплошной/ традиционной, и использование административных записей было очень ограниченным (только Padrón)
- - Перепись 2011 года была комбинированной. Она основывалась не только на Padrón, но и на других административных записях, и крупное обследование было проведено на почти 10 процентах населения.

Планы относительно Переписи 2021 года заключаются в том, чтобы попытаться провести операцию только на основе административных записей, включая: Padrón, Социальное обеспечение, Налоговое агентство и Министерство образования.

Среди всех переменных, которые должна предоставить Перепись, переменные, связанные с мигрантами, включают: Год прибытия в Испанию/регион/муниципалитет/жилище, предыдущее место жительства, место жительства 1 и 10 лет назад.

Однако у Padrón есть ограничение: его начальная точка – 1996 год, поэтому информация за предыдущие годы отсутствует. Следовательно, стратегия или решение, которое будет принято, заключается в связывании информации Padrón с Переписью 2001 года (которая была сплошной), и, таким образом, информация станет доступной для всех лиц в сфере охвата.

Связывание данных между Переписью и Padrón производилось с использованием различных переменных. В первую очередь, с использованием уникальных ссылочных номеров. Затем, для несвязанных записей, с использованием вспомогательных переменных, таких как имя, фамилия и дата рождения. Наконец, для заполнения пустой или ошибочной информации будут использоваться процедуры вероятностной подстановки (из других похожих записей).

Важно иметь в виду, что переменные, связанные с миграцией, должны соответствовать определенным правилам (например, год прибытия в Испанию должен быть до или равен году прибытия в жилище), и если они не выполняются из источников данных, как упоминалось ранее, их выполнение будет принудительно обеспечено с помощью вероятностных подстановок.

В Канаде Лонгитюдная база данных по иммигрантам (IMDB) включает интеграцию файлов по иммиграции, начиная с 1952 года, с ежегодными налоговыми файлами. Эта интеграция данных не охватывает население, родившееся в Канаде. Чтобы иметь возможность устранить это ограничение, IMDB часто используется параллельно с Лонгитюдным административным банком данных (LAD) – 20-процентной выборкой всех налогоплательщиков в Канаде. LAD связывается с IMDB для идентификации тех налогоплательщиков, которые охвачены IMDB.

В преддверии Переписи 2021 года Статистическое управление Канады изучает возможность замены вопросов о годе иммиграции и иммиграционном статусе с использованием интегрированных административных записей. Однако административные записи доступны только с 1952 года. В подходе, аналогичном описанному для Испании, чтобы устранить это ограничение, Статистическое управление Канады изучает возможность объединения административных данных с ответами из предыдущих переписей, которые включали бы годы иммиграции до 1952 года.

В Венгрии установление правил редактирования и подстановка пропущенных значений являются важнейшими элементами процесса очистки данных. Основные проблемы качества данных связаны с пропущенными или противоречивыми датами миграционных событий (например, отсутствует информация об иммиграции/ возвращении между множественными эмиграциями одного и того же лица).

3.4.3 Отчет о качестве данных исходных файлов

216. Наконец, оценка качества исходных файлов должна быть хорошо документирована для будущих пользователей результирующей базы данных.

217. В качестве первого шага, если в базу данных были внесены какие-либо изменения, которые повлияли бы на охват или измерение после предварительной обработки (раздел [3.4.2]), следует повторить проведенную оценку качества (раздел 3.4.1).

218. Показатели качества данных, такие как предложенные в разделе 3.4.1, должны быть хорошо документированы до и после любой предварительной обработки качества данных. Например, если данные подставлены для корректировки пропущенных значений, распределения до и после подстановки следует сравнивать с источниками данных для сравнения (т.е. альтернативными источниками данных с аналогичными измерениями на схожих совокупностях, которые можно

использовать для целей сравнения). Это позволяет проанализировать, улучшили ли шаги обработки охват или качество измерения.

219. Уровни предварительной обработки качества данных (например, уровни подстановки) должны документироваться и табулироваться по соответствующим переменным (например, временной период и значения переменных). Например, если пол подставляется в ежегодном файле результатов, интерес могут представлять уровень подстановки по годам и по итоговому значению пола.
220. Наконец, отчет должен включать руководство по качеству и пригодности к использованию различных полей в файлах когорты и результатов. Это руководство будет поддерживать будущий результирующий анализ, информируя пользователей о сильных сторонах и ограничениях исходных файлов данных. Отчет более подробно объясняется на Этапе 6.

3.5 ЭТАП 3: ИНТЕГРАЦИЯ ДАННЫХ ДЛЯ ПРОДОЛЬНЫХ ДАННЫХ⁷

221. Этот этап охватывает собственно процесс интеграции данных проекта. В этом разделе будут изложены методы интеграции и представлен обзор источников ошибок интеграции. Приведены примеры из разных стран, включая подходы, используемые для ограничения ошибок интеграции.
222. Этот раздел не охватывает административные процессы, которые могут потребоваться перед началом любого проекта интеграции данных. Для защиты конфиденциальности национальные статистические управления имеют строгие правила, регулирующие проекты интеграции данных. Они выходят за рамки рассмотрения доступа к самим исходным файлам, включая переменные, необходимые для завершения интеграции данных. Эти процессы зависят от страны, но их следует учитывать перед любым проектом интеграции данных.
223. Для целей данного раздела «интеграция данных» относится к микроданным комбинации различных наборов данных с целью создания лонгитюдной базы данных. «Сопоставление записей» относится к процессу микроданных по объединению наборов данных и в основном используется для обозначения вероятностного связывания в отличие от детерминистического, совпадающего по одному уникальному идентификатору. Однако возможно проведение детерминистического связывания по имени, дате рождения, полу и географии перед вероятностным связыванием.
224. Конечным результатом этого этапа является наличие ключа связывания, соединяющего файлы когорты и исходов, вместе с дополнительной документацией о качестве интеграции данных.

3.5.1 Методы интеграции данных

«Принцип сопоставления записей заключается просто в сравнении и объединении записей из двух (или более) источников, которые, как полагают, относятся к одному и тому же объекту реального мира и чье (предположительно перекрывающееся) описание может быть сопоставлено таким образом, что они затем могут рассматриваться как одна запись» (Berti-Equille, 2007:112).

225. Интеграция данных может быть осуществлена с использованием двух различных подходов:

- Точное совпадение с использованием общего уникального индивидуального идентификатора (для регистров или других источников данных с общим идентификатором)
- Вероятностное сопоставление путем применения конкретного алгоритма и использования различных переменных (таких как имя, пол, дата рождения и т.д.)⁸

226. Однако предлагается, чтобы до этого этапа были проведены внутренние связывания исходных файлов для идентификации дубликатов (см. раздел 3.4.2).

Например, в Канаде административные записи Иммиграционного, беженского и гражданского ведомства Канады (IRCC) используют общие идентификаторы личности, которые позволяют связывать различные файлы с людьми. Что касается Лонгитюдной базы данных по иммигрантам (IMDB), которая объединяет связанные административные файлы данных по иммиграции и налогам⁹, ежегодные налоговые файлы могут быть связаны с людьми с течением времени через их номер социального страхования (SIN). Однако между иммиграционными и налоговыми записями нет общего уникального индивидуального

идентификатора. Следовательно, требуется вероятностное сопоставление записей. Перед этим проводится внутреннее связывание, как описано в 3.4.2, для идентификации дубликатов в исходных файлах.

Этот процесс сопоставления записей использует переменные, доступные в обоих источниках данных (например, имена, дата рождения и т.д.), и в настоящее время использует Социальную среду связывания данных (SDLE) для облегчения связывания. Поскольку характеристики мигрантов со временем меняются чаще, Статистическое управление Канады нашло полезным полагаться на лонгитюдные источники связывания. Это подразумевает источник, который может содержать изменения информации о человеке с течением времени, но связывает эти изменения через индивидуальный идентификатор. Использование такого файла позволяет устанавливать связи, даже если имена или адреса изменились. В случае IMDB исторически этим лонгитюдным источником связывания были налоговые данные, связанные с одним SIN (или идентифицированной группой SIN, принадлежащих одному и тому же лицу).

В Германии Центральный реестр иностранцев (AZR) – административный реестр, содержащий всех иностранцев, в настоящее время зарегистрированных в Германии (активные) или ранее зарегистрированных в Германии (неактивные) – имеет сквозной и единообразный персональный идентификатор (PI) с 2007 года. Этот PI представляет собой псевдо-нумеризованный ссылочный номер реестра.

Итальянский Национальный институт статистики (ISTAT) уже некоторое время работает с административными данными, применяя методы сопоставления записей для увеличения ценности информации с лонгитюдной точки зрения. Он внедрил новую систему SIM (Интегрированная система микроданных) с целью создания уникального статистического идентификационного кода. Данные, касающиеся видов на жительство, были интегрированы с данными о приобретении гражданства, реестром населения и реестром социального обеспечения для мониторинга процесса включения граждан не из ЕС.

--- --

7. Данный раздел основан на ЕЭК ООН (2019), который рекомендуется использовать в дополнение к настоящему Руководству.

Для базы данных видов на жительство до сих пор использовался налоговый код, но они заменяются на SIM. Что касается обследования интеграции второго поколения, оно было связано с Реестрами населения и Реестрами учащихся. Две базы данных интегрируются через детерминистическое сопоставление записей с использованием налогового кода.

В Испании НСУ (INE) отвечает за координацию единого национального реестра населения (Padrón), который представляет собой комбинацию реестров каждого муниципалитета в Испании. Каждый человек в реестре имеет идентификационный номер. Для большинства людей (включая легальных иностранцев в стране) это номер национального удостоверения личности; для тех, кто находится в нерегулярном статусе, сохраняется номер их паспорта.

Есть две ситуации, когда может иметь смысл проводить вероятностное связывание. Первая – когда Padrón включает информацию из других вспомогательных источников. Padrón содержит информацию только о нескольких переменных (дата рождения, место рождения, пол и гражданство). Следовательно, для включения большего количества переменных из разных источников необходимо вероятностное сопоставление записей. Вторая ситуация, когда необходимо вероятностное связывание, – когда член Padrón с идентификатором X покидает Испанию и возвращается через несколько лет. По возвращении ему будет присвоен другой идентификатор Padrón, например Y. Следовательно, для построения лонгитюдного реестра для всего населения сопоставление записей может подтвердить, что лицо с идентификатором X – это то же самое, что и лицо с идентификатором Y.

Что касается Швейцарии, наличие уникального идентификационного номера личности (PIN) – соответствующего номеру социального страхования – позволяет связывать Центральный реестр иностранцев (ZAR, впоследствии ZEMIS) и Статистику населения и домохозяйств (STATPOP), которые основаны на муниципальных регистрах населения для построения Швейцарской лонгитюдной демографической базы данных (SLDD). Этот 13-значный PIN (известный как NAVS13) является полностью анонимным и был введен в конце первого десятилетия XXI века. Лицам требуется NAVS13 для работы в Швейцарии, и он часто запрашивается при обращении в органы власти и административные учреждения для таких целей, как получение страховки, пособий или медицинской помощи. Поэтому даже нелегальные мигранты часто имеют такой номер. Номер присваивается каждому иностранцу, въезжающему в страну (либо через иммиграцию, либо по рождению). Для статистических целей этот идентификатор доступен для любого, кто проживал в Швейцарии в любое время с 31 декабря 2010 года. Различные реестры связывались или сравнивались с использованием двух различных подходов: 1) точное совпадение с использованием общего идентификатора (PIN) и 2) сопоставление путем применения конкретного алгоритма. Связывание обоих регистровых источников (ZAR и STATPOP) с помощью PIN было невозможно для бывших просителей убежища, получивших вид на жительство или разрешение на поселение до 2008 года. В тех случаях требовалось сопоставление на основе наличия либо немодифицируемых переменных (т.е. дата рождения, пол, гражданство на момент прибытия в Швейцарию), либо редко изменяемых переменных (т.е. место жительства, семейное положение) в обоих реестрах. Этот алгоритм учитывает две характеристики в частности: территориальные изменения и

изменение гражданства человека (например, югослав принимает сербско-черногорское гражданство, затем сербское гражданство) с течением времени.

3.5.1.1 Ошибки связывания

227. Валидация результатов всегда является важным шагом производства данных. Следует учитывать административные ошибки, такие как дубликаты, пробелы, записи разных дат и противоречивые события. В силу природы административных данных переменные, такие как имя и адрес, фиксируются в рамках административного процесса и поэтому могут регистрироваться с задержкой. Для мигрантов информация может меняться чаще, что приводит к более высокому риску ошибок связывания. В частности, на связывание мигрантов может непропорционально сильно влиять различные факторы, такие как:

- Принятие мигрантами имени, более привычного для новой страны проживания
- Смена фамилии мигрантами после замужества (браки могут быть причиной миграции)
- Изменение мигрантами контактной информации и адреса после прибытия в новую страну проживания
- Незнание мигрантами местного алфавита – приводит к различным вариантам написания имен
- Незнание мигрантами местного календаря – приводит к неопределенности даты рождения

8. Это также может включать детерминистическое или иерархическое детерминистическое сопоставление по таким переменным, как имя, пол, дата рождения и т.д.

9. IMDB объединяет иммиграционную информацию от IRCC (записи об иммиграции, временном проживании и гражданстве), данные о налогообложении от Агентства налоговых поступлений Канады и дату смерти из Базы данных о смертности Статистического управления Канады.

228. Ожидаемым источникам ошибок связывания, таким как перечисленные выше факторы, следует уделять особое внимание на протяжении всей оценки качества.

229. Ошибки связывания возникают при некорректной классификации пары записей. Ложные совпадения (также известные как ложноположительные) могут возникать, когда записи разных людей ошибочно связываются. Пропущенные совпадения (ложноотрицательные) могут происходить, когда записи одного и того же человека не связываются из-за ошибок отчетности (например, опечаток), изменений с течением времени (например, фамилии замужних женщин, адресов) или пропущенных значений. Ошибки связывания могут угрожать надежности результатов, основанных на анализе связанных административных данных. По возможности разрабатываются стратегии для устранения этих ошибок связывания. В любом случае ошибки связывания следует оценивать и хорошо документировать, чтобы информировать о пригодности использования результирующих данных.

230. Для оценки качества сопоставления записей рекомендуемые практики включают:

- Качественную оценку связанных и несвязанных пар записей для оценки уровня ложноотрицательных и ложноположительных результатов
- Количественную оценку уровней связывания. Уровни связывания должны рассчитываться с точки зрения исходных файлов (т.е. исходные файлы должны предоставлять знаменатели при расчете уровней связывания).

231. Что касается коэффициентов связи, то расчеты следует производить с точки зрения каждого исходного файла. То есть исходный файл содержит знаменатель, а коэффициент связи - это процент от этого знаменателя, который связан. Затем показатели необходимо интерпретировать через призму исходных файлов. Следует ожидать, что показатели связи приблизятся к 100 процентам только в том случае, если все данные в исходном файле находятся в области, которую можно связать. Более низкие показатели связи могут быть вызваны ошибкой связи, а также относительными различиями в присутствии отдельных лиц в других исходных файлах. Эти различия могут быть вызваны ошибкой в охвате или выборке или просто отсутствием отдельных лиц в определенных связанных исходных файлах (например, отсутствие связи между отдельными лицами и больничными данными может просто означать, что люди здоровы и, следовательно, не нуждаются в посещении больницы).

Например, в Канаде Лонгитюдная база данных по иммигрантам (IMDB) проходит ежегодное сопоставление записей для добавления новых когорт и новых результатов. Качество связывания оценивается с помощью различных шагов:

1. Качественная оценка связанных и несвязанных пар записей для оценки уровня ложноотрицательных и ложноположительных результатов
2. Общие уровни связывания между файлами результатов и когорты с Социальной средой связывания данных (SDLE), которая используется для облегчения сопоставления записей
3. Количественный анализ уровней связывания, рассчитанных с каждым из исходных файлов в знаменателе.

4. Сравнительный анализ между новыми и предыдущими результатами связывания

В частности, много времени уделяется уровням связывания с точки зрения файлов по иммиграции (файлы когорты) с учетом переменных, представляющих ключевой аналитический интерес, а также переменных, которые могут быть релевантны для ошибок связывания. К ним относятся, но не ограничиваются, год иммиграции или прибытия, год рождения, пол, категория допуска и страна рождения.

Процесс и результаты документируются в Техническом отчете по Лонгитюдной базе данных по иммигрантам (IMDB) (Evra & Prokopenko, 2019).

В Испании наиболее распространенные ошибки включают составные имена, которые по-разному отображаются в каждом источнике, смену фамилии, дату рождения, равную 1 января¹⁰, брата или сестру-близнеца с очень похожим именем, изменения гражданства и изменения идентификационного номера. Для решения этих проблем обычно проводится вероятностная связь между источниками, и связи выше определенного порога достоверности принимаются.

В Германии Федеральное статистическое управление не занимается напрямую ошибками связывания, потому что они не отвечают за реестры. Ежегодно они получают выписку из реестра от органа, ведущего реестр, а именно Федерального ведомства по миграции и беженцам. Это означает, что они получают персональный идентификатор (PI) людей, а также дополнительные переменные для статистического анализа. Последние могут быть как зависящими от времени (например, статус проживания и вид на жительство с датой выдачи), так и не зависящими от времени (например, дата рождения, дата первого въезда и т.д.). Несоответствие может произойти только в том случае, если не зависящие от времени переменные меняются со временем. Изменения гражданства имели место, но это не обязательно свидетельствует о несоответствии. Также происходили изменения дат рождения и дат первого въезда в результате неполных данных (например, отсутствующий день и/или месяц) в очень старых данных, которые были заполнены в более поздних записях. Это также не является однозначным доказательством несоответствия.

Несоответствия маловероятны, поскольку исходный ID создается как временная метка на день первой регистрации плюс уникальный счетчик, генерируемый базой данных. Возможно, что человек был зарегистрирован как новый случай, даже если он уже был зарегистрирован ранее. Однако это означало бы, что у человека был новый паспорт, так что номер паспорта и предыдущие записи о визе или виде на жительство были больше не видны. В обычных обстоятельствах такие множественные регистрации происходят не часто, поскольку эмигрировавший иностранец остается «неактивной» записью в реестре в течение дополнительных 12 лет после его/ее эмиграции. Однако во время увеличения числа прибывающих беженцев в 2015-2016 годах, многие без паспорта, власти были перегружены, и работа была делегирована неподготовленному персоналу (например, вооруженным силам), что привело ко многим множественным регистрациям. Офис реестра недавно начал выявлять двойные/множественные регистрации путем проверки имен и даты рождения. На более позднем этапе они будут включать данные отпечатков пальцев для этой цели.

Что касается Швейцарии, были проведены различные тесты для проверки данных, в частности контроль ошибочных связей, правдоподобие последовательностей событий и их полнота. После завершения этих тестов вторая фаза включала идентификацию источников ошибок путем консультаций с Швейцарским федеральным статистическим управлением (SFSO) или проведения проверок правдоподобия с контрольными переменными. Третья фаза состоит в принятии решения о том, как исправить данные. Дублирующие записи были удалены. В некоторых случаях записи, касающиеся двух разных лиц, были связаны вместе. Сложность в таких случаях заключается в определении, являются ли записи дублированными (т.е. один человек записан дважды) или номер PIN был некорректным (т.е. два человека имеют один и тот же номер PIN). Такие случаи передавались в FSO, который имеет доступ к дополнительной информации (например, имя человека) и мог исправить данные. Когда существует несоответствие между информацией из разных источников, приоритет отдается информации, доступной в файле «наличного населения» (картотеке).

Оценка полноты траекторий основана на различных правилах относительно начала и конца траектории. Начало траектории определяется иммиграционным движением или рождением. Для иностранцев, уже проживавших в Швейцарии до начала наблюдения (1998), записи «наличного населения» по состоянию на 31 декабря 1997 года представляют начало траектории. Конец траектории определяется как выбытие, смерть или натурализация. Для людей, проживающих в Швейцарии в конце периода наблюдения (2017), записи «наличного населения» по состоянию на 31 декабря 2017 года представляют текущий конец

траектории. Присутствие только в один момент времени без какой-либо информации о въезде или выбытии не считается траекторией и поэтому было удалено.

Что касается Италии, базы данных связываются кодом Интегрированной системы микроданных (SIM). Код SIM является результатом алгоритма, который применяет уникальный идентификационный номер к каждой индивидуальной записи.

Одна из проблем в Италии связана с поддержанием хороших связей с течением времени. Итальянский закон о конфиденциальности предусматривает серьезные ограничения в отношении связывания и использования идентифицирующей информации о лицах. Даже в этом случае идентификаторы с большей вероятностью будут более низкого качества из-за проблем, связанных с использованием разных алфавитов. Кроме того, налоговый код, используемый в качестве ключа связывания для видов на жительство, отсутствует для некоторых лиц, особенно для мигрантов, прибывших в Италию в качестве просителей убежища.

В дополнение к связыванию баз данных процесс анонимизирует содержащуюся информацию и удаляет налоговый код. Этот процесс проводится ISTAT (Итальянским национальным институтом статистики). ISTAT еще не полностью доверяет результату этого алгоритма, и результат связывания проверяется с помощью детерминистического и вероятностного сопоставления записей. В некоторых случаях ошибки обнаруживаются при связывании с идентификатором SIM. Когда это происходит, ISTAT информирует офис, занимающийся SIM, который затем проверяет и фиксирует эти ошибки, обновляя таблицу перекодировки, которая позволяет отслеживать изменения и исправлять ошибки.

3.5.2 Документирование ошибок интеграции

232. Документирование процесса интеграции должно включать обзор методологии интеграции, а также подробные результаты по качеству интеграции.

233. Ошибки интеграции, включая оценки ложноположительных и ложноотрицательных связей, должны быть хорошо документированы. Уровни ошибок должны быть табулированы по временным периодам и ключевым переменным интереса. Любые подгруппы интереса, которые, по-видимому, непропорционально сильно затрагиваются ошибками интеграции, должны быть хорошо документированы.

234. Наконец, пользователям следует предоставить руководство по пригодности использования результатов. В частности, следует дать указания, если определенные подгруппы или временные периоды могут иметь более низкое качество интеграции. Например, в Канаде иммиграционные записи с 1961 по 1972 год часто не включают полную дату рождения, что приводит к более низким уровням связывания для этого периода иммиграции при связывании с другими файлами, такими как Перепись 2016 года. Эти результаты могут либо привести к исключению определенных групп из результатов, либо потребовать новых методов для корректировки на ошибку интеграции.

235. Следующие 2 этапа решают нерешенные проблемы, созданные в процессе интеграции. Этап 4 описывает сложности, связанные с поддержанием согласованного уникального идентификатора записи в лонгитюдном разрезе (особенно после повторных интеграций), в то время как Этап 5 описывает методы, необходимые для подготовки окончательной интегрированной базы данных, включая корректировки для учета ошибок интеграции.

3.6 ЭТАП 4: НАЗНАЧЕНИЕ ЛОНГИТЮДНЫХ ИНДИВИДУАЛЬНЫХ ИДЕНТИФИКАТОРОВ

236. После Этапа 3 устанавливается ключ связывания между файлами когорты и исходов после интеграции данных. На этом этапе ключ рассматривается в лонгитюдном контексте. Результатом этого этапа является преобразование ключей связывания (устраняются дубликаты и другие лонгитюдные проблемы) в анонимизированный лонгитюдный индивидуальный идентификатор, готовый к использованию для создания аналитической базы данных. Этот новый идентификатор затем следует присоединить к файлам когорты и результатов (часто заменяя любые существующие идентификаторы) для облегчения анализа лонгитюдной базы данных. В этом разделе будут рассмотрены общие проблемы и подходы, используемые разными странами для решения различных задач.

237. Существует несколько преимуществ присвоения нового обезличенного индивидуального идентификатора статистическими управлениями. Это включает добавление уровня

конфиденциальности к результирующей базе данных. Анонимизированный идентификатор ничего не будет значить для пользователей базы данных, и его создание устраняет необходимость наличия каких-либо персональных идентификаторов в интегрированных файлах данных. С другой стороны, официальные идентификаторы, сгенерированные для административных целей, могут быть использованы для идентификации лиц.

238. Критическим элементом любой лонгитюдной базы данных является возможность связывать исходы с течением времени с одной и той же единицей (человеком, семьей и т.д.). Это облегчается за счет использования уникальных идентификаторов записей для единиц исследования.
239. Когда лонгитюдная база данных разрабатывается с использованием интегрированных источников данных, необходимо установить согласованные идентификаторы единиц. В некоторых случаях, и как это часто бывает для баз данных, созданных из регистров населения, общий уникальный идентификатор уже существует (т.е. один и тот же уникальный идентификатор существует во всех источниках, которые интегрируются). В других случаях интегрируемые отдельные файлы могут иметь независимые уникальные идентификаторы, которые необходимо будет объединить.
240. Дополнительная сложность возникает для лонгитюдных баз данных, где уникальный идентификатор корректируется с течением времени для учета обнаруженных дубликатов (т.е. нескольких идентификаторов для одной и той же единицы) или разделений (т.е. нескольких единиц для одного и того же идентификатора). Хотя эта ситуация будет возникать для людей только при административной ошибке (например, ошибочное создание нового идентификатора для кого-то, у кого он уже есть), семейные единицы более сложны, так как они могут объединяться и разделяться с течением времени.

3.6.1 Интеграция данных с общими уникальными идентификаторами

241. Если лонгитюдная база данных разрабатывается с использованием интегрированных источников данных с общими уникальными идентификаторами, возможно, нет необходимости создавать новый идентификатор. Общий уникальный идентификатор может быть повторно использован или преобразован (например, анонимизирован) для создания нового уникального идентификатора записи для целей лонгитюдной базы данных. Рабочим предположением в этом сценарии будет то, что этот уникальный идентификатор записи существует и является уникальным для каждой единицы в сфере охвата.
242. Однако транзитный характер мигрантов может влиять на их идентификаторы в национальных базах данных. Например, если человек мигрирует в страну два или более раз отдельно (например, циркулярная миграция), ему может быть выдано несколько идентификаторов. Может существовать административный процесс для обнаружения таких повторных въездов и присвоения одного единственного идентификатора лицу, но этот процесс может быть не исчерпывающим. Внутреннее связывание записей до Этапа 3 могло бы идентифицировать такие случаи. Это упоминалось как рекомендуемая практика в разделе 3.4.2.
243. Если будет обнаружено, что люди имеют несколько индивидуальных идентификаторов в базе данных, этот идентификатор следует преобразовать, чтобы иметь один идентификатор на человека. В противном случае будет разрыв между различными событиями для одного и того же человека.

В Швейцарии Швейцарская лонгитюдная демографическая база данных (SLDD) строится в 2 этапа: Создается демографическая основа, в которой связываются Центральный реестр иностранцев (ZAR) и Статистика населения и домохозяйств (STATPOP)
Другие наборы данных, такие как Структурное обследование (RS) и данные реестра социального страхования, включаются в эту демографическую основу.
Эти связывания выполняются с использованием уникального идентификационного номера личности (PIN) – соответствующего номеру социального страхования – который доступен в каждой из баз данных.
Этот 13-значный PIN (известный как NAVS13) является полностью анонимным и был введен в 2010 году. NAVS13 необходим для возможности работать в Швейцарии и часто запрашивается при обращении в органы власти и административные учреждения для таких целей, как получение страховки, пособий или медицинской помощи. Поэтому даже нелегальные мигранты часто имеют такой номер. Номер присваивается каждому лицу, входящему в население (либо через иммиграцию, либо по рождению). Для статистических целей этот идентификатор доступен для любого, кто проживал в Швейцарии в любое время с 31 декабря 2010 года.

В Испании Padrón (Реестр населения) создается и управляется национальным статистическим управлением, INE. Хотя существует столько реестров, сколько муниципалитетов в Испании, с 1996 года действует закон, который обеспечивает интеграцию этих муниципальных списков в единую национальную базу данных. Также существуют юридические процедуры, которые поддерживают эту базу данных и муниципальные файлы взаимосвязанными и обновляемыми ежемесячно.

INE отвечает за координацию единого национального реестра населения. Он проверяет изменения, произведенные в каждом муниципалитете, чтобы избежать дублирования, и обновляет его, чтобы включить смерти, рождения и приобретение испанского гражданства, которые INE получает ежемесячно из Гражданского реестра. Каждый человек в реестре имеет идентификационный номер. Для большинства людей (включая легальных мигрантов в стране) это номер национального удостоверения личности; для тех, кто находится в нерегулярном статусе, сохраняется номер их паспорта.

INE постоянно получает информацию о регистрациях и снятиях с регистрации от каждого муниципалитета, так что, чтобы избежать дубликатов, как только поступает уведомление о новой регистрации человека в муниципалитете или консульском отделе, этот человек удаляется в муниципалитете происхождения. Когда INE обнаруживает дубликат, затронутые муниципалитеты информируются о обнаруженном дублировании регистраций и принимаются меры для выяснения, какая запись является правильной.

Чтобы избежать использования одного и того же внешнего идентификатора (национальный документ, вид на жительство или паспорт) для нескольких людей, INE генерирует свой собственный внутренний идентификатор для каждого человека в Padrón. Этот идентификатор остается неизменным, пока человек остается в стране. Если человек покидает страну, потому что он/она переводится за границу, а затем возвращается в Испанию, будет сгенерирован другой внутренний идентификатор, даже если это тот же человек. Внутреннее связывание записей, как описано в 3.4.2, может быть использовано для восстановления его/ее истории.

В Германии Центральный реестр иностранцев также строится с использованием связанных административных записей на основе псевдонимизированного идентификатора личности.

В Италии ISTAT внедрил новую систему, SIM (Интегрированная система микроданных), с целью производства уникального статистического идентификационного кода. Для видов на жительство до сих пор использовались налоговые коды, но они заменяются на SIM. Идентификационный код SIM полностью управляется и генерируется ISTAT с помощью особого алгоритма и использовался в качестве замены процесса анонимизации, который требуется по закону для связывания и интеграции административных баз данных в одну уникальную Интегрированную систему микроданных (SIM).

Дубликаты по коду SIM перепроверяются, поскольку как ложноотрицательные, так и ложноположительные результаты были выявлены при связывании микроданных с использованием детерминистических и вероятностных методологий. В некоторых случаях дубликаты по кодам SIM на самом деле определяются как разные люди. Эти случаи возвращаются в офис, который управляет алгоритмом. Эти коммуникации между производителями и пользователями кода имеют важнейшее значение для улучшения процесса идентификации и избежания ошибок в будущем, ошибок, которые затрагивают в основном определенные подгруппы населения, такие как иностранцы.

В Канаде данные Иммиграционного, беженского и гражданского ведомства Канады (IRCC) интегрируются с использованием общих идентификаторов личности, которые позволяют связывать различные файлы с людьми. Эти идентификаторы выдаются в административном порядке и предназначены быть единственными и уникальными для иммигрантов и временных резидентов в Канаде. При разработке Лонгитюдной базы данных по иммигрантам (IMDB) Статистическое управление Канады выявило дубликаты (случаи, когда у одного человека было несколько идентификаторов личности IRCC) с помощью внутреннего связывания записей. Хотя они редки для иммигрантов (рабочее предположение состоит в том, что они не возникают), они более распространены среди временных резидентов. Статистическое управление Канады устраняет эти дубликаты, только для временных резидентов, так как предполагается, что дублирующих записей для иммигрантов нет, при замене идентификатора личности IRCC на новый анонимизированный идентификатор IMDB.

В Канаде налоговые данные о физических лицах могут быть со временем привязаны к их номеру социального страхования (SIN). Налоговое управление Канады (CRA) сообщило Статистическому управлению Канады о случаях, когда у физических лиц было несколько SIN. Это особенно распространено среди лиц, которые переходят от статуса временного жителя к статусу постоянного жителя, поскольку им выдается временный паспорт, прежде чем они становятся постоянными жителями. Статистическое управление Канады также провело внутреннюю проверку записей для выявления дополнительных дубликатов (или дополнительных случаев, когда у человека было несколько грехов). Эти взаимосвязи и вытекающие из них открытия и корректировки проводятся ежегодно.

В Казахстане также используются общие уникальные идентификаторы для избежания дубликатов. Три статистические единицы производятся на основе юридической единицы: человек, семья и домохозяйство. В первом случае Идентификационный номер личности (PIN) – соответствующий официальному

удостоверению резидентов Республики Казахстан – генерируется Министерством юстиции для Государственной базы данных физических лиц, одного из основных источников информации для Статистического регистра населения. Этот PIN присваивается человеку в Государственной базе данных физических лиц и передается в Статистический регистр населения. Для решения некоторых несоответствий между разными источниками информации для одного и того же человека была определена система приоритета между источниками.

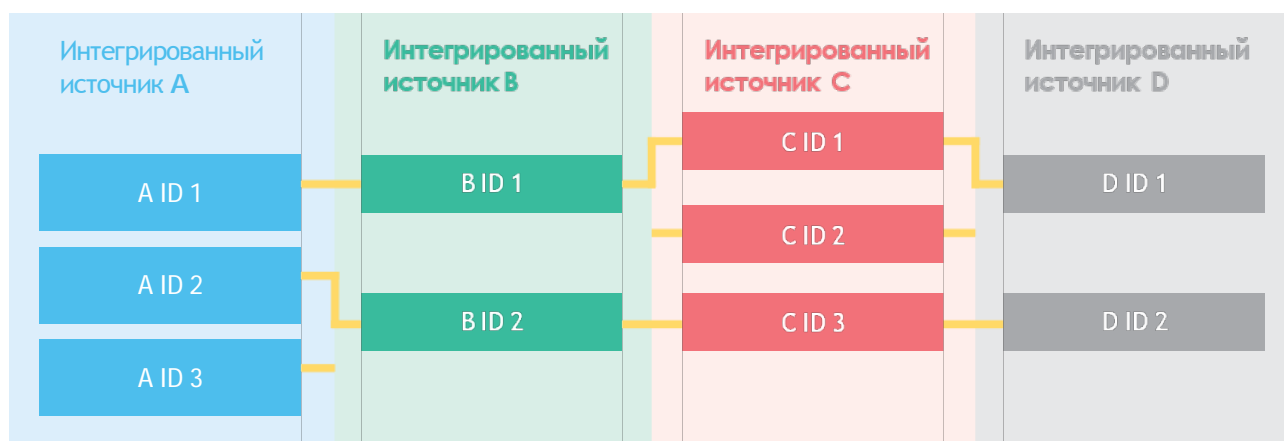
В Турции система регистрации населения на основе адресов (ABPRS) является уникальной системой, с помощью которой распространяется ежегодная демографическая статистика. В рамках этой системы на основе ABPRS и других дополнительных административных источников формируется ежегодная информация о численности и потоках международной миграции в разбивке по гражданству и стране рождения. Постоянное сохранение личных идентификационных данных в системе ABPRS позволило осуществлять долгосрочное наблюдение за отдельными лицами. Однако изменение идентификационного номера после получения гражданства может повлиять на отдельных лиц (т.е. потеря данных о натурализованном лице).

3.6.2 Интеграция данных без общих уникальных идентификаторов

244. Существуют страны, в которых отсутствует регистр населения или возможность связывать файлы с помощью общего уникального идентификатора. Для лонгитюдных наборов данных, полученных из интегрированных источников без каких-либо общих уникальных идентификаторов, необходимо сгенерировать уникальный идентификатор. В некоторых случаях, если один из интегрированных источников служит основой для лонгитюдной базы данных (т.е. там, где будут определяться когорты), можно использовать уникальный идентификатор этой основы. Альтернативно, может быть использована или преобразована (например, анонимизирована) комбинация идентификаторов из разных источников.

245. В этих условиях сгенерированные уникальные идентификаторы должны учитывать дубликаты (т.е. несколько идентификаторов для одной и той же единицы) или разделения (т.е. несколько единиц для одного и того же идентификатора), некоторые из которых могут быть выявлены в самом процессе интеграции. В лонгитюдных базах данных, которые со временем обновляются за счет новых связей, могут быть обнаружены дополнительные дубликаты и разделения. Следует установить процедуру для устранения будущих обнаруженных несоответствий.

РИСУНОК 9: РАЗЛИЧИЯ В ИДЕНТИФИКАТОРАХ МЕЖДУ ИНТЕГРИРОВАННЫМИ ИСТОЧНИКАМИ



Источник: Статистическое управление Канады.

246. Для выявления дубликатов в исходных файлах перед Этапом 3 следует провести внутреннее связывание записей. Это упоминалось как передовая практика в разделе 3.4.2.

247. Например, на Рисунке 9 показана интеграция 4 источников данных с разными идентификаторами записей. Эти источники могут быть связаны во времени для создания лонгитюдного источника данных (где источник А представляет самые ранние события, а источник D — самые поздние). Процесс интеграции выявил несоответствия между идентификаторами. Например, согласно связи

между источником А и источником В, А ID 2 и А ID 3 — это один и тот же человек (В ID 2). Необходимо принять решение о том, как устранить эти несоответствия.

В Канаде Лонгитюдная база данных по иммигрантам (IMDB) включает интеграцию налоговых данных с номерами социального страхования (SIN) в качестве уникального идентификатора и административных данных по иммиграции с идентификаторами лиц IRCC. В IMDB предполагается, что идентификаторы лиц IRCC являются единичными и уникальными (после учета дублирующихся идентификаторов, связанных с временными резидентами), и они заменяются анонимизированным идентификатором лица в IMDB. Затем, в процессе связывания записей, обнаруживаются новые дубликаты и разделения для SIN. В результате этих обнаружений SIN или идентифицированная группа SIN заменяется идентификатором лица в IMDB.

В Италии ISTAT разработал Anvis, систему демографического учета в разрезе физических лиц, с целью обеспечения согласованности всех индивидуальных потоков, которые пополняют и/или изменяют численность резидентного населения. Посредством связывания данных можно отслеживать миграционную историю каждого человека и связывать все события, относящиеся к одному и тому же лицу.

3.6.3 Идентификаторы семей

248. В отличие от отдельных лиц, семьи действительно могут объединяться и разделяться с течением времени; это добавляет новый уровень сложности к идентификаторам семей для целей обработки лонгитюдных данных.

В Канаде IMDB идентифицирует семьи как в иммиграционных записях, так и в налоговых записях. В иммиграционных записях семьи идентифицируются как те, кто был принят в рамках одного и того же разрешения на постоянное проживание. В некоторых случаях (например, воссоединение семьи) не все члены семьи будут иметь одно и то же разрешение. Эти идентификаторы семей не меняются со временем, так как соответствуют одному разрешению. Однако в налоговых записях семейные единицы измеряются ежегодно (на основе супругов или родителей и детей, проживающих вместе в одном домохозяйстве) и могут меняться каждый год. IMDB преобразует идентификатор одного из взрослых иммигрантов в налоговой семье в идентификатор семьи для всех членов семьи. Это может приводить к изменению этих идентификаторов каждый год по мере изменения состава семей. Фактически, даже в случаях, когда идентификатор семьи остается неизменным из года в год, состав семьи может меняться (например, дети съезжают из домохозяйства).

В Казахстане на основе юридической единицы производятся три статистические единицы: человек, семья и домохозяйство. Уникальный 12-значный код автоматически присваивается Статистическим регистром населения каждой семье, и еще один код предоставляется для идентификации каждого домохозяйства. Для решения некоторых несоответствий между различными источниками информации об одном и том же лице была установлена система приоритетов между источниками.

249. Эти сложности затрудняют проведение лонгитюдного анализа семей. В главе 4.1 кратко рассматриваются лонгитюдные показатели, связанные с семьей.

3.7 ЭТАП 5: СОЗДАНИЕ ОКОНЧАТЕЛЬНОЙ БАЗЫ ДАННЫХ

250. После Этапа 4 уникальный идентификатор индивида присваивается файлам Когорты и Результатов. На этом этапе файлы Когорты и Результатов дополнительно обрабатываются для создания окончательной базы данных. Это может включать настройку структуры базы данных, отбор записей, релевантных для базы данных, гармонизацию содержания и стандартизацию названий переменных по файлам и референсным периодам.

251. К концу этого этапа лонгитюдная база данных будет готова к заключительному этапу распространения.

3.7.1 Протоколы конфиденциальности

252. После добавления уникального идентификатора индивида, для защиты конфиденциальности, все персональные идентификаторы должны быть удалены из файлов Когорты и Результатов.

253. Кроме того, на этом этапе следует рассмотреть уровень детализации переменных. Переменные должны быть настолько детализированными, насколько это необходимо для статистических целей. Любые идентифицирующие переменные, такие как даты рождения, адреса и т.д., должны быть заменены более агрегированными переменными, такими как возраст или географический район проживания. Все переменные должны быть рассмотрены в этом контексте, обеспечивая соответствующую аналитическую ценность при защите конфиденциальности индивидуальных записей.

Например, в Лонгитюдной базе данных по иммигрантам (IMDB) в Канаде переменные с датами (например, дата приема) заменяются на годы или годы и месяцы в зависимости от аналитической потребности.

254. Наконец, должны быть установлены требования к доступу, чтобы обеспечить доступ к базе данных только для тех, кто имеет соответствующее разрешение.

3.7.2 Структура базы данных

255. Существует множество вариантов структуры лонгитюдной базы данных.

256. Если количество результатов относительно невелико, может быть предпочтительнее создать единый плоский файл, содержащий переменные Когорты и Результаты. В этом сценарии файл может быть широким (где лонгитюдные результаты обозначаются столбцами или переменными, специфичными для каждого результата / референсного периода) или длинным (где каждый индивид имеет новую строку для каждого результата / референсного периода).

257. Например, базовая база данных, объединяющая детали иммиграции (год иммиграции) с ежегодными лонгитюдными данными о занятости, может быть представлена как широкий файл:

ID номер лица	Год иммиграции	Пол	Принят на работу в 2009 году	Принят на работу в 2010 году	Принят на работу в 2011 году	Принят на работу в 2012 году
1	2009	Мужской	0	1	1	
2	2011	Женский			1	1
3	2009	Женский	0	0	1	1
4	2010	Мужской		0	0	0

Или как длинный файл:

Отчетный год	ID номер лица	Год иммиграции	Пол	Принят на работу
2009	1	2009	Мужской	0
2010	1	2009	Мужской	1
2011	1	2009	Мужской	1
2011	2	2011	Женский	1
2012	2	2011	Женский	1
2009	3	2009	Женский	0
2010	3	2009	Женский	0
2011	3	2009	Женский	1
2012	3	2009	Женский	1
2010	4	2010	Мужской	0
2011	4	2010	Мужской	0
2012	4	2010	Мужской	0

258. Однако часто количество результатов существенно, и может потребоваться другой подход. Один из вариантов — поддерживать реляционную базу данных, связанную с помощью индивидуального идентификатора. Файлы Когорты и Результаты будут храниться отдельно, но объединяться в зависимости от проводимого анализа.

В Канаде Лонгитюдная база данных по иммигрантам (IMDB) имеет большой файл Когорты, состоящий из множества переменных, связанных с характеристиками иммигрантов при приеме в Канаду, а также некоторыми демографическими данными. Файлы Результаты — это ежегодные налоговые файлы с сотнями различных показателей, охватывающих такие темы, как доход, район проживания, семья и

налоговые льготы. Файлы Результатов сами по себе разделены, так как существует один файл на каждый налоговый референсный год.

Этот формат реляционной базы данных также гибок для новых дополнений. Недавно к базе данных IMDB были подключены данные о разрешениях на временное проживание (которые сами по себе представляют собой набор данных о лицах, получающих несколько разрешений с течением времени) с использованием идентификатора на уровне личности. Вместо того чтобы интегрировать эту информацию в файл когорты (что значительно снизило бы аналитический потенциал), она была просто добавлена в качестве отдельной новой реляционной части базы данных.

На рисунке ниже показана текущая структура IMDB. Файлы постоянных резидентов, файлы гражданства, записи об услугах по обустройству и разрешения временных резидентов (включая лица, ищущие убежища) интегрируются для создания связанных файлов:

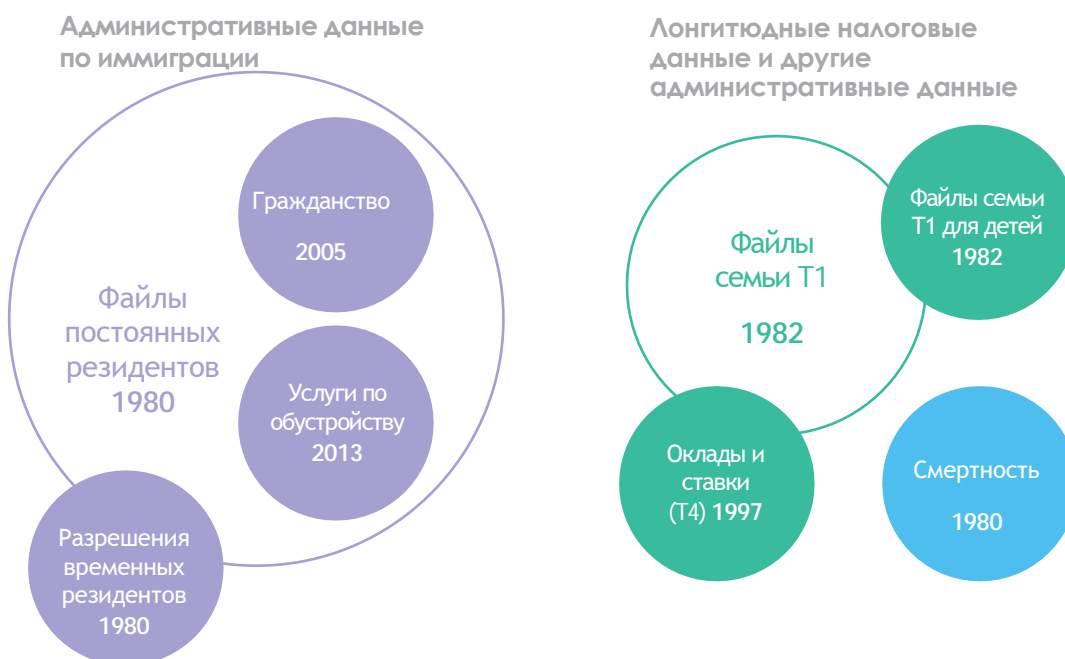
1. Файлы физических лиц (единица анализа: человек)
 - Содержит базовую демографическую информацию
 - Содержит детали (если применимо), связанные с условиями получения статуса постоянного резидента
 - Содержит информацию о гражданстве (если применимо)
 - Содержит суммарные переменные опыта временного проживания в Канаде (если применимо)
 - Содержит суммарные переменные использования услуг по обустройству в Канаде (если применимо)
 - Содержит дату смерти (если применимо) из файла результатов по смертности
2. Разрешения временных резидентов (единица анализа: разрешение или запрос на убежище)
 - Содержит детали всех выданных временных разрешений и запросов на убежище
 - Указаны даты, связанные с разрешениями, что позволяет проводить анализ в 'реальном времени'
3. Услуги по обустройству (единица анализа: услуга)
 - Содержит детали полученных услуг по обустройству

Для каждого отчетного года, начиная с 1982 года, существуют отдельные файлы результатов. Также существуют отдельные файлы для основных налоговых файлов (семейные файлы T1), налоговые файлы для детей и файлы отчетов о вознаграждениях по заработной плате.

259. На этом этапе необходимо учитывать периодичность Результатов. Может быть проще создавать периодические сводные файлы вместо перечисления результатов в реальном времени. Кроме того, могут существовать различия в периодичности между файлами, которые можно устранить путем создания таких сводных файлов.

260. Однако эти решения должны быть взвешены с учетом первоначальных целей создания базы данных, изложенных на Этапе 1.

РИСУНОК 10: СТРУКТУРА ЛОНГИТЮДНОЙ БАЗЫ ДАННЫХ ПО ИММИГРАНТАМ КАНАДЫ



3.7.3 Отбор записей

261. После определения структуры базы данных может возникнуть необходимость ограничить файлы Когорты или Результаты (или оба) подмножеством записей, релевантных целям базы данных. Это может означать исключение определенных групп населения из-за их характеристик в файлах Когорты или Результаты или ограничение базы данных только теми записями, которые были связаны. Это решение должно быть мотивировано статистическим дизайном базы данных, изложенным на Этапе 1. Однако некоторые решения могут диктоваться практическими соображениями, такими как доступность других источников данных по сравнению с техническими требованиями к хранению большой базы данных.

Например, Лонгитюдная база данных по иммигрантам (IMDB) в Канаде сохраняет налоговые записи только для тех, кто связан с иммиграционной записью. Однако все иммигранты и временные резиденты сохраняются в базе данных, независимо от того, были ли они связаны с налоговой записью. Это обусловлено тем, что существуют другие источники данных для сравнения налоговых результатов для иммигрантского и неиммигрантского населения в Канаде, а также тем, что сохранение дополнительных налоговых записей повлияло бы на удобство использования базы данных. IMDB особенно поддерживается Лонгитюдным административным банком данных (LAD), который содержит 20-процентную выборку лонгитюдных налоговых файлов Канады с флагом, идентифицирующим тех, кто связан с IMDB.

262. Другие решения, возможно, все же потребуются принять в случае оставшихся дубликатов. В частности, если один человек связан с несколькими записями результатов в одном и том же референсном периоде, необходимо решить, как поступать в таких ситуациях.

263. Наконец, это этап, на котором к полученной базе данных можно применить любое окончательное взвешивание или калибровку. Конкретные методы взвешивания и калибровки не рассматриваются в данном Руководстве. Однако следует обратить внимание на то, как эти методы повлияют на предполагаемый лонгитюдный анализ. В частности, в контексте статистики миграции методы борьбы с "исчезновением" отдельных лиц из файлов результатов должны учитывать, является ли это результатом отсутствия результата или выбытия из миграционного потока (эмиграции).

264. Все эти решения должны быть документированы и изложены в отчете. Это должно включать анализ возможных последствий принятых решений или рекомендации по использованию других источников данных для поддержки сравнительного анализа.

3.7.4 Гармонизированное содержание и стандартные названия переменных

265. На этом этапе важно вспомнить некоторую информацию, полученную на Этапах 1 и 2, при оценке исходных файлов на предмет их способности удовлетворять потребностям статистического дизайна.

266. В частности, необходимо изучить случаи, когда понятия измеряются неодинаково в разных файлах или в разные референсные периоды. По возможности, соответствующие переменные следует гармонизировать для обеспечения согласованности. Если это невозможно, различия между понятиями должны быть надлежащим образом задокументированы. Следует рассмотреть все переменные с общими (или относительно схожими) понятиями между файлами или референсными периодами.

267. Гармонизация переменных во времени значительно упростит лонгитюдный анализ, поскольку понятия будут определяться одинаковым образом. Если этого не сделать, наблюдаемые со временем различия могут быть вызваны изменением классификаций, а не реальными изменениями.

268. Гармонизация может включать использование таблиц соответствия для корректировки различий в классификации или преобразование переменных между файлами или референсными периодами для отражения одного и того же показателя.

269. Еще одним важным шагом является определение соглашения об именах для тех переменных, которые сопоставимы между референсными периодами. Для облегчения будущего анализа рекомендуется

сохранять точно такое же имя переменной для всех референсных периодов. Для дальнейшего упрощения будущего анализа референсный период также может быть включен в имя переменной.

270. Например, в приведенном выше примере с общим файлом в качестве переменной результата можно было бы использовать ГОД, где ГОД является базовым периодом.

271. Наконец, это этап, на котором должна быть выполнена любая окончательная обработка переменных. Это может включать в себя условное вычисление, не проведенное на Этапе 2.

Статистическое управление Канады в сотрудничестве с Министерством по делам иммиграции, беженцев и гражданства Канады (IRCC) согласовывает переменные с течением времени и данные по иммиграции и временному проживанию. Это включало обновление баз данных с учетом изменений в системах, используемых IRCC для сбора административных данных. В большинстве случаев это просто изменения в наборах кодов.

Для создания IMDB и Лонгитюдного банка административных данных (LAD), в которых используется тот же подход, Статистическое управление Канады сначала преобразует необработанные налоговые данные в T1FF (семейные файлы T1), которые содержат дополнительную информацию о составе семьи, а затем преобразует эти T1FF в лонгитюдные файлы, которые имеют согласованные переменные с течением времени.

Все налоговые переменные имеют стандартный протокол присвоения имен (например, T4_I2005 предоставляет доход от трудовой деятельности T4 на индивидуальном уровне или уровне "Г" – данные также доступны на семейном уровне – и за базовый 2005 год). Такой подход позволяет легко извлекать одну и ту же переменную из нескольких лет в любой статистической программе. Словари данных содержат информацию о том, какие отчетные годы содержат определенные переменные.

В Венгрии административные данные из Национального фонда медицинского страхования (NHIF) используются для получения данных исключительно об эмиграции и возвратной миграции венгерских граждан. Следует, однако, отметить, что в системе регистрации NHIF не требуется минимальная продолжительность пребывания для миграционных событий. Поэтому в этом отношении необходимы статистические корректировки для соответствия статистическому определению миграции. Например, продолжительности пребывания за границей (или в Венгрии между возвращением и повторной миграцией), которые составляют менее одного года, будут исключены из дальнейших этапов производства данных.

3.8 ЭТАП 6: РАСПРОСТРАНЕНИЕ РЕЗУЛЬТАТОВ

272. После Этапа 5 сама база данных готова к распространению. Однако есть несколько других элементов, которые должны быть установлены до распространения любых результатов и до проведения любого анализа.

273. На этом этапе будут выполнены заключительные элементы подготовки лонгитюдной базы данных к распространению. К ним относятся:

- Определение правил конфиденциальности для доступа и распространения
- Окончательная оценка качества базы данных
- Подготовка словаря(ей) данных
- Подготовка или обновление технического отчета / руководства пользователя / службы поддержки пользователей

274. На протяжении всех этапов этой главы качество данных неоднократно проверялось, но на разных стадиях. На этом этапе окончательная база данных будет оценена в последний раз на пригодность для использования, а также все соображения по качеству данных, отмеченные на предыдущих этапах, будут включены в официальную документацию — либо в словарь(и) данных (когда соображения качества данных относятся к отдельным переменным), либо в технический отчет / руководство пользователя. Технический отчет / руководство пользователя также должен содержать рекомендации о том, как следует или не следует использовать базу данных, исходя из выявленных различных ограничений качества данных.

275. После этого этапа база данных действительно готова к распространению. В следующей главе представлен обзор ключевых лонгитюдных показателей и передовых практик, связанных с распространением результатов из лонгитюдной базы данных, относящейся к мигрантам.

3.8.1 Определение правил конфиденциальности для доступа и распространения

276. Крайне важно установить строгие правила для защиты конфиденциальности результирующей базы данных. База данных, ориентированная на такие субпопуляции, как мигранты, уже рассматривает малые популяции. Лонгитюдный аспект базы данных предоставляет больше информации, чем простое перекрестное обследование. Кроме того, использование административных данных также требует особого рассмотрения.
277. В качестве передовой практики рекомендуется сотрудничать с методистами, работающими в этой конкретной области национальной статистики, и оставлять за ними окончательные решения по правилам конфиденциальности. Правила должны учитывать, как будут осуществляться доступ к данным и анализироваться агрегированные результаты. Риск раскрытия необходимо оценивать с учетом полезности данных.
278. Для агрегированных результатов могут действовать правила, учитывающие результаты, специфичные для уровня географии (например, правило доминирования дохода), и они могут подразумевать округление, подавление или возмущение данных. Даже содержание файлов микроданных должно быть тщательно проверено, чтобы гарантировать, что те, кто имеет к ним доступ, ограничены информацией, необходимой им для проведения их анализа, и ничем более. После Этапа 4 все идентифицирующие переменные могут быть удалены и заменены случайно сгенерированным индивидуальным идентификатором. Уровень детализации переменных также следует учитывать.

Например, в Канаде для Лонгитюдной базы данных по иммигрантам (IMDB) установлены правила подавления малых чисел или оценок, основанных на малых популяциях, округления других оценок и удаления выбросов. Эти методы изложены в Техническом отчете по IMDB (Evra & Prokopenko, 2019).

279. Наконец, доступ необходимо продолжать контролировать, чтобы ограничить базу данных только одобренными лицами, которые доказывают, что им требуется доступ для определенных статистических целей.

3.8.2 Окончательная оценка качества базы данных

280. На этом этапе следует оценить охват базы данных после интеграции данных. Следует учитывать пропущенные значения или наблюдения с течением времени, а также несоответствия между измеренными событиями (например, смерть до миграции). Если проводилась подстановка данных, следует учитывать процент подстановленных значений и наличие меток подстановки. Наконец, следует оценить пригодность базы данных для использования путем проведения предварительного анализа, ориентированного на ключевые цели, заявленные на Этапе 1.
281. Оценка качества должна ссылаться на рамки, определенные во введении этой главы. Как эти ошибки могли возникнуть в процессе? Были ли они идентифицированы и устранены?
282. Оценка должна использовать другие источники данных для целей сравнения или сопоставления. Эти источники данных не обязательно должны предоставлять такую же широту информации, как интегрированный источник, и они не обязательно должны быть золотым стандартом качества. Их роли могут варьироваться от исследования конкретной возможной ошибки до сравнения точечных оценок и сравнения общих тенденций с течением времени. Эти источники могут включать:
- Сами файлы Когорты и Результаты (до интеграции)
 - Они предоставляют ключевую информацию о характеристиках тех, кто охвачен (и не охвачен) интеграцией данных
 - Переписи и оценки численности населения
 - Обследования со схожим содержанием
 - Административные данные со схожим содержанием
 - Другие источники данных, включая «Большие данные»
283. Для оценки охвата вернитесь к источникам, использованным в анализе охвата на Этапе 2. Повторите этот анализ, но сосредоточьтесь только на связанной популяции. Временной элемент, добавленный файлами Результаты, предоставляет средства для установления того, что популяция Когорты входит

в сферу охвата для тех же референсных периодов. Это облегчает сравнение с другими источниками данных за тот же референсный период.

Например, в Канаде Лонгитюдную базу данных по иммигрантам (IMDB) можно сравнить с оценками переписи, ограничив сравнение связанным населением, подающим налоговые декларации, за тот же референсный год. Эти сравнения затем могут быть выполнены по различным характеристикам, общим для обоих источников.

284. Следует изучить закономерности в файлах Результатов, а также то, как они соотносятся с файлами Когорты. Отсутствуют ли люди в некоторых референсных периодах? Есть ли признаки исчезновения отдельных лиц? Появляются ли результаты в бессмысленном порядке? Убедитесь, что результаты начинаются после прибытия мигрантов и заканчиваются до их отъезда или смерти. Все эти вопросы должны быть документированы в техническом отчете / руководстве пользователя. В некоторых случаях они могут потребовать дополнительных указаний для пользователей, чтобы избежать определенных типов анализа.

В Венгрии хорошо известно, что административные источники недооценивают реальный размер потоков эмиграции, поскольку мигранты часто не сообщают о своих миграционных событиях властям. Однако предполагается, что административные данные по различным подгруппам не смещены одинаковым образом. Чтобы выявить подгруппы мигрантов, миграционные потоки которых систематически недооцениваются в административных источниках, административные данные сравнивались с данными обследования Microcensus (микрперепись). Это сравнение выявило основные слабости административных источников, особенно при работе с различными формами множественной миграции.

285. Любые ограничения, выявленные на этапах 1 или 2, следует пересмотреть с учетом окончательных результатов. Есть ли условно подставленные записи? Вызывают ли они со временем ложные результаты, которыми нельзя пренебречь? Были ли устранены выявленные проблемы с охватом или измерением?

286. Если база данных обновляется, проверьте результаты напрямую по сравнению с предыдущей версией. Есть ли заметные различия в связанной популяции? Есть ли заметные различия в результатах?

287. Наконец, необходимо рассмотреть базу данных с учетом целей, определенных в Статистическом дизайне. Для выполнения этой задачи начните предварительный анализ в соответствии с целями, заявленными на Этапе 1. Если доступны альтернативные источники данных, сравните полученные результаты. Сравните результаты для различных когорт и подмножеств населения. В частности, если существуют определенные закономерности, ожидаемые существующей литературой, проверьте, сохраняются ли эти закономерности. Графический анализ может помочь выявить любые аномальные результаты с течением времени.

288. Задokumentируйте выводы в отчете и установите, можно ли использовать эту базу данных для достижения целей, изложенных на этапе 1. Если есть типы анализа, которых следует избегать или которые необходимо модифицировать, это должно быть четко указано в техническом отчете / руководстве пользователя.

3.8.3 Подготовка словаря данных и технического отчета/ руководства пользователя

289. Теперь пришло время предоставить пользователям соответствующие справочные материалы для проведения их анализа. Это будет включать словарь(и) данных, которые предоставляют определения и охватывают любые проблемы качества данных, абсолютные или специфичные для определенного периода времени, относящиеся к переменным. Технический отчет/ руководство пользователя даст общий обзор базы данных, различных соображений по качеству данных, выявленных на протяжении этой главы, и рекомендации о том, как следует использовать базу данных для целей, определенных на этапе Статистического дизайна.

290. В качестве напоминания, вот ключевые проблемы качества данных, отмеченные на предыдущих этапах, которые должны быть задokumentированы либо в словаре(ях) данных, либо в техническом отчете/ руководстве пользователя.

Этап 1: Статистический дизайн

- Различия между понятиями, представляющими интерес, и измеряемыми понятиями
- Различия между популяциями, представляющими интерес, и измеряемыми популяциями
- Различия между временными элементами, представляющими интерес, и измеряемыми референсными периодами

Этап 2: Оценка и предварительная обработка исходных файлов

- Ошибки охвата исходных файлов до и после предварительной обработки
- Ошибка выборки исходных файлов до и после предварительной обработки
- Пропущенные записи (или отсутствие ответа) в исходных файлах до и после предварительной обработки
- Ошибки измерения или пропущенные значения в исходных файлах до и после предварительной обработки
- Дубликаты, выявленные в исходных файлах

Этап 3: Интеграция данных для продольных данных

- Коэффициенты ошибок связывания (ложные срабатывания и ложные отрицания)
- Коэффициенты связывания (тщательно продумайте, какие отдельные источники использовать в знаменателе)

Этап 4: Назначение лонгитюдных индивидуальных идентификаторов

- Проблемы, связанные с индивидуальными идентификаторами с течением времени

Этап 5: Создание окончательной базы данных

- Проблемы с гармонизацией содержания
- Проблемы отбора записей

Этап 6: Распространение окончательных результатов

- Проблемы охвата после интеграции данных
- Пропущенные или противоречивые наблюдения после интеграции данных
- Пригодность базы данных для использования

291. Словарь данных должен содержать запись для каждой переменной из файлов Результаты и Когорты. Каждая запись должна включать, как минимум:

- Имя переменной (в соответствии со стандартным соглашением об именах, установленном на Этапе 5)
- Определение или описание переменной
- Источник переменной (наряду с любыми соображениями по гармонизации содержания на Этапе 5)
- Референсные периоды, для которых переменная доступна
- Любые соображения относительно качества данных, специфичные для переменной
- Ссылки на соответствующие классификации

292. Словари данных не должны углубляться в рекомендации по использованию отдельных переменных для анализа или подробно описывать концептуальные различия между идеальным и измеряемым. Однако запись о переменной должна четко указывать, что измеряется и как.

293. Технический отчет / руководство пользователя должен охватывать гораздо более подробную информацию. Он должен включать разделы, описывающие исходные файлы, методы, использованные для предварительной обработки, методологию, использованную для интеграции данных, и все другие шаги, предпринятые при построении этой базы данных. Он также должен охватывать все элементы качества данных, рассмотренные выше и на протяжении этой главы. Наконец, он должен предоставлять пользователям рекомендации по правильному использованию базы данных. Эти рекомендации должны включать объяснения того, где понятия отличаются от измеряемого. Эти различия не обязательно должны препятствовать проведению анализа, но они могут означать, что анализ необходимо соответствующим образом скорректировать.

Например, Статистическое управление Канады публикует подробный технический отчет по Лонгитюдной базе данных по иммигрантам (IMDB).

294. Описания лонгитюдных показателей в Главе 4 включают некоторые примеры случаев, когда измеряемое не полностью соответствовало понятиям, представляющим интерес. Однако анализ все же может продолжаться с другим подходом и давать полезные результаты.
295. Разработка источника данных всегда является сложной задачей, но разработка лонгитюдного источника данных, связанного с мигрантами, посредством интеграции данных включает множество различных препятствий, не типичных для простого проведения перекрестного выборочного обследования общего населения. Следует отметить, что хотя эти проблемы существуют, и хотя конечный продукт не будет идеальным, он все же может быть мощным источником данных и давать ценную информацию о мигрантах и их условиях. Следующая глава показывает, как вывести эту базу данных на регулярное распространение.

3.9 РЕЗЮМЕ

296. В этой главе было представлено руководство о том, как разработать лонгитюдный набор данных для статистики миграции на основе интегрированных данных.
297. Сначала была изложена рамка ошибок для лонгитюдных данных из интегрированных источников, разработанная Соединенным Королевством (Blackwell & Rogers, 2019). Цель применения рамки погрешностей административных данных — обеспечить оптимизацию качества данных на всех этапах статистического проекта. Рамка предлагает поэтапный подход к пониманию качества данных, рассматривая сначала отдельные наборы данных, прежде чем оценивать ошибки в производстве наборов данных, созданных путем интеграции нескольких источников.
298. Затем в главе рассмотрены шесть этапов разработки лонгитюдного набора данных, суммированных на Рисунке 4 на странице 29:
 1. Статистический дизайн
 2. Оценка и предварительная обработка исходных файлов
 3. Интеграция данных для лонгитюдных данных
 4. Назначение лонгитюдных индивидуальных идентификаторов
 5. Создание окончательной базы данных
 6. Распространение результатов
299. Приведены конкретные примеры из национальных статистических управлений Бельгии, Канады, Германии, Венгрии, Италии, Казахстана, Испании, Швейцарии, Турции и Соединенного Королевства.

ГЛАВА 4

РАСПРОСТРАНЕНИЕ РЕГУЛЯРНОЙ СТАТИСТИКИ МИГРАЦИИ ИЗ ИСТОЧНИКОВ ПРОДОЛЬНЫХ ДАННЫХ

300. В этой главе представлены рекомендации о том, как можно распространять регулярную статистику миграции с использованием лонгитюдных источников данных. В частности, эта глава будет включать два раздела:
- Ключевые лонгитюдные показатели
 - Передовая практика распространения лонгитюдной статистики
301. В первом разделе будет представлено разнообразие лонгитюдных (продольных) показателей, которые могут быть произведены для статистики миграции. Эти показатели включают темы, связанные с самим процессом миграции, а также с социально-экономическими результатами мигрантов с течением времени. Представлены особые проблемы, связанные с каждой темой показателя, и предложены возможные подходы к решению этих проблем с использованием примеров из разных стран. Проблемы и примеры сосредоточены на использовании административных данных или интегрированных источников данных.
302. Во втором разделе будут рассмотрены передовые практики распространения лонгитюдной статистики миграции. Это будет включать обзор ключевых аудиторий для этой статистики, а также сводку различных методов распространения, включая новые подходы, такие как интерактивные приложения и инфографика.

4.1 КЛЮЧЕВЫЕ ЛОНГИТЮДНЫЕ ПОКАЗАТЕЛИ

303. Основываясь на публикации ЕЭК ООН «Измерение изменений в социально-экономических условиях мигрантов» (ЕЭК ООН 2015), в этом разделе предлагаются показатели по лонгитюдным результатам, то есть по результатам, которые происходят в течение определенного периода времени.
304. В контексте данной публикации «показатели» относятся к статистике общего интереса, которую можно регулярно воспроизводить. Они отличаются от статистики, созданной для целей конкретного исследования. Показатели могут использоваться для информирования государственной политики или программных решений, но также могут служить другим заинтересованным сторонам, включая общественные организации и самих мигрантов.
305. В качестве отправной точки перечислены ключевые темы показателей для основных областей интереса для лонгитюдных данных. В рамках каждой темы может быть предложено несколько конкретных показателей. В то время как ключевые темы показателей могут относиться к общим терминам, таким как «продолжительность времени до события», конкретные показатели могут включать долю мигрантов, которые испытали это событие через X лет, среднее время до события или другие временные меры. Набор показателей может быть определен для подмножеств мигрантов (например, лица, ищущие убежища, временные резиденты и т.д.) и может быть перекрестно табулирован по демографическим или социально-экономическим статусным переменным (например, возраст, пол, страна рождения, образование, условия приема и т.д.).
306. Для целей этой главы термин «мигрант» будет использоваться в общем смысле для набора тем показателей и анализов. Однако некоторые показатели могут применяться только к определенным субпопуляциям. Например, не все мигранты будут входить в сферу охвата показателей, связанных с изменением правового статуса или гражданства. Важно учитывать интересующие популяции для каждого набора показателей, где это применимо. В примерах конкретные популяции мигрантов указаны четко.
307. Каждый показатель будет описан, и будут рассмотрены проблемы, связанные с его получением. Будут предоставлены практические примеры, включая случаи, когда практический показатель отклоняется от желаемого из-за ограничений источников данных. Цель этой главы — предоставить примеры описательной статистики, которые могут использоваться в качестве регулярных лонгитюдных показателей. Более сложные лонгитюдные аналитические методы, такие как анализ

выживаемости или обобщенные оценочные уравнения, не рассматриваются в этой главе, но могут быть использованы для получения более глубоких сведений о лонгитюдных результатах мигрантов. Наборы показателей, определенные в этой главе, могут быть использованы в качестве результирующих переменных с использованием этих методов.

308. Важно учитывать общие ограничения доступных источников данных при рассмотрении того, какие показатели можно оценить. В Главе 2 изложены некоторые сильные стороны и ограничения обследований, административных данных и регистров, а в Главе 3 более подробно рассматриваются ограничения, которые могут существовать для интегрированных источников данных.
309. Одной из глобальных проблем является временная мера, необходимая для лонгитюдных показателей. Если данные наблюдаются не в реальном времени, а в периодических (например, ежегодных) сегментах, время не может быть измерено так точно. Другая сложность может существовать в определении момента начала отсчета — начинается ли он после первого прибытия мигранта после того, как он становится постоянным резидентом, и т.д.? Эти вопросы необходимо решить для всех этих показателей.
310. Лонгитюдные показатели можно классифицировать на три основные категории:
- Миграционные модели
 - Социально-экономические результаты
 - Семейная миграция
311. В таблице под каждым показателем указаны страны, которые представляют конкретные примеры этих статистических данных в сочетании со своими существующими данными на продольной или поперечной основе. Следует отметить, что доступность показателей не обязательно связана с регулярной публикацией этих статистических данных. Хотя многие страны могут располагать доступными данными, эти показатели могут подготавливаться не на регулярной основе, а, например, в рамках отдельной или разовой публикации.

ТАБЛИЦА 6: КЛЮЧЕВЫЕ ЛОНГИТЮДНЫЕ ПОКАЗАТЕЛИ

Категория	Тема	Показатель
Миграционные модели	Продолжительность проживания в стране	Доля мигрантов, все еще находящихся в стране по истечении определенного периода времени
		Среднее время, в течение которого мигранты покидают страну (Суммарная) продолжительность пребывания с момента первого прибытия
		Продолжительность пребывания с момента последнего прибытия
	Средняя продолжительность пребывания (при более чем одном пребывании на душу населения)	Доля мигрантов, остающихся на территории региона по истечении определенного периода времени
	Среднее время, в течение которого мигранты покидают регион	Среди тех, кто остается в стране, доля мигрантов, которые остаются в субнациональном образовании по прошествии определенного периода времени
	Продолжительность проживания в субнациональном образовании	Среди тех, кто остается в стране, среднее время, за которое мигранты покидают субнациональное образование
	Время до смены места жительства или правового статуса, включая гражданство	Доля временных или краткосрочных резидентов, которые становятся постоянными или долгосрочными по истечении определенного периода времени
		Среднее время, за которое временные или краткосрочные резиденты становятся постоянными или долгосрочными
		Доля лиц, ищущих убежища, которым разрешается постоянное проживание по истечении определенного периода времени
		Среднее время, в течение которого просители убежища получают разрешение на постоянное проживание

		Доля мигрантов иностранного происхождения, которые приобретают национальное гражданство по истечении определенного периода времени Среднее время, в течение которого мигранты иностранного происхождения приобретают национальное гражданство
	Показатели циркулярной миграции	Среднее число пребываний на одного «циркулярного мигранта» (требует дальнейшего определения, например, для цензурированных наблюдений, периода времени, требуемой минимальной продолжительности пребывания и т.д.) Среднее число пребываний на одного «циркулярного мигранта»
Социально-экономические результаты	Языковые навыки	Доля мигрантов, которые могут говорить на официальных языках страны по истечении определенного периода времени Среднее время, в течение которого мигранты смогут говорить на официальных языках страны, Доля мигрантов, которые говорят на официальном языке (языках) страны дома по истечении определенного периода времени Среднее время, в течение которого мигранты будут говорить дома на официальном языке (языках) страны
	Владение домом	Доля мигрантов, которые приобретут свой дом в собственность по истечении определенного периода времени Среднее время, в течение которого мигранты впервые приобретут свой дом в новой стране проживания Срок владения жильем (например, арендное или собственное) мигрантами с течением времени Доля мигрантов, проживающих в субсидируемом или финансируемом государством жилье по истечении определенного периода времени Распределение по типам жилой недвижимости (например, квартиры, двухквартирные дома и т.д.) по истечении определенного периода времени
	Трудоустройство	Доля мигрантов, которые трудоустраиваются по истечении определенного периода времени Среднее время, в течение которого мигранты впервые получают работу в новой стране проживания Соотношение полной и частичной занятости мигрантов с течением времени Статус мигрантов как рабочей силы с течением времени
	Послешкольное образование	Доля мигрантов, получивших аттестат о среднем образовании, ученую степень или диплом о высшем образовании в своей новой стране проживания по истечении определенного периода времени в разбивке по возрасту Посещаемость школы мигрантами с течением времени в разбивке по возрасту
	Доход	Средний и медианный совокупный доход по истечении определенного периода времени Средний и медианный доход от трудовой деятельности по истечении определенного периода времени Доля мигрантов, получающих доход от трудовой деятельности по истечении определенного периода времени, доля мигрантов, получающих доход от самостоятельной занятости или бизнеса по истечении определенного периода времени Доля мигрантов, получающих социальную или государственную помощь по истечении определенного периода времени Время, пока средний или медианный совокупный доход мигрантов не сравняется со средним или медианным совокупным доходом населения в целом Время, пока средний или медианный доход мигрантов от занятости не сравняется со средним или медианным совокупным доходом населения в целом Доля мигрантов с низким уровнем дохода по прошествии X лет
	Владение бизнесом/	Доля мигрантов, которые стали самозанятыми или владеют

предпринимательство	<p>бизнесом по истечении определенного периода времени</p> <p>Среднее время, в течение которого мигранты становятся самозанятыми или владеют бизнесом</p> <p>Число лиц, нанятых мигрантами по истечении определенного периода времени</p>
Здоровье	<p>Доля мигрантов, которые регистрируются у врача в системе здравоохранения принимающей страны по прошествии определенного периода времени (или времени до регистрации)</p> <p>Доля мигрантов, которые были госпитализированы в течение X лет с момента прибытия в принимающую страну проживания, в целом или по поводу различных заболеваний, например, хронических или инфекционных по своей природе (или время до первой госпитализации)</p> <p>Доля мигрантов, умерших с момента прибытия в принимающую страну проживания от различных заболеваний (по прошествии определенного периода времени) (или времени до смерти)</p>
Семейная миграция	<p>Среднее время между прибытием первого члена семьи в страну и прибытием последнего члена семьи (т.е. временной лаг в воссоединении семьи)</p> <p>Экономические результаты семей мигрантов после определенного периода пребывания в стране</p>

4.1.1 Миграционные модели

312. Лонгитюдные показатели, связанные с миграционными моделями, могут включать показатели, связанные с эмиграцией, циркулярной миграцией, внутренней миграцией и изменениями в правовом или резидентском статусе. Что касается показателей циркулярной миграции, читателям рекомендуется обратиться к публикации ЕЭК ООН "Определение и измерение циркулярной миграции" (ЕЭК ООН, 2016) и последующим работам Евростата (2017а, 2018).

313. Кроме того, поскольку данный раздел выходит за рамки демографических показателей и также охватывает изменения в правовом или административном статусе, следует отметить, что государственные департаменты, министерства и агентства, ответственные за администрирование вопросов, связанных с миграцией, играют важную роль в сотрудничестве с НСУ в определении показателей — особенно тех, которые имеют непосредственное значение для политики.

314. Показатели, связанные с этими темами, были недавно включены в Международные рекомендации по статистике беженцев (EGRIS, 2018) и Международные рекомендации по статистике внутренне перемещенных лиц (EGRIS, 2020). Глава 5 Международных рекомендаций по статистике беженцев включает некоторые предлагаемые лонгитюдные показатели по этой теме, в том числе (а) долю лиц в стране в поисках международной защиты, которые оставались в течение 5, 10 или более 10 лет, и (б) коэффициент признания.

315. Следующие темы были определены как ключевые лонгитюдные показатели, представляющие интерес для стран-членов ЕЭК ООН, связанные с миграционными моделями:

- Продолжительность проживания в стране
- Продолжительность проживания в субнациональном образовании
- Время до смены места жительства или правового статуса, включая гражданство
- Показатели циркулярной миграции.

316. В таблице 7 приведены примеры, предоставленные странами по этим темам.

ТАБЛИЦА 7: КЛЮЧЕВЫЕ ТЕМЫ ПОКАЗАТЕЛЕЙ ДЛЯ МИГРАЦИОННЫХ МОДЕЛЕЙ С ПРИМЕРАМИ ИЗ СТРАН

Ключевые темы показателей	Канада	Германия	Италия	Испания	Швейцария	Турция	Великобритания
Продолжительность проживания в стране	x			x	x	x	x

Продолжительность проживания в субнациональном образовании	x	x	x				
Время до смены места жительства или правового статуса, включая гражданство	x	x	x		x		x
Показатели циркулярной миграции*							

* См. материалы ЕЭК ООН (2016) и Евростат (2017а и 2018).

4.1.1.1 Продолжительность проживания в стране

317. В конечном счете, модели международной миграции могут быть сложными, и лица могут въезжать в страну и выезжать из нее несколько раз. Отслеживание множества возможных моделей по своей природе является лонгитюдным. В данном разделе рассматриваются простые случаи единичных измерений (например, время с момента последнего прибытия против времени с момента первого прибытия). Более сложные модели рассматриваются в существующих отчетах, подготовленных ЕЭК ООН и Евростатом (подробнее см. раздел 4.1.1.4 Лонгитюдная статистика циркулярной миграции).

318. Для измерения продолжительности проживания в стране предлагаются следующие показатели:

- Доля мигрантов, все еще находящихся в стране спустя определенное время
- Среднее время до того, как мигранты покидают страну
- (Совокупная) продолжительность проживания с момента первого прибытия
- Продолжительность проживания с момента последнего прибытия
- Средняя продолжительность проживания (когда более одного пребывания на человека)

319. Эти показатели имеют большое значение для понимания миграционных моделей и их влияния на принимающие страны. Они могут использоваться для понимания всего: от текущего состава населения до факторов, влияющих на удержание мигрантов. Поскольку проживание может быть связано с различным законодательством и политикой (например, получение права на гражданство), странам важно учитывать сложный характер этих измерений. В зависимости от потребностей, наиболее интересным показателем может быть время с момента первого прибытия или время с момента последнего прибытия. Также может представлять интерес краткосрочное, но повторяющееся пребывание.

320. В этом контексте также важно знать, возвращаются ли мигранты в свою первоначальную страну происхождения или переезжают в третью страну. Страна происхождения, которая может относиться к стране их рождения или стране последнего проживания, может быть переменной, перекрестно табулированной с этими показателями, чтобы добавить такую информацию. Страны назначения эмигрантов дают ценную информацию о демографическом обмене между странами. Точно так же страна происхождения может информировать о продолжительности проживания перед эмиграцией (например, прибытие непосредственно из их страны происхождения или через другую промежуточную страну).

321. Это показатели, которые по своей сути являются лонгитюдными, поскольку место жительства лиц (внутри или вне страны) необходимо наблюдать с течением времени. Однако, если изменения места жительства не наблюдаются в режиме реального времени, существуют некоторые ограничения для разработки этих показателей.

322. Даже среди стран с регистрами населения может быть запаздывание между выбытием и его обнаружением в регистре — это может привести к завышению оценки времени, которое лица остаются в стране. Более того, в некоторых случаях выбытие может вообще не фиксироваться. Например, Статистическое управление Нидерландов (Prins, 2016) сообщает, что: «[...] около одного из трех жителей, которые покидают страну для проживания за границей, не уведомляют местное правительство о своем отъезде».

323. Для стран, полагающихся на ежегодные административные данные, отсутствие информации, специально предназначенной для измерения эмиграции, снижает точность, с которой может быть

произведен показатель. Например, отсутствие конкретной даты эмиграции в ежегодных административных данных может привести к годовым показателям, а не к более точному измерению времени (например, среднее число лет против среднего числа дней до того, как мигранты покинут страну). Во многих случаях административные данные опираются на концепции, которые могут не отражать истинную эмиграцию, как ее определяют демографы. Они также могут быть ограничены обычными недостатками административных данных, такими как охват населения и запаздывания.

324. Другое ограничение для стран без регистра населения заключается в том, что это измерение, вероятно, будет ограничено наличием и отсутствием сигналов. В этих случаях может быть некоторая ошибка измерения, когда отсутствие сигналов не указывает на истинное отсутствие (а в некоторых случаях наличие сигналов может не отражать истинное присутствие). Более того, в некоторых случаях разные сигналы могут давать противоречивую информацию. Смерть является потенциально искажающим фактором, поскольку она также может привести к отсутствию сигналов, если данные не могут полностью учесть это событие. Например, в Канадской лонгитюдной базе данных по иммиграции мигрант, который перестал подавать налоговые декларации, может сигнализировать о том, что это лицо умерло, эмигрировало или продолжает жить в стране, но просто не подало налоговую декларацию. Также следует отметить, что технологическое развитие позволяет постоянно растущие возможности быть подключенным удаленно, без необходимости фактического присутствия в интересующем географическом районе. Это может снизить актуальность сигналов административной жизни как индикатора фактического присутствия с течением времени.

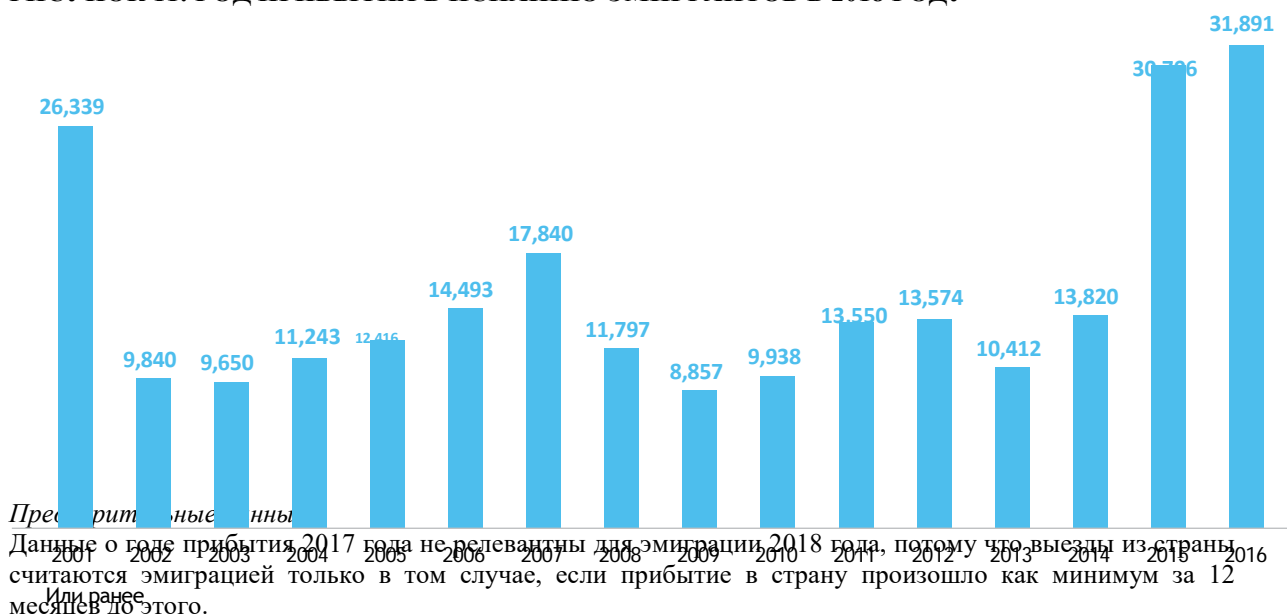
Испания публикует показатели, связанные с этой темой, включая:

- Эмигранты по годам, году прибытия в Испанию и гражданству
- Эмигранты по годам, году прибытия в Испанию и стране рождения

Основным источником информации, касающейся как запасов населения, так и статистики миграции в Испании, является регистр населения, называемый Padrón на испанском языке. Padrón — это официальный список жителей в каждом из 8 124 муниципалитетов (по состоянию на 1 января 2018 года) в Испании. Но Padrón также является лонгитюдной базой данных, включающей 46,7 миллиона человек, проживающих в Испании (по состоянию на 1 января 2018 года), и более 265 миллионов записей с лонгитюдной информацией о лицах, проживающих в Испании с 1996 по 2018 год. Также фиксируются все предыдущие места жительства (в пределах Испании), что позволяет проводить лонгитюдный анализ и очень точный мониторинг внутренней миграции. Хотя потенциал Padrón для лонгитюдного анализа велик, на данный момент Национальный институт статистики Испании (INE) делает лишь первые шаги.

В Испании статистика миграции, распространяемая INE с 2013 года, включает информацию о количестве иммигрантов и эмигрантов с разбивкой по возрасту/ поколению, полу, национальности, стране рождения, стране происхождения и стране назначения. Впервые в июне 2019 года была распространена информация о продолжительности проживания в Испании тех, кто уже покинул страну. Эта информация помогает обогатить информацию, касающуюся эмигрантов. На графике ниже показано распределение эмигрантов 2018 года по году прибытия.

РИСУНОК 11: ГОД ПРИБЫТИЯ В ИСПАНИЮ ЭМИГРАНТОВ В 2018 ГОДУ



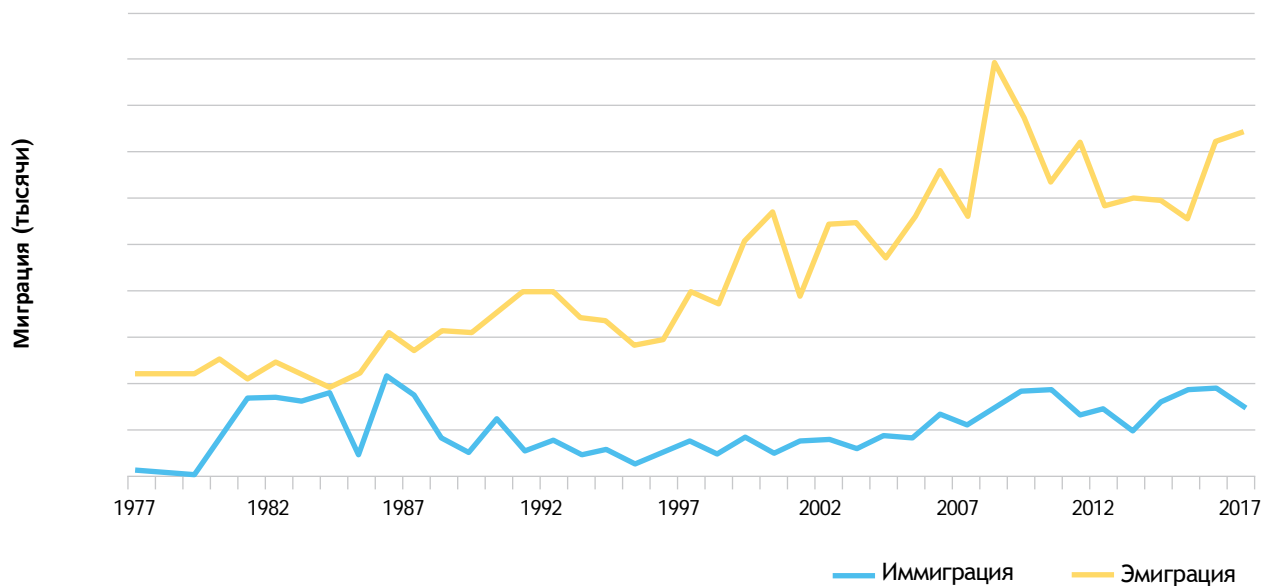
Источник: Национальный институт статистики Испании (INE), Padrón, 2018.

325. Более подробная информация об этих показателях доступна в INE (2018).

В настоящее время Соединенное Королевство публикует данные о запасах и потоках международных мигрантов. Не существует единого источника, который мог бы обеспечить измерение всех перемещений людей в Великобританию и из нее. В настоящее время Великобритания использует комбинацию данных из разных источников для расчета официальных оценок долгосрочной международной миграции (LTIM) Великобритании, которые публикуются каждые три месяца в Ежеквартальном отчете по статистике миграции Управления национальной статистики. Оценки LTIM примерно на 90 процентов основаны на данных Обследования международных пассажиров (IPS), выборочного обследования, проводимого во всех основных портах Великобритании. IPS фиксирует намерения мигрантов и используется для идентификации международных мигрантов, въезжающих в Великобританию и выезжающих из нее. Для оценки LTIM данные IPS дополняются данными Министерства внутренних дел и Агентства статистики и исследований Северной Ирландии (NISRA), и вносятся несколько корректировок для получения более полной оценки миграции.

Также публикуются наблюдаемые данные о продолжительности миграции для людей, въезжающих в Великобританию или выезжающих из нее. Эти оценки основаны на IPS. Фактическая продолжительность пребывания устанавливается путем задания ретроспективного вопроса в IPS либо о годе выезда из Великобритании (для тех возвращающихся иммигрантов, которые ранее выехали как долгосрочные эмигранты), либо о годе прибытия для тех, кто выезжает из Великобритании на долгосрочной основе. Фактическая продолжительность пребывания рассчитывается простым вычитанием года прибытия/отъезда из года проведения интервью (Рисунок 12).

РИСУНОК 12: ОЦЕНКИ ДОЛГОСРОЧНОЙ МЕЖДУНАРОДНОЙ МИГРАЦИИ В ВЕЛИКОБРИТАНИЮ И ИЗ НЕЕ, 1975-2017 ГГ. НЕБРИТАНЦЫ, ФАКТИЧЕСКАЯ ПРОДОЛЖИТЕЛЬНОСТЬ ПРЕБЫВАНИЯ 1-4 ГОДА



Источник: Управление национальной статистики Великобритании, Обследование международных пассажиров, 1975-2017 (2019).

Национальные оценки краткосрочной международной миграции производятся непосредственно на основе данных Обследования международных пассажиров (IPS) и публикуются ежегодно в Бюллетене по краткосрочной международной миграции (STIM) для Англии и Уэльса. Оценки STIM производятся для Англии и Уэльса и для местных органов власти. Оценки краткосрочной миграции доступны в форме потоков (общее количество перемещений за установленный период) и запасов (среднее количество краткосрочных мигрантов в стране в средний день за 12-месячный период). Следует отметить, что человек может мигрировать более одного раза в один и тот же период, поэтому данные представляют собой подсчет перемещений мигрантов, а не отдельных лиц. Великобритания также публикует данные о жителях Великобритании по стране рождения и гражданству, как сообщается в Ежегодном обследовании

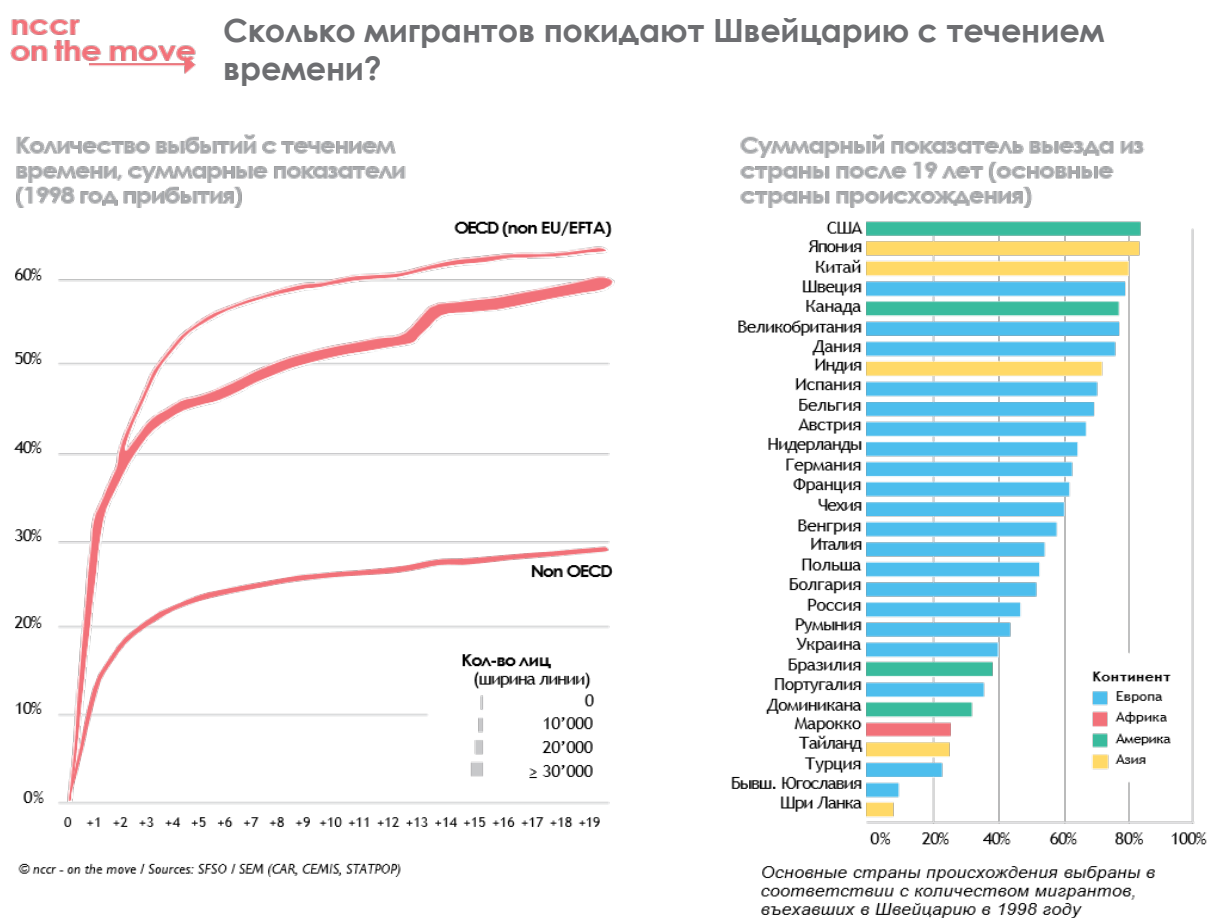
населения (APS). Жители могли прибыть недавно или проживать в Великобритании в течение ряда лет, и они представляют собой подсчет запасов нетитульного населения.

В будущем Управление национальной статистики Великобритании намерено объединить обследования и административные данные для формирования лонгитюдных наборов данных и понимания «следа», который мигранты оставляют, находясь в стране, например, путем измерения продолжительности пребывания и идентификации возвратной или циркулярной миграции иностранных сезонных рабочих.

Швейцария располагает тщательно разработанным набором показателей, ориентированных на внешнюю миграцию. Они основаны на швейцарской лонгитюдной демографической базе данных (SLDD), связанной базе данных на основе регистров, недавно разработанной в рамках национального научного проекта NCCR «В движении», где NCCR является Национальным центром научных исследований в области миграции и мобильности. Он создан путем объединения данных Центрального реестра иностранных граждан (ZAR) и базы данных статистики населения и домашних хозяйств (STATPOP). Она распространяется на всех швейцарских граждан и иностранцев, проживающих в Швейцарии.

Первый анализ, основанный на SLDD, подчеркнул тот факт, что мигрантское население в Швейцарии довольно мобильно, поскольку мигранты обычно не остаются в Швейцарии всю жизнь, а только на несколько лет. Например, из иммигрантов, прибывших в Швейцарию в 1998 году, более 80 процентов покинули страну в течение 17 лет после их прибытия. Поскольку краткосрочные мигранты могут быть менее склонны к интеграции в страну назначения, этот вывод актуален для оценки в свете текущей политики интеграции. Швейцария публикует коэффициенты выбытия с течением времени в интерактивном графике через партнерство «В движении». График ниже демонстрирует, что продолжительность пребывания или коэффициент выбытия варьируется в зависимости от страны происхождения.

РИСУНОК 13: КУМУЛЯТИВНЫЕ КОЭФФИЦИЕНТЫ ВЫБЫТИЯ ДЛЯ МИГРАНТОВ, ПРИБЫВШИХ В ШВЕЙЦАРИЮ В 1998 ГОДУ, ПО ГОДАМ С МОМЕНТА ПРИБЫТИЯ И СТРАНЕ ПРОИСХОЖДЕНИЯ



Источник: NCCR «В движении», Швейцарская лонгитюдная демографическая база данных (SLDD), 2017.

Дополнительная информация доступна на:

<https://nccr-onthemove.ch/publications/towards-a-new-data-set-for-the-analysis-of-migration-and-integration-in-switzerland/>

http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.10/2017/mtg1/2017_UNECE_Migration_WP_07_Switzerland_WannerHeiniger_ENG.pdf

Демографическая программа оценок Канады требует оценок запасов и потоков временных жителей (NPRs — также называемых временными резидентами) для оценки населения страны между переписями. Для этого Статистическое управление Канады использует ежемесячную информацию об административных операциях (записи о легализации, ходатайства о предоставлении убежища, разрешения на работу, разрешения на учебу, выдворения и заявления на легализацию) от агентства Иммиграции, беженцев и гражданства Канады (IRCC).

Статистическое управление Канады интегрирует эти файлы и строит лонгитюдные профили NPR. Используя некоторые базовые предположения, биографии из лонгитюдных данных могут быть использованы для измерения административной продолжительности пребывания каждого NPR с целью производства оценок запасов и потоков. Оно также учитывает, если лицо имеет более одного разрешения в один и тот же период.

Эти данные имеют два основных ограничения. Во-первых, аналогично административным данным по иммиграции, данные NPR собирают предполагаемое место назначения, а не место, где иммигрант фактически поселился. Во-вторых, данные NPR собирают только дату начала и окончания действия разрешения, а не фактические даты прибытия и отъезда NPR. В результате продолжительность пребывания в стране аппроксимируется с использованием срока действия разрешения(ий), имеющихся у NPR. Более того, отсутствует информация о датах отъезда NPR, поскольку некоторые из них покидают страну до окончания срока действия их разрешений. Поскольку не требуется уведомлять федеральное правительство перед отъездом из страны, Статистическое управление Канады в настоящее время не имеет дополнительных источников для получения этой информации.

В Турции Система регистрации населения на основе адресов (ABPRS) является уникальной системой, на основании которой распространяется ежегодная статистика населения. В этом контексте ежегодная информация о запасах и потоках международной миграции по гражданству и стране рождения производится на основе ABPRS и других дополнительных административных источников.

В 2019 году в рамках исследований по разработке статистики, получаемой из административных регистров, было завершено исследование «продолжительности проживания иностранных жителей», и был получен показатель «иностранное население по первому году проживания в Турции». Охват включал иностранное население, зарегистрированное в ABPRS, и «первый год проживания» здесь относится к году, когда иностранный житель начал свое непрерывное проживание до даты отчета, т.е. 31 декабря 2018 года для первого исследования.

Информация была произведена на основе ретроспективного наблюдения и анализа иностранного населения. В анализе использовались не только последние, но и предыдущие годы ABPRS вместе с данными по международной миграции и записями разрешений на проживание/работу за соответствующие годы.

Более подробная информация о показателе доступна на http://www.turkstat.gov.tr/PreTablo.do?alt_id=1067

4.1.1.2 Продолжительность проживания в субнациональном образовании

326. Аналогично продолжительности проживания в стране, для измерения продолжительности проживания в субнациональном образовании предлагаются следующие показатели:

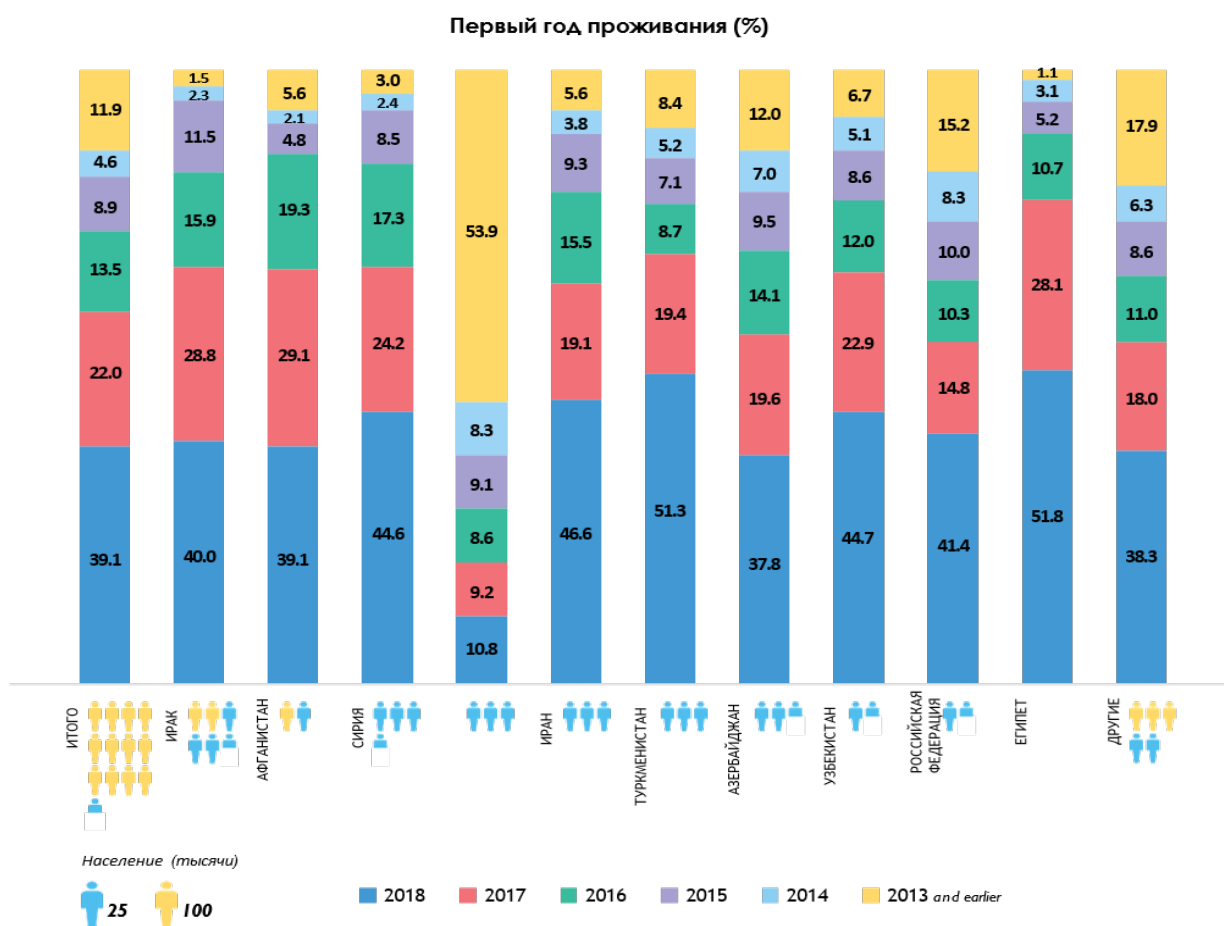
- Доля мигрантов, которые остаются в субнациональном образовании по прошествии определенного периода времени
- Среднее время до того, как мигранты покидают субнациональное образование
- Среди тех, кто остается в стране, доля мигрантов, которые остаются в субнациональном образовании по прошествии определенного периода времени
- Среди тех, кто остается в стране, среднее время до того, как мигранты покидают субнациональное образование

327. Кроме того, могут быть некоторые более специфичные для страны вопросы, такие как:

- Доля мигрантов, которые переезжают из сельских в городские районы по прошествии определенного периода времени
- Доля мигрантов, которые переезжают из восточных в западные регионы по прошествии определенного периода времени

328. Для показателей, которые должны учитывать международные перемещения, применимы проблемы, поднятые выше для «Продолжительности проживания в стране».
329. В противном случае, модели субнациональной миграции накладывают некоторые новые (или усугубляющие) ограничения, связанные с географией. Изменение географических границ создает проблему для лонгитюдного анализа миграции. При изменении границ отдельные лица могут выступать в качестве переселенцев, даже если они остались в том же месте жительства. Это связано с тем, что их официальная географическая зона могла измениться в результате изменения границы. Хотя это также может быть проблемой для международной миграции, изменение внутренних границ может быть гораздо более распространенным явлением.
330. Другая проблема может возникнуть при объединении различных источников географической информации. Если продольные данные являются результатом интеграции нескольких источников данных, возможно, что некоторые источники используют разные географические определения, чем другие. Названия городов могут меняться от одного источника к другому, и источники могут предоставлять географическую информацию, которую нельзя сравнивать. Например, школьные или медицинские округа могут пересекаться с почтовыми индексами или границами городов. Столичный регион и муниципалитет могут иметь общее название, но относиться к разным географическим границам (например, переписная столичная зона Торонто, Канада, имела численность населения по переписи 2016 года 5 928 040 человек, но переписное подразделение (город) с тем же названием имело население 2 731 571 человек).

РИСУНОК 14: ИНОСТРАННОЕ НАСЕЛЕНИЕ ПО СТРАНЕ ГРАЖДАНСТВА И ПЕРВОМУ ГОДУ ПРОЖИВАНИЯ В ТУРЦИИ, ПО СОСТОЯНИЮ НА 31 ДЕКАБРЯ 2018 Г.



Источник: TurkStat, Система регистрации населения на основе адресов (ABPRS), 2013-2018.

331. Важно также учитывать перемещение в пределах географических границ. В некоторых случаях, особенно в случае крупных территорий, миграция в пределах географического района может

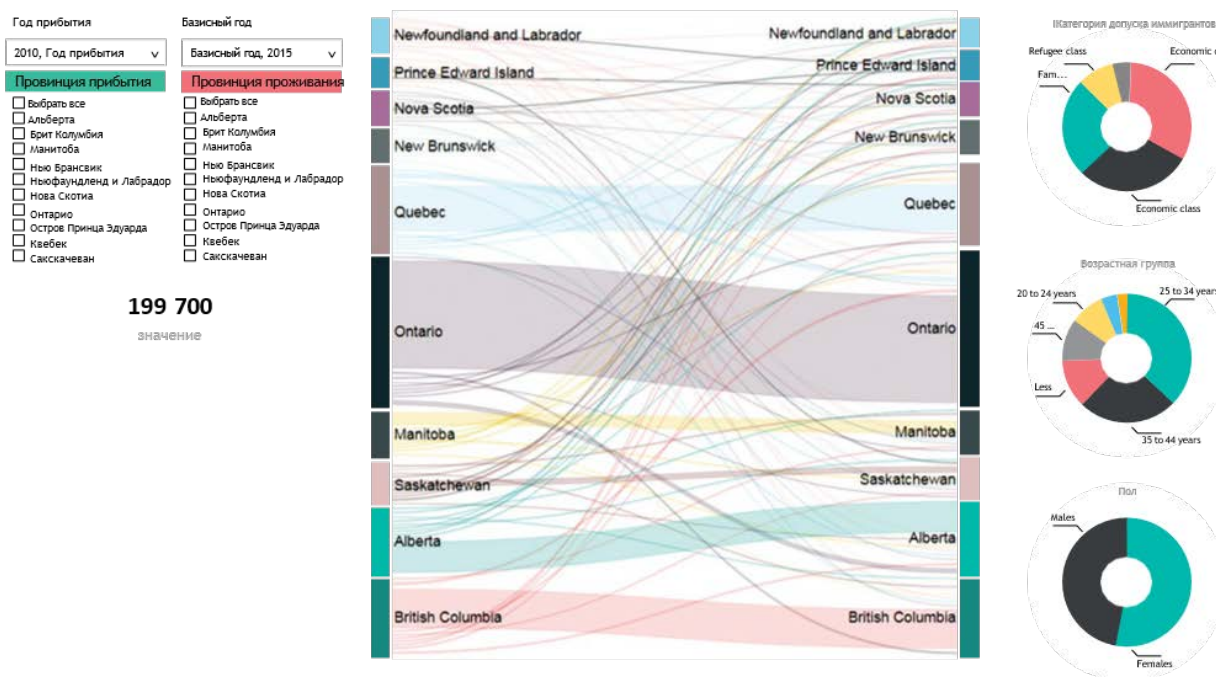
представлять собой перемещение на большие расстояния. С другой стороны, миграция из одного района в другой может осуществляться на короткие расстояния, если она происходит вблизи границ районов. Переезды на большие расстояния, вероятно, будут оказывать более глубокое воздействие на мигрантов, поскольку они могут переехать в сообщества с различными рынками труда, различными доступными услугами и различными социальными условиями.

Канада регулярно производит статистику по внутренним миграционным моделям иммигрантов на основе Лонгитюдной базы данных по иммиграции (IMDB). IMDB представляет собой интеграцию административных данных по иммиграции и ежегодных налоговых файлов. Ежегодные налоговые файлы предоставляют текущую географию иммигрантов по годам, а иммиграционные файлы предоставляют географию назначения, когда иммигрант был допущен в Канаду.

Одна проблема, специфичная для Канады, заключается в том, что в административных данных по иммиграции иммигрантов спрашивают, где они намерены проживать. Затем это сравнивается с их фактической географией проживания согласно их ежегодным налоговым файлам. Различия могут быть обусловлены реальными перемещениями между допуском в Канаду и подачей налоговых деклараций, но они также могут быть обусловлены концептуальными различиями (т.е. предполагаемая провинция проживания против первой фактической провинции проживания).

Несмотря на эти проблемы, IMDB является ключевым источником данных для понимания удержания иммигрантов по провинциям или субпровинциальным географическим областям. Диаграмма Санкей ниже иллюстрирует сдвиги между предполагаемыми провинциями проживания для иммигрантов-налогоплательщиков, допущенных в 2010 году, и их фактической провинцией проживания 5 лет спустя, в 2015 году. Эта диаграмма также позволяет пользователям понять характеристики тех лиц, которые остались в своей предполагаемой провинции проживания, по сравнению с теми, кто теперь живет в другом месте Канады.

РИСУНОК 15: ИЗМЕНЕНИЯ МЕЖДУ ПРЕДПОЛАГАЕМОЙ ПРОВИНЦИЕЙ И МЕСТОМ ЖИТЕЛЬСТВА ЧЕРЕЗ 5 ЛЕТ ПОСЛЕ ВЪЕЗДА В КАНАДУ ДЛЯ ИММИГРАНТОВ-НАЛОГОПЛАТЕЛЬЩИКОВ, ПРИНЯТЫХ В 2010 ГОДУ



Источник: Статистическое управление Канады, Лонгитюдная база данных по иммиграции, 2016.

В Италии связывание записей между видами на жительство позволяет связать международную миграцию с внутренней мобильностью. На рисунке ниже показан пример китайской миграционной сети, как внешней, так и внутренней. Красные линии указывают на прибытия из Китая в итальянские провинции в 2007 году, а зеленые линии представляют внутреннюю мобильность между 2007 и 2012 годами.

Германия в настоящее время отслеживает ежегодные компоненты изменения запасов населения (включая внутреннюю миграцию). Есть идеи расширить охват, рассматривая эти факторы с лонгитюдной

перспективы (Brückner, 2018). Набор данных позволит, например, отслеживать внутреннюю мобильность после иммиграции.

Предварительные результаты основаны на иностранцах, которые впервые мигрировали в Германию после 2006 года или позже и непрерывно оставались резидентами до конца 2017 года. Модели внутренней мобильности после миграции анализировались на уровне 401 округа. Использовались следующие потенциальные показатели для внутренней мобильности после миграции:

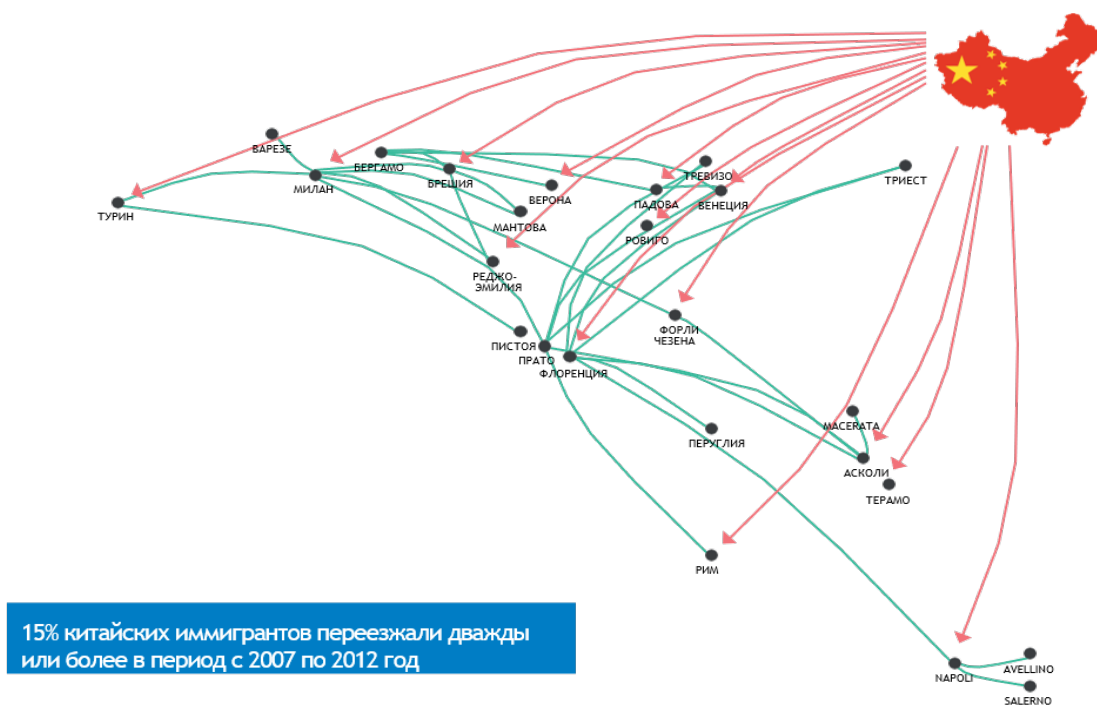
- количество перемещений через границы округов за определенный период (по продолжительности проживания),
- абсолютная региональная продолжительность проживания (годы, проведенные в том же округе), и относительная региональная продолжительность проживания (годы, проведенные в том же округе, в процентах от общей продолжительности проживания в Германии).

Германия рассматривает количество перемещений через границы округов как хороший показатель внутренней миграции, если он нормализован по общей продолжительности проживания иностранца. Кроме того, она рассматривает относительную региональную продолжительность проживания как наиболее гибкий показатель, поскольку его можно осмысленно интерпретировать, даже если общая продолжительность проживания различается.

Предварительные результаты обнадеживают: Относительная региональная продолжительность проживания варьируется в зависимости от характеристик иностранца — в частности, пола, возраста, семейного положения, гражданства и статуса вида на жительство. Продолжительность проживания самая длинная среди пожилых замужних женщин, имеющих гражданство государства-члена ЕС и, следовательно, свободу поселения ЕС в качестве статуса постоянного вида на жительство.

Относительная региональная продолжительность проживания также варьируется в зависимости от характеристик региона. Она самая длинная в регионах с высокой плотностью населения, высоким процентом иностранцев и высокими значениями ВВП на душу населения; она самая короткая в регионах с низкой плотностью населения, низким процентом иностранцев и низкими значениями ВВП на душу населения.

РИСУНОК 16: МИГРАЦИОННАЯ СЕТЬ ДЛЯ КИТАЙСКИХ МИГРАНТОВ, ПРИБЫВШИХ В ИТАЛИЮ В 2007 ГОДУ, С ВНУТРЕННЕЙ МОБИЛЬНОСТЬЮ МЕЖДУ 2007 И 2012 ГОДАМИ



Источник: ISTAT, Conti et al (2012).

4.1.1.3 Время до изменения правового статуса

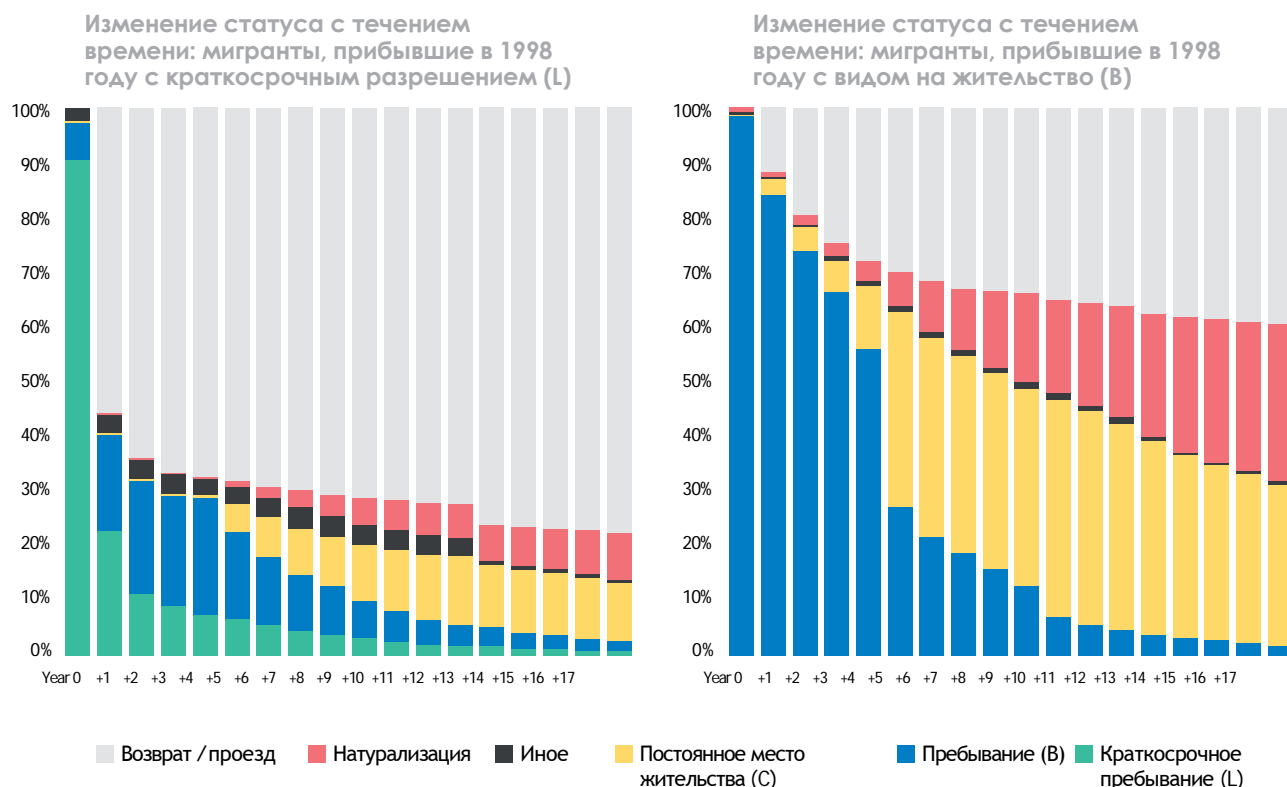
332. Типичные примеры изменения правового статуса мигранта включают получение статуса постоянного жителя, удовлетворение ходатайства лица, ищущего убежища, или получение гражданства. Для лонгитюдного измерения предлагаются следующие показатели:
- Доля временных или краткосрочных резидентов, которые становятся постоянными или долгосрочными резидентами по прошествии определенного периода времени
 - Среднее время до того, как временные или краткосрочные резиденты становятся постоянными или долгосрочными резидентами
 - Доля лиц, ищущих убежища, которые получают разрешение на постоянное поселение по прошествии определенного периода времени
 - Среднее время до того, как лица, ищущие убежища, получают разрешение на постоянное поселение
 - Доля мигрантов, родившихся за границей, которые приобретают гражданство страны по прошествии определенного периода времени
 - Среднее время до приобретения гражданства страны мигрантами, родившимися за границей
333. Если доступные административные данные связывают соответствующую информацию (например, записи временных и постоянных резидентов), эти показатели должны быть довольно простыми в производстве.
334. Временные резиденты могут вносить некоторую сложность для временного измерения, поскольку не всегда очевидно, когда началось их проживание в стране. Другой проблемой может быть наличие международных перемещений между первым временным проживанием лица в стране и моментом, когда они становятся постоянными или долгосрочными резидентами (например, кто-то приезжает в качестве временного студента, покидает страну, возвращается несколько лет спустя в качестве временного работника, затем становится постоянным резидентом). Таким образом, необходимо принять решения о том, когда время будет начинаться для этих показателей.
335. Этот раздел также включает показатели, относящиеся к натурализации. Для целей данного раздела натурализация относится к процессу, в ходе которого мигрант, родившийся за границей, то есть тот, кто либо родился иностранным гражданином, либо апатридом, становится гражданином своей новой страны проживания. Коэффициенты натурализации относятся к доле мигрантов, родившихся за границей, которые стали гражданами. Знаменатели для таких коэффициентов могут варьироваться в зависимости от того, включают ли они или исключают тех лиц, которые не имеют права на гражданство.
336. В зависимости от страны, не все мигранты, родившиеся за границей, могут иметь возможность стать гражданами. Например, в некоторых случаях может потребоваться минимальное количество лет проживания. В результате расчет этих показателей может быть ограничен теми мигрантами, которые находятся в стране. Например, доля мигрантов, ставших гражданами, была бы ниже, чем при аналогичном расчете, но ограничивалась бы теми, кто имеет право стать гражданами. Аналитикам следует обратить особое внимание на правовые механизмы, связанные с натурализацией в интересующей стране, чтобы тщательно выявлять мигрантов, которые имеют право на получение гражданства.
337. Аналогично другим показателям, если события наблюдаются не в режиме реального времени, а вместо этого в ежегодных сводках, временное измерение должно будет измениться с конкретных дат на годы и приведет к меньшей точности (например, количество лет до того, как временные резиденты становятся постоянными резидентами, количество лет до того, как постоянные резиденты становятся гражданами).
338. В течение некоторого времени подчеркивалась необходимость применения когортного подхода к изучению натурализации (Peglin, 2006). Эта точка зрения недавно была вновь предложена Райхелем (2011), который подчеркнул, что «для этого необходимо было бы основывать ставку на иностранном населении, фактически имеющем право на натурализацию, или, как это называют статистики, на населении, подверженном риску пережить какое-либо событие». Как указывает Перрен (2006), «когортный подход к измерению коэффициентов приобретения гражданства позволил бы рассчитать вероятность получения гражданства для отдельных когорт иммигрантов, что обеспечило бы гораздо лучшую меру воздействия политических мер на модели приобретения гражданства». Тот же автор также указывает, что отсутствие соответствующих данных затрудняет расчет лонгитюдных коэффициентов натурализации, которые требуют наличия информации о годе иммиграции. Однако это то направление, в котором движутся несколько стран, также благодаря доступности

статистических источников на основе регистров, которые позволяют извлекать лонгитюдные данные (Perrin, 2006).

Швейцария, через Швейцарскую лонгитюдную демографическую базу данных (SLDD), имеет показатели, связанные с тем, как правовой статус мигрантов развивается с течением времени, связанные с административными переменными, касающимися основания для допуска мигрантов в Швейцарию.

В дополнение к приведенным ниже графикам, для получения дополнительной информации см.: <https://indicators.nccr-onthemove.ch/how-does-the-migrants-legal-status-evolve-over-time/>

РИСУНОК 17: ИЗМЕНЕНИЕ ПРАВОВОГО СТАТУСА МИГРАНТОВ, ПРИБЫВШИХ В ШВЕЙЦАРИЮ В 1998 ГОДУ С КРАТКОСРОЧНЫМ РАЗРЕШЕНИЕМ (L) ИЛИ ВИДОМ НА ЖИТЕЛЬСТВО (B), ДО 2017 ГОДА



Источник: NCCR «В движении», Швейцарская лонгитюдная демографическая база данных (SLDD), 2017.

Швейцария также проявляет особый интерес к мониторингу коэффициентов натурализации и имеет несколько показателей по натурализации.

На следующих рисунках представлены коэффициенты натурализации по странам происхождения, кантону проживания и с течением времени, с 1998 по 2017 год.

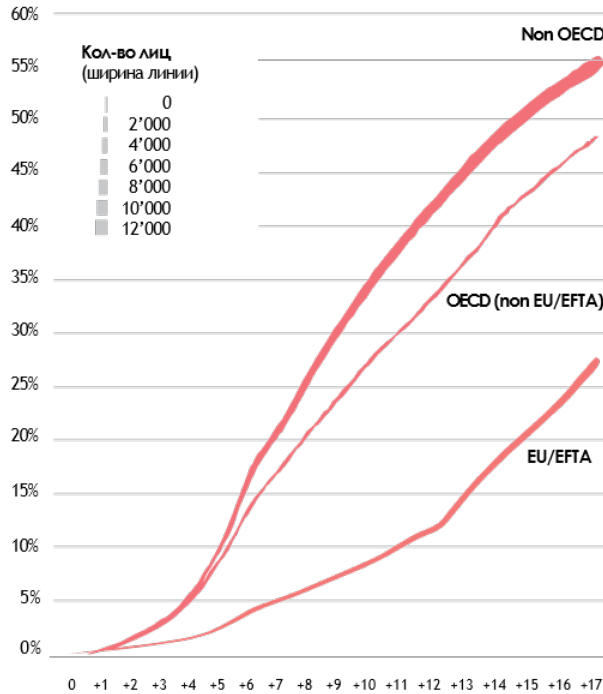
Для получения дополнительной информации:

http://nccr-onthemove.ch/DataManagement/Visualization/Embed/Naturalization_Rates.html

<https://indicators.nccr-onthemove.ch/where-in-switzerland-are-migrants-naturalized-most-often/>

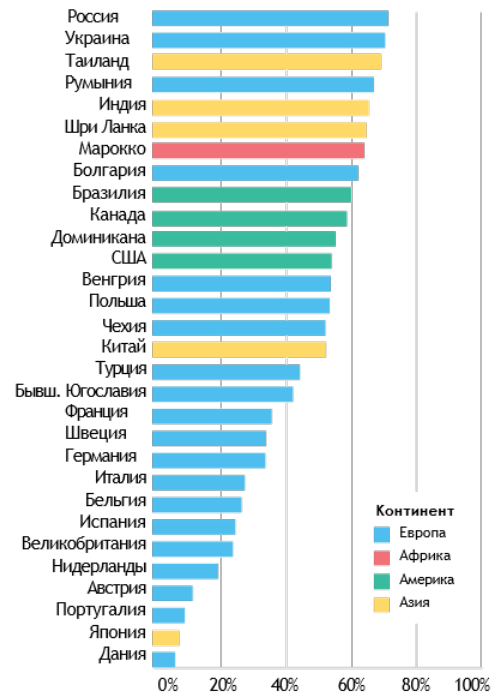
РИСУНОК 18: КУМУЛЯТИВНЫЕ КОЭФФИЦИЕНТЫ НАТУРАЛИЗАЦИИ ДЛЯ МИГРАНТОВ, ПРИБЫВШИХ В ШВЕЙЦАРИЮ В 1998 ГОДУ, ПО ГОДАМ С МОМЕНТА ПРИБЫТИЯ И СТРАНЕ ПРОИСХОЖДЕНИЯ

Натурализация с течением времени, кумулятивные коэффициенты (год прибытия 1998)



© nccr - on the move / Sources: SFSO / SEM (CAR, CEMIS, STATPOP)

Кумулятивный коэффициент натурализации через 17 лет (основные страны происхождения)



Основные страны происхождения выбраны в соответствии с количеством мигрантов, въехавших в Швейцарию в 1998 году

ТАБЛИЦА 8: ДОЛЯ ДОЛГОСРОЧНЫХ ВИДОВ НА ЖИТЕЛЬСТВО ОТ (1) ОБЩЕГО КОЛИЧЕСТВА ВИДОВ НА ЖИТЕЛЬСТВО И (2) РАЗРЕШЕНИЙ, ДЕЙСТВИТЕЛЬНЫХ БОЛЕЕ 5 ЛЕТ, ВЫБРАННЫЕ ГРАЖДАНСТВА, ИТАЛИЯ, ПО СОСТОЯНИЮ НА 1 ЯНВАРЯ 2018 Г.

Гражданство	1	2
	Долгосрочные резиденты в % от общего числа видов на жительство	Долгосрочные резиденты в % от числа видов на жительство, выданных до 2013 года
Марокко	70.3	84,1
Албания	71.6	84,6
Китай	56.0	63,7
Украина	72.3	86,1
Филиппины	62.2	72,5
Индия	59.0	78,5
Египет	64.6	87,2
Бангладеш	54.9	85,4
Молдавия	75.6	83,8
Пакистан	49.8	83,2
Итого	61 7	80 0

В таблице 9 иллюстрируется лонгитюдный подход из Италии. Он отслеживает две «когорты» новых разрешений для граждан не из ЕС, выданных в 2011 и 2012 годах. Для обоих годов когорты показаны абсолютные числа выданных разрешений, процент тех, кто все еще присутствовал в Италии в январе 2018 года, и, среди тех, кто все еще присутствовал, процент тех, кто стал долгосрочным резидентом. В таблице показаны различные модели пребывания и перехода к статусу долгосрочного резидента в Италии среди мигрантов, имеющих разное гражданство.

ТАБЛИЦА 9: НОВЫЕ ВИДЫ НА ЖИТЕЛЬСТВО, ВЫДАННЫЕ В 2011 И 2012 ГОДАХ, ПРОЦЕНТ МИГРАНТОВ, ВСЕ ЕЩЕ НАХОДЯЩИХСЯ В ИТАЛИИ НА НАЧАЛО 2018 ГОДА, И ПРОЦЕНТ ДОЛГОСРОЧНЫХ РЕЗИДЕНТОВ СРЕДИ ПРИСУТСТВУЮЩИХ, В РАЗБИВКЕ ПО СТРАНАМ ГРАЖДАНСТВА

Гражданство	Год получения первого разрешения 2011			Год получения первого разрешения 2012			Первые разрешения 2011 и 2012		
	Итого	% оставшихся на 1.01.2018	% постоянных жителей из числа оставшихся	Итого	% оставшихся на 1.01.2018	% постоянных жителей из числа оставшихся	Итого	% оставшихся на 1.01.2018	% постоянных жителей из числа оставшихся
Китай	26,850	47.8	21.1	25,153	47.7	11.2	52,003	47.7	16.3
Марокко	30,516	48.7	52.3	21,146	45.2	36.0	51,662	47.3	45.9
Албания	24,621	50.2	48.0	18,608	48.8	32.3	43,229	49.6	41.4
Индия	18,371	49.1	48.1	11,666	47.0	26.2	30,037	48.3	39.8
США	14,402	4.4	39.4	14,329	4.5	29.2	28,731	4.5	34.2
Тунис	19,454	19.6	40.4	6,364	33.8	36.4	25,818	23.1	39.0
Молдавия	16,463	38.4	51.4	8,779	35.8	40.3	25,242	37.5	47.7
Украина	15,595	53.1	54.3	8,596	50.8	42.3	24,191	52.3	50.2
Бангладеш	13,743	56.0	56.3	9,191	54.5	28.6	22,934	55.4	45.4

(продолжение Таблицы 9)

Гражданство	Год получения первого разрешения 2011			Год получения первого разрешения 2012			Первые разрешения 2011 и 2012		
	Итого	% оставшихся на 1.01.2018	% постоянных жителей из числа оставшихся	Итого	% оставшихся на 1.01.2018	% постоянных жителей из числа оставшихся	Итого	% оставшихся на 1.01.2018	% постоянных жителей из числа оставшихся
Египет	13,082	20.6	46.5	9,834	24.7	20.9	22,916	22.3	34.4
Филиппины	12,987	48.7	23.9	8,804	49.3	11.7	21,791	49.0	19.0
Пакистан	9,971	43.0	47.1	9,541	39.3	26.5	19,512	41.2	37.5
Нигерия	11,625	31.8	27.5	7,734	32.2	19.2	19,359	32.0	24.1
Шри-Ланка	8,951	52.4	33.0	6,721	51.6	16.0	15,672	52.1	25.8
Сенегал	8,167	47.8	48.1	6,177	41.4	28.2	14,344	45.0	40.2
Перу	8,661	43.4	40.5	5,174	41.6	24.8	13,835	42.7	34.8
Бразилия	7,154	25.8	60.5	5,626	25.0	57.7	12,780	25.4	59.2
Гана	6,375	33.4	32.8	4,528	36.7	20.7	10,903	34.8	27.5
Российская Федерация	5,493	37.8	61.0	4,506	35.3	52.3	9,999	36.7	57.2
Республика Македония	4,956	35.7	54.0	3,570	36.7	35.2	8,526	36.1	46.0
Другие страны	80,336	27.6	41.6	64,986	29.5	30.2	145,322	28.5	36.3
Итого	357,773	37.8	43.4	261,033	37.4	27.8	618,806	37.6	36.9

Показатель для приобретения гражданства, предоставленный Италией Евростату для Регламента (ЕС) № 862/2007, в настоящее время рассчитывается как соотношение между числом жителей, которые приобрели гражданство страны в течение календарного года, и общим числом иностранных резидентов в этой стране на начало года. Лонгитюдный подход может представлять особый интерес в этом случае, особенно для приобретений, которые получают по месту жительства.

Лонгитюдный подход может представлять особый интерес в этом контексте, поскольку не все те, кто получил гражданство, остались в Италии. Для расчета коэффициента натурализации необходимо учитывать население, подверженное риску натурализации. В Италии доступность источников данных, основанных на регистрах, позволяет извлекать данные о натурализации, увязывая событие натурализации с въездом в страну и выездом из нее.

ТАБЛИЦА 10: ИНОСТРАННОЕ НАСЕЛЕНИЕ И ИТАЛЬЯНЦЫ В РАЗБИВКЕ ПО ПОЛУЧЕНИЮ ГРАЖДАНСТВА, ПОДДАНСТВУ ИЛИ НАЦИОНАЛЬНОСТИ ПРОИСХОЖДЕНИЯ ПО СОСТОЯНИЮ НА 1 ЯНВАРЯ 2018 ГОДА, АБСОЛЮТНЫЕ ЗНАЧЕНИЯ И ПРОЦЕНТНОЕ СООТНОШЕНИЕ

Гражданство/ гражданство происхождения	Иностранное население (количество)	Иностран ное население (%)	Итальянцы при получении гражданства (количество)	Итальянцы при получении гражданства (%)
Румыния	1,190,091	23.1	77,046	5.7
Албания	440,465	8.6	169,644	12.6
Марокко	416,531	8.1	184,333	13.7
Китай	290,681	5.6	12,552	0.9
Украина	237,047	4.6	23,096	1.7
Филиппины	167,859	3.3	16,725	1.2
Индия	151,791	3.0	39,360	2.9
Бангладеш	131,967	2.6	22,394	1.7
Молдавия	131,814	2.6	18,654	1.4
Египет	119,513	2.3	24,125	1.8
Другие страны	1,866,681	36.3	757,332	56.3
Итого	5,144,440	100 0	1,345,261	100 0

Для расчета коэффициентов натурализации индивидуальные данные переписи 2011 года были привязаны к тем, кто получил гражданство после даты переписи. В июне 2019 года Национальное статистическое управление Италии (ISTAT) впервые распространило данные о численности "иностранцев по рождению" (таблица 10, выше). Статистика рассчитывает число лиц, относящихся к категории "иностранцы при рождении", по данным последней традиционной переписи населения (9 октября 2011 года), с учетом административных потоков (получение гражданства и демографических потоков).

Учитывая, что в Италии граждане, не входящие в ЕС, могут получить гражданство для постоянного проживания как минимум после 10 лет проживания, основное внимание уделяется тем, кто проживал на момент переписи и уже находился в Италии в 2006 году.

В период с 2011 по 2017 год 679 470 граждан, не являющихся гражданами ЕС, получили итальянское гражданство. Доля людей, получивших итальянское гражданство, значительно различается в зависимости от гражданства происхождения: около 30% в Марокко, Албании и Индии и менее 3% в Китае и Украине (таблица 11). Следует отметить, что законы и нормативные акты в упомянутых странах могут сильно отличаться, особенно в том, что касается сохранения гражданства страны происхождения, а также различий в геополитических условиях. Например, Китай и Украина не признают двойного гражданства, но то же самое верно и для Индии, граждане которой с большей вероятностью приобретут итальянское гражданство. В то же время граждане Румынии, являющиеся частью Европейского союза, могут больше не быть заинтересованы в получении гражданства, но граждане Польши, все еще являющейся страной ЕС, проявляют более высокую склонность.

ТАБЛИЦА 11: ПОКАЗАТЕЛИ СОХРАНЕНИЯ И СВЯЗАННОГО С ЭТИМ ПРИОБРЕТЕНИЯ ГРАЖДАНСТВА ГРАЖДАНАМИ СТРАН, НЕ ВХОДЯЩИХ В ЕС, ПРИБЫВШИМИ В ИТАЛИЮ ДО 2006 ГОДА И ЗАРЕГИСТРИРОВАННЫМИ В ХОДЕ ПЕРЕПИСИ НАСЕЛЕНИЯ 2011 ГОДА, В РАЗБИВКЕ ПО ГРАЖДАНСТВУ

Гражданство	Всего	% оставшихся на 1.01.2018	% новых граждан от общего числа	% новых граждан от оставшихся
Албания	301,606	92.7	25.2	27.2
Марокко	232,269	88.1	24.9	28.3
Китай	133,143	90.4	1.7	1.9
Украина	120,990	90.0	2.8	3.1
Филиппины	92,783	91.7	4.6	5.0
Молдавия	68,219	89.6	10.5	11.8
Перу	60,629	92.3	18.0	19.6
Эквадор	59,762	89.5	11.1	12.4
Индия	58,497	85.9	25.4	29.7
Тунис	50,789	83.4	14.5	17.4
Македония	48,417	79.2	19.9	25.1
Шри-Ланка	46,281	89.2	6.5	7.4
Сенегал	45,820	89.1	17.7	19.9
Бангладеш	43,151	80.0	26.2	32.9
Египет	37,827	74.1	11.4	15.4
Пакистан	34,600	84.5	27.7	32.8
Сербия	31,954	68.6	11.7	17.1
Нигерия	27,355	84.2	14.5	17.3
Гана	25,210	80.7	31.3	38.9
Косово	23,624	60.9	16.5	27.2
<i>Другие страны</i>	<i>255,707</i>	<i>82.6</i>	<i>13.6</i>	<i>16.5</i>
Итого	1,798,633	87.0	16.2	18.6

Примечание: гражданство приобретено между 2012 и 2017 годами

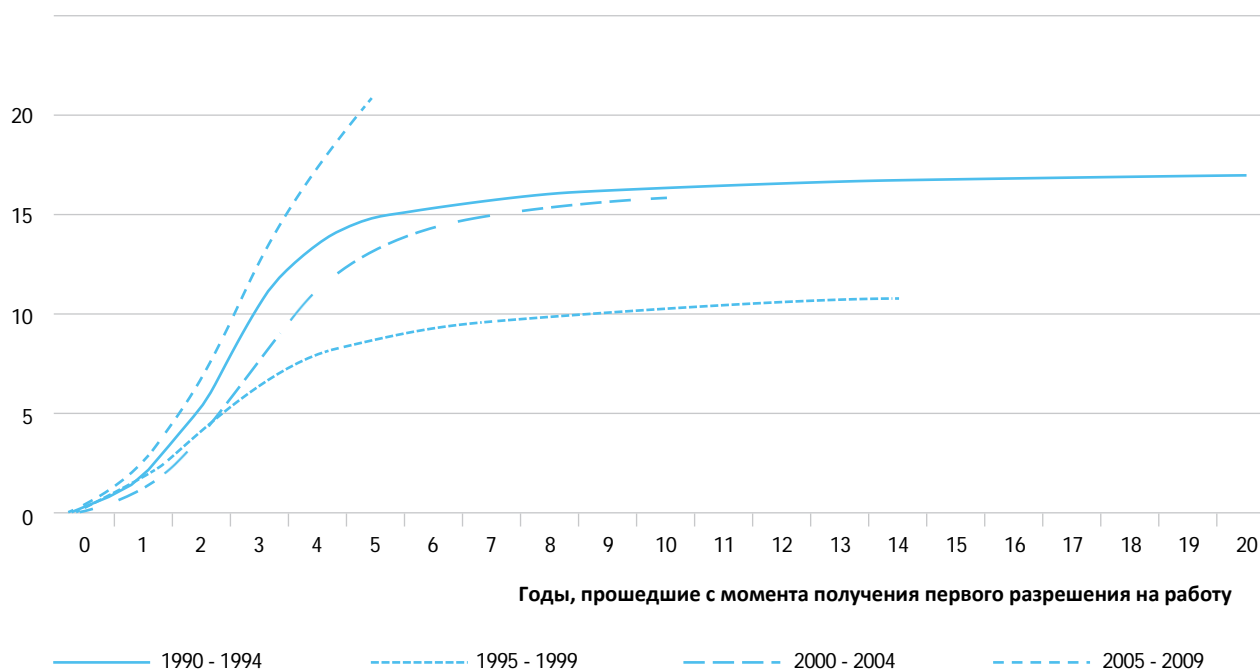
Источник: ISTAT, предварительные данные.

В Канаде Лонгитюдная база данных по иммиграции (IMDB) фиксирует даты всех разрешений временного проживания и даты, когда лица становятся постоянными жителями Канады. Это позволяет провести анализ времени до получения постоянного места жительства. На графике ниже показаны изменяющиеся во времени коэффициенты перехода временных иностранных работников к постоянному месту жительства на основе разных когорт.

На графике видно, что менее четверти временных иностранных работников, как правило, становятся постоянными жителями, но это изменилось с течением времени, и более поздние работники с большей вероятностью переходят. Когда это происходит, этот переход с большей вероятностью происходит в течение нескольких лет после их первого разрешения на работу в Канаде.

РИСУНОК 20: СОВОКУПНЫЕ ПОКАЗАТЕЛИ ПЕРЕЕЗДА ВРЕМЕННЫХ ИНОСТРАННЫХ РАБОТНИКОВ В КАНАДУ НА ПОСТОЯННОЕ МЕСТО ЖИТЕЛЬСТВА В РАЗБИВКЕ ПО ГОДАМ, ПРОШЕДШИМ С МОМЕНТА ПОЛУЧЕНИЯ ПЕРВОГО РАЗРЕШЕНИЯ НА РАБОТУ, И ПЕРИОДУ ДЕЙСТВИЯ ПЕРВОГО РАЗРЕШЕНИЯ НА РАБОТУ

Процент
переезда



Источник: Статистическое управление Канады, Лонгитюдная база данных по иммиграции, 2014: Lu, Y. & Hou, F. (2017) «Переезд временных иностранных работников на постоянное место жительства, 1990–2014» Серия исследовательских работ Аналитического отдела, каталог Статистического управления Канады № 11F0019M.

Канада недавно добавила дату получения гражданства в свою Лонгитюдную базу данных по иммиграции (IMDB). Эта информация доступна для иммигрантов, которым было предоставлено гражданство с 2005 года. На данный момент никакие показатели не публикуются. Эта новая информация предоставляет возможность охватить несколько показателей, включая:

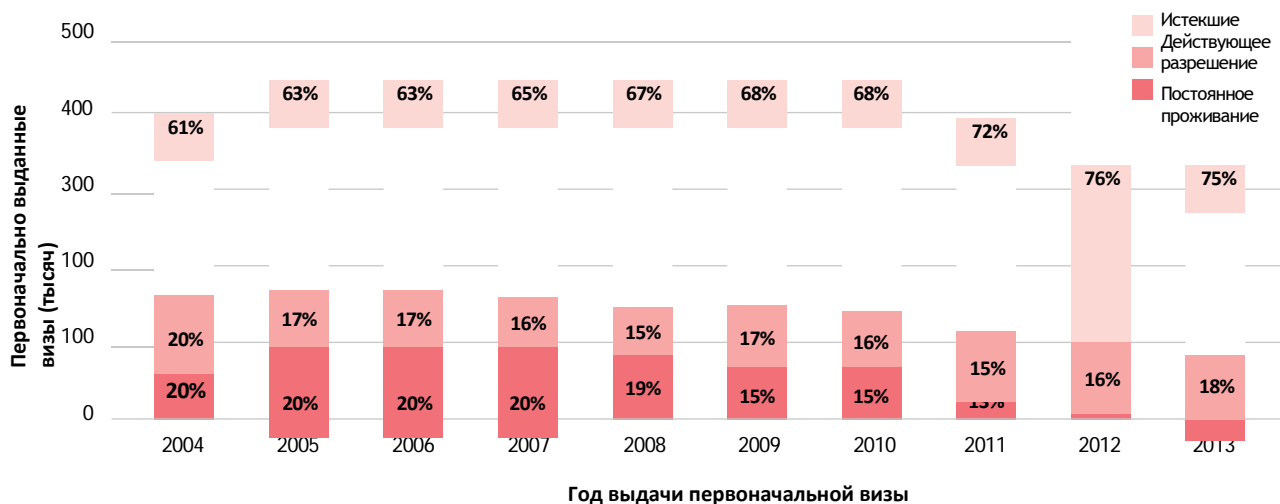
- Долю иммигрантов, которые стали гражданами путем натурализации с 2005 года, по годам с момента допуска
- Коэффициенты приобретения гражданства с 2005 года в соответствии с различными ключевыми характеристиками (например: знание официальных языков страны, страна исхода, пол)

Например, в статье, опубликованной в декабре 2018 года для выпуска IMDB, заработная плата иммигрантов, приобретших гражданство, сравнивалась с заработной платой иммигрантов, принятых в том же году, но не ставших гражданами к 2016 году. В 2016 году медианная заработная плата мужчин-иммигрантов, принятых в 2006 году, которые получили гражданство к 2016 году, составляла 40 500 долларов, в то время как для неграждан она составляла 31 200 долларов. Для женщин-иммигрантов медианная заработная плата составляла 28 100 долларов, если они получили гражданство, и 21 600 долларов, если нет.

В Соединенном Королевстве Министерство внутренних дел Великобритании публикует ежегодный отчет «Путь мигранта», в котором исследуются изменения в визовом статусе и статусе пребывания мигрантов. В отчете основное внимание уделяется поездкам, где перерыв между периодами пребывания составляет не более 12 месяцев. Последний отчет, охватывающий период до конца 2018 года, представляет данные двумя взаимодополняющими способами. Прямой анализ исследует иммиграционный статус мигрантов, которым были выданы визы в Великобританию в период с 2004 по 2016 год. Этот анализ исследует изменения иммиграционного статуса по категориям въезда и исключает посетителей. Последние выводы обсуждают тех, кому была выдана виза в 2013 году, с акцентом на их статус через 5 лет.

Обратный анализ исследует тех, чей въезд в Великобританию в конечном итоге привел к предоставлению им права на постоянное проживание (разрешение на постоянное проживание в Великобритании) по типу визы, с которой они изначально приехали.

РИСУНОК 21: ЧИСЛО МИГРАНТОВ, ПОЛУЧИВШИХ ПЕРВОНАЧАЛЬНУЮ ВИЗУ С 2004 ПО 2014 ГОД, ПО ИММИГРАЦИОННОМУ СТАТУСУ ЧЕРЕЗ 5 ЛЕТ В СОЕДИНЕННОМ КОРОЛЕВСТВЕ



Источник: Министерство внутренних дел Великобритании (2019) Путешествие мигранта: Отчет за 2018 год, Статистический бюллетень от 07/19.

Германия еще не публикует регулярно лонгитюдные показатели для статистики миграции. Существуют предварительные результаты мониторинга изменений правового статуса проживания после иммиграции и внутренней мобильности (Brückner, 2018).

Эти предварительные результаты основаны на иностранцах, которые впервые иммигрировали в Германию после 2006 года и непрерывно оставались резидентами до конца 2017 года. Статус вида на жительство анализируется отдельно для граждан государств-членов ЕС и граждан государств, не входящих в ЕС. Гражданам ЕС предоставляется свобода поселения ЕС и, следовательно, статус постоянного проживания с момента иммиграции. Однако предоставление статуса постоянного проживания является важным событием для всех граждан государств, не входящих в ЕС. В своем тематическом исследовании Brückner (2018) измеряет вероятность предоставления статуса постоянного проживания двумя показателями:

- процент иностранцев, (не) имеющих статус постоянного проживания после X лет
- среднее «время ожидания», то есть промежуток времени между иммиграцией и предоставлением статуса постоянного проживания (в годах)

Время ожидания является несколько проблематичным показателем, поскольку его значение основано на тех случаях, когда статус постоянного жителя предоставляется в течение контролируемого периода времени. В случае Германии максимальный срок наблюдения составил одиннадцать лет, и по истечении этих одиннадцати лет процентная доля лиц, не являющихся гражданами ЕС, имеющих статус постоянного жителя, составила всего 48,1%. Для получения объективных данных о времени ожидания может потребоваться период мониторинга в 25 лет и более. Доля иностранцев, имеющих статус постоянного жителя, может послужить хорошей заменой в предыдущие годы.

На основе дат выдачи видов на жительство можно отслеживать переход от обычного правового статуса к «статусу защиты» как лица, ищущего убежища, или беженца, и обратно, а также время, проведенное в каждом статусе.

Предварительные результаты показывают, что большинство, но не все лица, ищущие убежища, имеют гуманитарный статус защиты с момента иммиграции. 8,9 процента всех иностранцев в статусе защиты имели обычный статус проживания до этого — они, вероятно, подали заявление на убежище после истечения срока действия их вида на жительство, чтобы избежать депортации. Немецким властям потребовалось в среднем 0,8 года для обработки заявления на убежище, 0,9 года, если в убежище было отказано. Это время варьируется в зависимости от пола, возраста, семейного положения и гражданства. В среднем иностранцы, имевшие статус защиты в любые годы между 2007 и 2017 годами, провели 3,7 года

из 4,4 лет проживания в статусе защиты (84,1 процент). Из этих 3,7 лет они провели 0,8 года в ожидании решения по их заявлению, 2,4 года с предоставленной защитой и 0,5 года с отклоненной защитой.

4.1.1.3 Лонгитюдная статистика циркулярной миграции

- Среднее число пребываний на одного «циркулярного мигранта» (требует дальнейшего определения, например, для цензурированных наблюдений, периода времени, требуемой минимальной продолжительности пребывания и т.д.)*
- Средняя продолжительность пребывания на одного «циркулярного мигранта»*

* См. материалы ЕЭК ООН (2016) и Евростату (2017а и 2018).

339. Приведенные выше измерения относятся только к лицам в процессах циркулярной миграции, которые являются подмножеством всего набора мигрантов. Для правильной интерпретации этой статистики рекомендуется классификация по стране рождения и гражданству.

4.1.2 Социально-экономические результаты

340. Лонгитюдные показатели, связанные с социально-экономическими результатами, могут включать широкий спектр показателей, связанных с условиями мигрантов после прибытия. Это может включать такие темы, как знание официальных языков, занятость, жилье, доход и здоровье. Более подробная информация о показателях, связанных с социально-экономическими условиями мигрантов, содержится в публикации ЕЭК ООН «Измерение изменений в социально-экономических условиях мигрантов» (2015).

341. В отличие от показателей, связанных с миграцией, некоторые показатели, связанные с социально-экономическими результатами, могут не адекватно измеряться с использованием административных данных. В этих случаях для измерения этих результатов могут потребоваться обследования или административные прокси.

342. Следующие темы были определены как ключевые лонгитюдные показатели, представляющие интерес для стран-членов ЕЭК ООН, связанные с социально-экономическими результатами:

- Время до того, как мигранты начнут владеть официальными языками принимающей страны
- Время до того, как мигранты станут владельцами жилья
- Статус мигрантов на рынке труда с течением времени
- Время для получения высшего и среднего специального образования
- Доход мигрантов с течением времени
- Владение бизнесом / предпринимательство для мигрантов с течением времени
- Здоровье мигрантов с течением времени.

ТАБЛИЦА 12: КЛЮЧЕВЫЕ ТЕМЫ ПОКАЗАТЕЛЕЙ ДЛЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ РЕЗУЛЬТАТОВ С ПРИМЕРАМИ ИЗ СТРАН

Ключевые темы показателей	Бельгия	Канада	Норвегия	Швейцария	Великобритания
Время до того, как мигранты начнут владеть официальными языками страны		x		x	x
Время до того, как мигранты станут владельцами жилья					
Статус мигрантов на рынке труда с течением времени		x	x		x
Время для получения послешкольного образования		x			
Доход мигрантов с течением времени		x			
Владение бизнесом/ предпринимательство для		x			

мигрантов с течением времени

Здоровье мигрантов с течением времени

x

x

x

4.1.2.1 Время до овладения мигрантами официальными языками принимающей страны

343. Это тема, которая может применяться ко всем мигрантам независимо от возраста или других факторов. Некоторые мигранты прибывают с владением (или действительно были отобраны благодаря владению) официальным(и) языком(и) страны. Однако для других неспособность говорить на официальных языках страны может быть барьером во многих отношениях, таких как получение работы, посещение школы или даже установление социальных связей. Таким образом, это может быть важной темой показателей, связанных с обустройством мигрантов в новой стране.

344. Некоторые мигранты будут проходить языковую подготовку, в то время как другие могут изучать язык (языки) с погружением или получать пользу от общения со своими детьми, которые изучают этот язык (языки) в школе. В целом, ожидается, что чем дольше мигранты будут находиться в стране, тем больше вероятность того, что они смогут говорить на официальном языке (языках). Таким образом, эта тема хорошо подходит для лонгитюдной статистики.

345. Для измерения языковых навыков предлагаются следующие лонгитюдные показатели:

- Доля мигрантов, которые владеют официальными языками страны по прошествии определенного периода времени
- Среднее время до того, как мигранты начнут владеть официальными языками страны
- Доля мигрантов, которые говорят на официальном(ых) языке(ах) страны дома по прошествии определенного периода времени
- Среднее время до того, как мигранты начинают говорить на официальном(ых) языке(ах) страны дома

346. Эти показатели являются лонгитюдными в том смысле, что способность говорить на официальных языках страны может меняться с течением времени. Следует позаботиться о том, чтобы исключить мигрантов из стран с тем же официальным языком, что и страна назначения, в этих расчетах, предполагая, что они уже свободно владеют им по прибытии.

347. Однако, в отличие от некоторых других показателей, их может быть сложно измерить лонгитюдно. В ходе обследования могут быть заданы вопросы о способности респондента говорить, писать или понимать язык или о том, используют ли они язык дома, в школе, на работе или в другом месте. Однако административные записи, которые были бы более практичными для лонгитюдного анализа, могут быть менее точными.

Например, в Канаде подробные языковые вопросы включены в вопросник Переписи населения. Однако административные записи часто ограничены языком обслуживания (например, язык административной формы).

Еще одним ограничением, связанным с административными данными, является то, влияет ли на ответы сам процесс сбора административных данных. Служба иммиграции, беженцев и гражданства Канады (IRCC) запрашивает у потенциальных иммигрантов информацию о знании официальных языков в форме заявления. Хотя эти ответы не используются в процессе отбора, возможно, что люди считают, что положительные ответы повысят их шансы на поступление. С другой стороны, IRCC также собирает результаты языковых стандартизированных тестов для определенных потенциальных иммигрантов. Эти результаты могут обеспечить более объективную оценку знания официальных языков, чем самоотчет в ходе опроса.

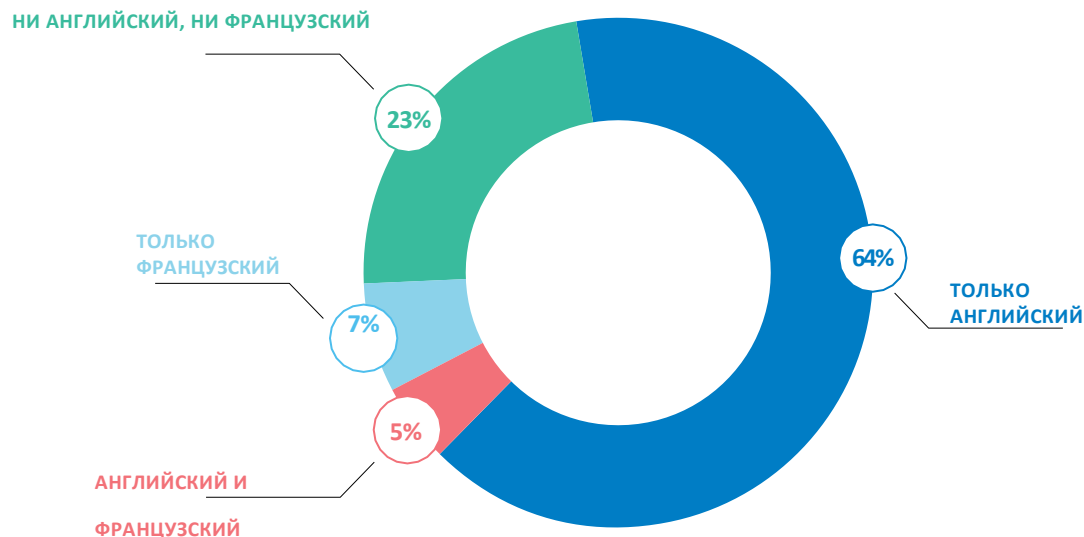
Хотя эти результаты собираются IRCC, когда иммигранты допускаются в Канаду, сложно сравнивать эти измерения с любыми последующими результатами обследований или других административных источников. Однако эти сравнения все же могут дать ценную информацию о владении языком.

Рисунок 22 дает ценную информацию о владении английским или французским языком среди недавних иммигрантов в Канаду. Например, среди тех иммигрантов, допущенных в период с 2006 по 2011 год, которые не могли говорить ни на английском, ни на французском по прибытии, более 75 процентов сообщили, что они могут говорить по крайней мере на одном официальном языке по состоянию на 10 мая 2011 года согласно Национальному обследованию домохозяйств 2011 года. Дальнейший анализ может

исследовать характерные различия между теми, кто овладел знанием английского и французского, по сравнению с теми, кто этого не сделал.

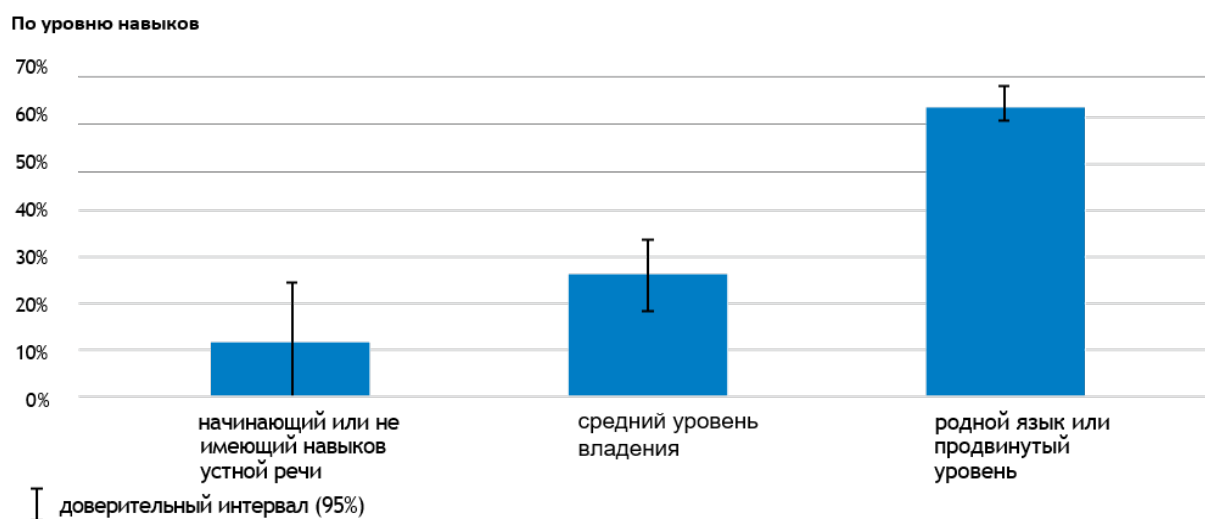
При оценке языковой интеграции может быть важна региональная специфика. Например, в Швейцарии доступны перекрестные результаты, чтобы дать снимок способности мигрантов говорить на официальном языке их кантона проживания.

РИСУНОК 22: ЗНАНИЕ ОФИЦИАЛЬНЫХ ЯЗЫКОВ В 2011 ГОДУ ДЛЯ НЕДАВНИХ (2006-2011) ИММИГРАНТОВ В КАНАДУ, НЕ ЗНАВШИХ АНГЛИЙСКИЙ ИЛИ ФРАНЦУЗСКИЙ НА МОМЕНТ ПРИЕМА



Источник: Статистическое управление Канады, Лонгитюдная база данных по иммиграции 2013 года, интегрированная с Национальным обследованием домохозяйств 2011 года.

РИСУНОК 23: УСТНЫЕ ЯЗЫКОВЫЕ НАВЫКИ ИММИГРАНТОВ В ШВЕЙЦАРИЮ НА ПРИМЕРЕ ОДНОГО ИЗ ОФИЦИАЛЬНЫХ ЯЗЫКОВ В КАНТОНЕ ПРОЖИВАНИЯ, 2014 ГОД



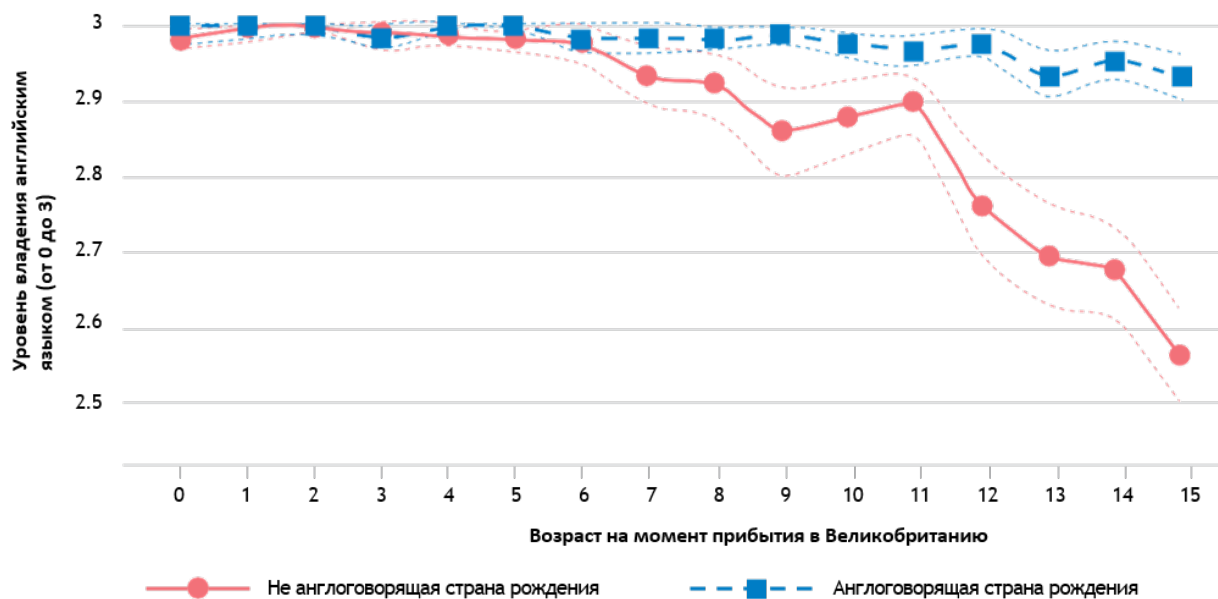
Источник: Федеральное статистическое управление Швейцарии, Швейцарское обследование рабочей силы (SLFS), модуль миграции, 2014.

В Соединенном Королевстве, хотя и не созданное специально для изучения миграции, Лонгитюдное исследование для Англии и Уэльса (с эквивалентами для Шотландии и Северной Ирландии) являются ключевыми лонгитюдными источниками данных, на примере которых можно наблюдать социально-экономические результаты с течением времени. Лонгитюдное исследование для Англии и Уэльса связывает данные переписи и административные данные с течением времени. Два вопроса об основном языке и знании английского были впервые включены в перепись 2011 года. Это позволило проанализировать влияние языковых навыков на социально-экономические результаты, такие как образование, здоровье и рождаемость.

В качестве примера, Лонгитюдное исследование Англии и Уэльса использовалось для определения причинного эффекта навыков английского языка на образование, здоровье и рождаемость иммигрантов в Англии и Уэльсе (Aoki and Santiago, 2015). Для начала исследование исследует знание английского языка среди людей в возрасте 20 лет и старше, которые были детьми-иммигрантами. Те, кто мигрировал из англоязычных стран, как правило, хорошо владели английским, поскольку они подвергались воздействию языка до их прибытия. Для тех, кто мигрировал из неанглоязычных стран, те, кто мигрировал в возрасте до восьми лет, как правило, владели английским так же хорошо, как и те, кто мигрировал из англоязычных стран. Однако показатели владения языком расходятся в возрасте девяти лет; чем позже они мигрировали, тем хуже было их знание английского языка, в среднем. Это проиллюстрировано на Рисунке 24.

Используя регрессионный анализ, в рамках этого исследования также удалось обнаружить, что лучшие навыки английского языка снижают вероятность того, что взрослые не имеют квалификации, и повышают вероятность получения ученых степеней. Влияние языка на рождаемость значительно. Лучшие знания английского значительно откладывают возраст при рождении первого ребенка для женщин, снижают вероятность того, что они станут матерями-подростками, и уменьшают рождаемость. Не было обнаружено связи между знанием английского языка и здоровьем детей или самооценкой здоровья взрослых.

РИСУНОК 24: ВОЗРАСТ ПРИБЫТИЯ И ЗНАНИЕ АНГЛИЙСКОГО ЯЗЫКА ВЗРОСЛЫМИ В 2011 ГОДУ ДЛЯ ДЕТЕЙ-МИГРАНТОВ В АНГЛИЮ И УЭЛЬС



Примечания: на рисунке показана средняя порядковая мера знания английского языка, где 3, 2, 1 и 0 соответствуют говорению «очень хорошо», «хорошо», «не очень хорошо» и «совсем не говорю» соответственно. Знание английского языка скорректировано по регрессии для возраста. Два набора внешних линий соответствуют 95-процентным доверительным интервалам. Выборка соответствует детям-иммигрантам в возрасте от 20 до 60 лет на момент переписи 2011 года.

Источник: Aoki & Santiago (2015) Education, Health and Fertility of UK migrants: The Role of English Language Skills, IZA Discussion Paper Series No.498.

В Германии нет лонгитюдных источников данных, из которых можно было бы получить социально-экономические результаты. Вместо этого для этой цели используются ретроспективные данные в популяционных обследованиях. Для иммигрантов микроперепись и немецкая версия Европейского обследования рабочей силы предоставляют данные о стране рождения и годе прибытия. Образовательный уровень, участие в рабочей силе, полученный доход и риск бедности регулярно анализируются для мигрантов по стране рождения, возрасту при первом прибытии и продолжительности проживания.

Германия предоставляет показатели, относящиеся к вышеуказанным тематическим областям для

- мигрантов первого и второго поколения, и
- с гражданством Германии и без него, соответственно.

Данные временных рядов по этим показателям доступны за годы с 2005 по 2018 год.

https://www.destatis.de/EN/Themes/Society-Environment/Population/Migration-Integration/_node.html#sprg265538

4.1.2.2 Время до приобретения собственного жилья мигрантами

348. Получение права собственности на жилье может свидетельствовать о том, что человек обосновывается в своей новой стране или районе проживания. Безусловно, есть отдельные лица и семьи, которые, как правило, предпочитают не иметь собственного жилья и арендовать его независимо от статуса мигранта. Однако владение домом является признаком того, что отдельные лица или семьи располагают финансовыми возможностями для приобретения недвижимости и что они готовы инвестировать в конкретное место для проживания.

349. Для измерения этой темы предлагаются следующие продольные показатели:

- Доля мигрантов, которые владеют своим жильем по прошествии определенного периода времени
- Среднее время до того, как мигранты впервые станут владельцами жилья в новой стране проживания
- Статус занятости жилья (напр., аренда vs. владение) мигрантов с течением времени
- Доля мигрантов, проживающих в субсидируемом или финансируемом государством жилье по прошествии определенного периода времени
- Владение по типам жилой недвижимости (напр., квартира, таунхаус и т.д.) по прошествии определенного периода времени

350. Владение жильем — это концепция, которая может измеряться административно или через регистр населения, в дополнение к обследованию. Административные записи должны быть четкими в отношении права собственности на жилище. В качестве альтернативы, административные записи о получателях государственного субсидируемого или финансируемого жилья или арендаторах были бы важны для определенных показателей.

351. Одним из элементов, который может быть сложным при использовании административных записей, является то, что право собственности на недвижимость может быть указано только на одно физическое лицо. В этом показателе также необходимо учитывать проживающих членов семьи. Этот последний компонент является сложным в долгосрочном плане, поскольку состав семейства может изменяться с течением времени (см. раздел 4.2.3).

4.1.2.3 Статус мигрантов на рынке труда с течением времени

352. Поиск оплачиваемой работы в новой стране проживания считается важным признаком экономической интеграции, что связано с другими показателями, перечисленными в этой главе. Чтобы получить работу, люди часто должны соответствовать определенным требованиям. Например, на многих работах от работника требуется умение эффективно общаться на официальном языке (языках) страны. Возможно, также требовались признанные документы, такие как высшее образование. Выход на рынок труда открывает новые возможности, а социальная сеть мигранта расширяется за счет коллег, клиентов и других людей.

353. Принадлежность мигрантов в рабочей силе можно считать важным показателем с национальной экономической точки зрения. Если мигранты допускаются специально для заполнения определенных пробелов на рынке труда, важно учитывать, работают ли те мигранты (и в какой профессии или отрасли) фактически после прибытия.

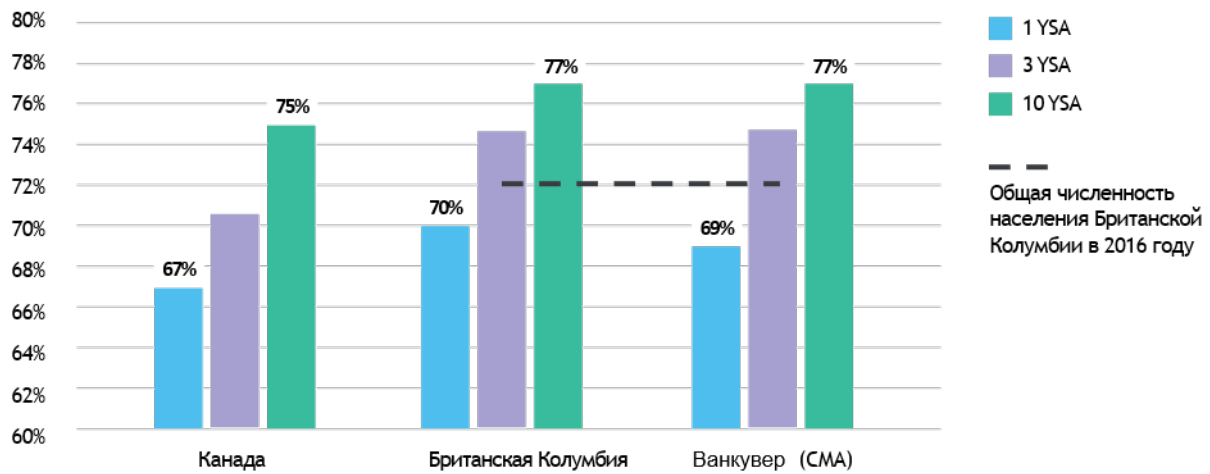
354. Многие мигранты прибывают в страну своего нового проживания, уже имея работу. Действительно, занятость по определенной профессии могла быть мотивирующим фактором миграции (как с точки зрения мигранта, так и с точки зрения страны). С другой стороны, многим мигрантам, включая беженцев, возможно, придется искать работу по прибытии.
355. Для измерения статуса на рынке труда предлагаются следующие лонгитюдные показатели:
- Доля мигрантов, которые трудоустроены по прошествии определенного периода времени
 - Среднее время до того, как мигранты впервые трудоустраиваются в новой стране проживания
 - Занятость на полную либо неполную ставку у мигрантов с течением времени
 - Статус на рынке труда мигрантов с течением времени
356. Подобно набору показателей, связанных с языками, измерение статуса на рынке труда может быть трудным для оценки с использованием административных данных.
357. В частности, хотя административные данные могут дать информацию о степени занятости конкретного человека, провести различие между безработными и теми, кто не входит в состав рабочей силы, более сложно.
358. С точки зрения лонгитюдной перспективы, в обследованиях, как правило, используется заранее определенный короткий период, например неделя, для определения статуса рабочей силы. С другой стороны, административные данные могут основываться на других отчетных периодах. Если эти учетные периоды охватывают более длительные периоды времени, может оказаться невозможным указать точное значение времени, прошедшего до того, как человек получил работу.
359. Одной из наилучших практик может быть игнорирование значений, если отчетный период совпадает с периодом въезда мигранта в страну. В противном случае значения за этот период будут искажены и будут зависеть от того, когда прибыли определенные подгруппы мигрантов. Например, если исходный период является годовым, то мигранты, прибывающие в январе, будут иметь более высокую вероятность трудоустройства в этом году, чем те, кто прибывает в декабре, независимо от других факторов.
360. Еще одним фактором, который следует учитывать при расчете этих показателей, является знаменатель, поскольку он должен включать только тех лиц, которые могут быть трудоустроены или безработными (например, минимальный возраст).

В Канаде административные налоговые документы могут служить основой для определения занятости. В частности, база данных по долгосрочным иммиграционным отношениям (IMDB) содержит годовые переменные для заработной платы, отрасли(ей) занятости и т.д. Однако эти переменные являются ежегодными сводками. Они могут быть использованы для определения того, сколько иммигрантов, подающих налоговые декларации, имели какую-либо работу в данном отчетном году (например, заработная плата > 0 за отчетный год), но не указано, была ли эта работа в январе или декабре. Подробная информация о отработанных часах или неделях также недоступна. Другие переменные, такие как наличие страховки при трудоустройстве, являются слабыми показателями безработицы и отражают другие события, такие как отпуск по уходу за ребенком.

Тем не менее, IMDB предоставляет показатели, касающиеся занятости иммигрантов за годы, прошедшие с момента их въезда в Канаду, несмотря на эти ограничения.

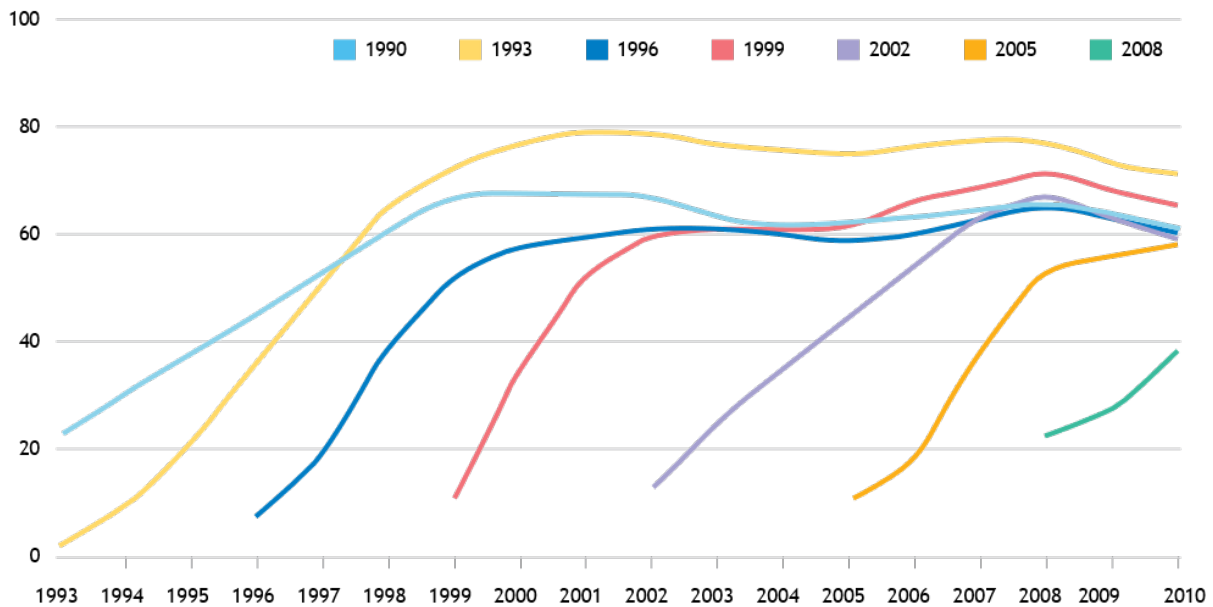
На следующей диаграмме, основанной на данных из базы данных по долгосрочным иммиграционным вопросам (IMDB), представлены ежегодные показатели трудоустройства беженцев, допущенных в Канаду в 2006 году, через 1 год, 3 года и 10 лет после въезда. Данные сравниваются по Канаде, провинции Британская Колумбия и столичному району Ванкувер. Данные о заболеваемости сравниваются с аналогичными расчетами для населения в целом, проживавшего в Британской Колумбии в 2016 году (что эквивалентно данным о беженцах за 2006 год, через 10 лет после приема). Анализ можно было бы упростить с учетом других факторов, таких как возраст и пол.

РИСУНОК 25: НАЛИЧИЕ ЗАНЯТОСТИ ДЛЯ БЕЖЕНЦЕВ-НАЛОГОПЛАТЕЛЬЩИКОВ, ДОПУЩЕННЫХ В КАНАДУ В 2006 ГОДУ, ПО ВЫБРАННОМУ МЕСТУ ПРОЖИВАНИЯ И ГОДАМ С МОМЕНТА ПРИЕМА



Источник: Статистическое управление Канады, Лонгитюдная база данных по иммиграции 2016 года, таблица 43-10-0014; Канадское обследование доходов 2017 года, таблица 11-10-0239-04.

РИСУНОК 26: ДОЛЯ БЕЖЕНЦЕВ В ВОЗРАСТЕ ОТ 17 ДО 36 ЛЕТ, КОТОРЫЕ ТРУДОУСТРОЕНЫ, ПО КОГОРТЕ И ГОДУ ДОХОДА, НОРВЕГИЯ, ПРОЦЕНТ



Источник: Статистическое управление Норвегии, Статистика населения и доходов.

В Соединенном Королевстве, хотя оно и не было специально создано для изучения миграции, лонгитюдное исследование Управления национальной статистики использовалось для изучения различий в уровне безработицы между мигрантами первого и второго поколений в Англии и Уэльсе. Авторы (Zuccotti and Platt, 2016) отмечают, что исследование показало, что потомки мигрантов демонстрируют заметные улучшения по сравнению со своими родителями в первом поколении с точки зрения неблагоприятных условий на рынке труда и неравенства. Они используют данные, которые увязывают результаты пяти последовательных переписей населения (с 1971 по 2011 год), для оценки семейных ресурсов отдельных лиц, проживавших в детстве с родителями-иммигрантами в Англии и Уэльсе, и увязывают их с результатами их трудоустройства во взрослом возрасте. Кроме того, сопоставляя

информацию о том, где проживали семьи, они могут определить степень депривации района, чтобы выяснить, может ли район, в котором выросли дети второго поколения, повлиять на их социальную мобильность.

Результаты показывают, что пакистанские и карибские мужчины примерно на 6 процентов чаще становятся безработными, чем британские мужчины. Когда они принимают во внимание социальное происхождение и депривацию района происхождения, этот «штраф» сокращается вдвое до 30 процентов. Они также показывают, что наличие университетского диплома значительно снижает штраф как для мужских, так и для женских групп.

Норвегия также опубликовала лонгитюдные анализы занятости мигрантов. На диаграмме ниже используется когортный подход для сравнения траекторий занятости между 7 когортами беженцев в возрасте 17-36 лет.

4.1.2.4 Время до получения послешкольного образования

361. Для мигрантов (особенно детей и молодых людей) получение послешкольного (высшего и среднего специального) образования в новой стране проживания является важным процессом, поскольку оно дает им возможность учиться на официальных языках страны, установить социальную сеть и, в конечном счете, предоставляет им признанные полномочия для поиска работы после окончания учебы. Время на завершение образования может быть затронуто различными факторами, такими как перерывы в их предыдущем образовании, связанные с событием миграции, доступность получения высшего образования по сравнению с поиском работы, способность учиться на официальных языках страны и т.д. В результате некоторым мигрантам может потребоваться больше времени, чем другим, чтобы завершить свое образование.

362. Для этой темы важно учитывать «время» как с возрастной точки зрения, так и с точки зрения времени с момента прибытия.

363. Послешкольное образование в новой стране проживания также может быть важным для мигрантов с предыдущим образованием. В частности, если их предыдущее образование не признается работодателями или если им необходимо получить переподготовку, чтобы открыть новые возможности на рынке труда. В некоторых случаях знание категории допуска заявителя (программа, на основании которой мигрант был допущен) может информировать о минимальных полномочиях, которыми обладает лицо, если никакая другая мера не существует в административных файлах миграции.

364. Некоторые мигранты уже установят планы поступления в университет или высшее учебное заведение до прибытия в страну. Подобно теме показателя об участии в рабочей силе, посещение университета или высшего учебного заведения может быть мотивирующим фактором самого события миграции.

365. Для измерения времени для получения послешкольного образования предлагаются следующие показатели:

- Доля мигрантов, которые получили аттестат, степень или диплом о послешкольном образовании в своей новой стране проживания по прошествии определенного периода времени по возрасту
- Посещение школы/университета/высшего учебного заведения мигрантами с течением времени по возрасту

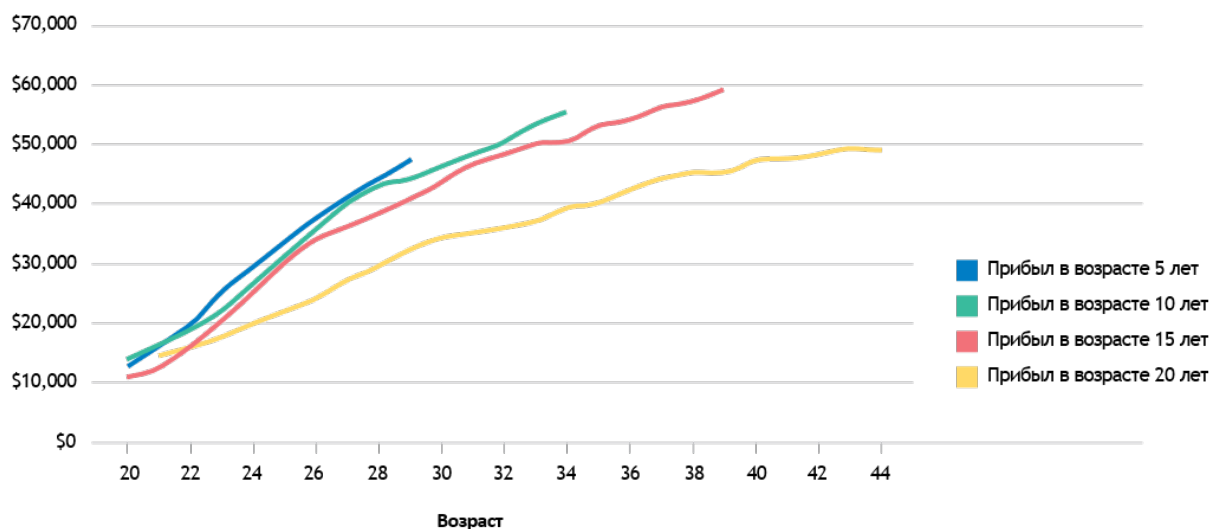
366. Если данные о зачислении из всех послешкольных учебных заведений страны собираются централизованно, эти показатели могут быть относительно простыми в производстве из статистического регистра или из интегрированных административных записей. В качестве альтернативы, если подмножества учреждений не охватываются государственными административными данными (например, частные учреждения), это предоставит ограничение для рассмотрения соответствующих показателей.

367. Это более сложная тема, поскольку она может включать по крайней мере две временные переменные, и объединение обеих в одном показателе может быть чрезмерным. Есть несколько возможных подходов к рассмотрению для упрощения показателей. Для молодых мигрантов (например, тех, кто прибывает в возрасте 18 лет или младше) возраст может быть более важным фактором, чем годы с момента прибытия. Таким образом, показатели можно упростить до:

- Посещение послешкольного учебного заведения по одному году возраста
- Получение послешкольного образования по одному году возраста

368. В этом случае все еще важно контролировать год прибытия. Однако вместо того, чтобы быть частью самого показателя, это могла бы быть переменная фильтра (например, рассчитать показатели для тех, кто прибыл в данном году). Вышеуказанные показатели также могут быть воспроизведены для общего населения, позволяя сравнивать послешкольное образование между молодежью-мигрантами и немигрантами.
369. Другая временная переменная, представляющая интерес, особенно при изучении результатов молодых мигрантов, — это возраст при прибытии. Чем моложе были мигранты при прибытии, тем большую долю обучения они получили в принимающей стране. Это может облегчить поступление в послешкольные учебные заведения в той же стране.
370. Например, на диаграмме ниже показаны траектории доходов иммигрантов, которые были допущены (т.е. легализованы) в Канаду в 1990 году по выбранным возрастам при допуске. На диаграмме показано, что в пределах одной и той же когорты дети-иммигранты, которые прибывают в более молодом возрасте, как правило, зарабатывают больше, когда вырастают, по сравнению с детьми-иммигрантами, которые прибывают в возрасте, близком к взрослому.

РИСУНОК 27: СРЕДНЯЯ ЗАРАБОТНАЯ ПЛАТА И ОКЛАДЫ ПО ВОЗРАСТУ И ВОЗРАСТУ ПРИ ДОПУСКЕ В КАНАДУ ДЛЯ ИММИГРАНТОВ-НАЛОГОПЛАТЕЛЬЩИКОВ, ДОПУЩЕННЫХ В 1990 ГОДУ (ДОЛЛАРЫ 2014 ГОДА)



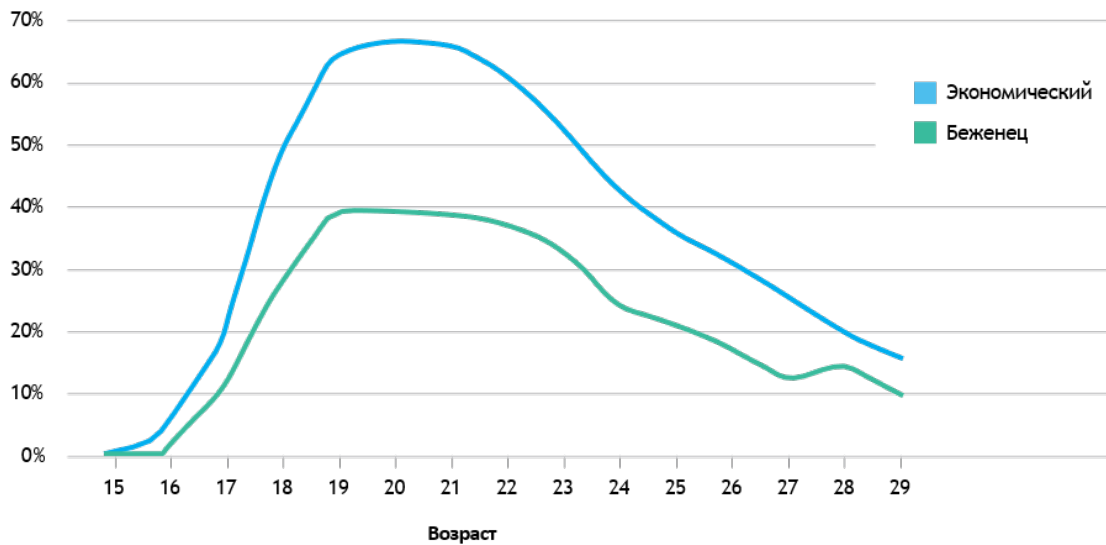
Источник: Статистическое управление Канады, Лонгитюдная база данных по иммиграции, 2014.

Для взрослых мигрантов годы с момента прибытия могут представлять больший интерес. В этом случае может следует ограничить анализ теми, кто был в возрасте не менее 25 лет (или другого возраста после того, как послешкольное образование обычно завершается) при прибытии. Цель этого ограничения — смягчить эффект возраста на соответствующий анализ.

В Канаде база данных по долгосрочным иммиграционным вопросам (IMDB) не содержит каких-либо административных данных, полученных непосредственно от высших учебных заведений, но налоговые органы декларируют суммы за обучение в высших учебных заведениях. Это может быть использовано в качестве подтверждения участия в программе после окончания средней школы. Однако в нем не указывается тип высшего или среднего специального образования (например, докторантура или профессиональная сертификация) и не указываются лица, завершившие обучение.

Следующая диаграмма, использующая IMDB, показывает участие в послешкольном образовании по одному году возраста для иммигрантов, допущенных в Канаду в 1999 году в возрасте 14 лет. Результаты, разделенные по категориям поступления, показывают, что дети беженцев с меньшей вероятностью будут учиться в высших учебных заведениях, чем дети экономических иммигрантов.

РИСУНОК 28: ПРОЦЕНТ ПОДАВШИХ ДЕКЛАРАЦИЮ НА УЧЕБНЫЙ ВЫЧЕТ ПО ВОЗРАСТУ И КАТЕГОРИИ ДОПУСКА ДЛЯ ИММИГРАНТОВ-НАЛОГОПЛАТЕЛЬЩИКОВ, КОТОРЫЕ БЫЛИ ДОПУЩЕНЫ В КАНАДУ В 1999 ГОДУ В ВОЗРАСТЕ 14 ЛЕТ

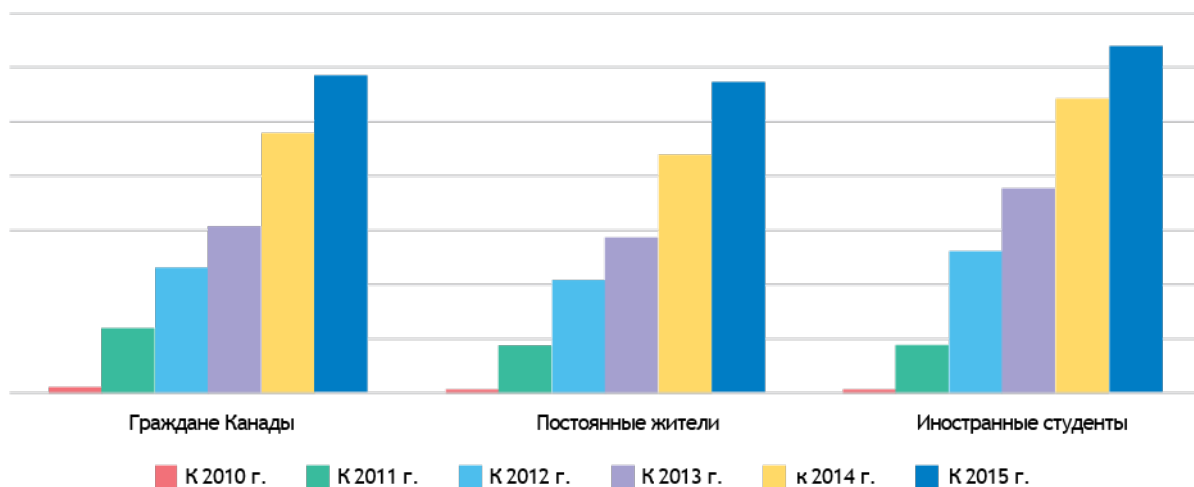


Источник: Статистическое управление Канады, Лонгитюдная база данных по иммиграции, 2014.

Рассматривая взрослых иммигрантов, исследование Беженцы и канадское послешкольное образование: Характеристики и экономические результаты в сравнении использовало тот же шаблон для изучения характеристик и экономических результатов тех, кто получал послешкольное образование в Канаде после допуска, по сравнению с теми, кто его не получал.

Другой канадский анализ (Frenette, Lu and Chan 2019) использовал административные данные о послешкольном обучении на основе Информационной системы о студентах послешкольных учебных заведений (PSIS). Статус иммигранта фиксировался административно учреждениями во время зачисления. На диаграмме ниже показаны коэффициенты выпуска по статусу иммиграции после определенного количества лет с момента поступления на программу.

РИСУНОК 29: КОЭФФИЦИЕНТЫ ВЫПУСКА ПО СТАТУСУ ИММИГРАЦИИ НОВЫХ ПОСТУПИВШИХ НА ПРОГРАММЫ В КАНАДЕ В 2010 ГОДУ



Источник: Статистическое управление Канады, Информационная система о студентах послешкольных учебных заведений, 2015.

4.1.2.5 Доход мигрантов по источникам с течением времени

371. Доход является мерой экономических результатов. При разбивке по ключевым источникам дохода можно понять, какие лица заняты, самозаняты или получают социальную или государственную помощь. Хотя сам по себе он не может идентифицировать тип работы отдельного лица, общая величина их дохода от занятости дает более широкое указание, сочетающее почасовую заработную плату с общим количеством отработанных часов.
372. Многие мигранты могут сталкиваться с препятствиями на рынке труда, как обсуждалось выше. В целом, неудивительно наблюдать, что мигранты зарабатывают меньше, чем основное население, когда они впервые прибывают. Но сохраняется ли это с увеличением количества лет в стране? Есть ли некоторые группы, которые с большей вероятностью превзойдут средний или медианный доход общего населения?
373. Также есть интерес к пониманию фискального воздействия иммиграции на принимающие страны. Это можно изучать различными способами. Например, можно рассмотреть общие налоги, уплаченные иммигрантами, по сравнению с получением государственной помощи.
374. Для измерения дохода предлагаются следующие продольные показатели:
- Средний и медианный общий доход по прошествии определенного периода времени
 - Средний и медианный доход от занятости по прошествии определенного периода времени
 - Доля мигрантов с доходом от занятости по прошествии определенного периода времени
 - Доля мигрантов с доходом от самозанятости или предпринимательства по прошествии определенного периода времени
 - Доля мигрантов, получающих социальную или государственную помощь по прошествии определенного периода времени
 - Время до того, как средний или медианный общий доход для мигрантов сравняется со средним или медианным общим доходом общего населения
 - Время до того, как средний или медианный доход от занятости для мигрантов сравняется со средним или медианным общим доходом общего населения
 - Доля мигрантов с низким доходом по прошествии определенного периода времени
375. Благодаря налогообложению, эти показатели должны быть простыми в производстве с использованием либо регистров, либо интегрированных административных записей. Конечно, любой доход, не задекларированный (например, возможно, иностранные источники дохода), может не включаться.
376. С долгосрочной точки зрения, для определения этих показателей важно учитывать отчетный период. Значения дохода будут относиться к отчетному периоду в целом и могут быть изменены, если доход был получен только за часть отчетного периода. Например, небольшое значение дохода от трудовой деятельности за отчетный период может отражать низкооплачиваемую работу или высокооплачиваемую работу, которая была активной только в течение части отчетного периода. Это особенно важно учитывать при изучении мигрантов, поскольку они, возможно, даже не находились в стране в течение всего отчетного периода.
377. Одной из наилучших практик может быть игнорирование значений, если отчетный период совпадает с периодом въезда мигранта в страну. В противном случае значения за этот период будут искажены и будут зависеть от того, когда прибыли определенные подгруппы мигрантов. Например, если исходный период является годовым, то мигранты, прибывающие в январе, будут иметь в среднем более высокие доходы в этом году, чем те, кто прибывает в декабре, даже если они будут работать на низкооплачиваемой работе.
378. Как и в случае с показателями рынка труда, важно учитывать население в рамках этих показателей. Показатели занятости или самозанятости должны быть ограничены населением, которое может работать. Другие показатели можно было бы рассматривать в большей степени на уровне семьи. Например, понятие «низкий доход», как правило, определяется на уровне домохозяйства, и тогда это будет общей характеристикой для всех членов домохозяйства. Социальная или государственная помощь также может рассматриваться как понятие на уровне домохозяйства или семьи. Однако

важно проводить различие между поддержкой семей с низким доходом и индивидуальной поддержкой (например, временная помощь по безработице).

379. Другое важное соображение, особенно для лонгитюдного анализа, — это эффект изменений стоимости жизни с течением времени. Может быть лучшей практикой корректировать денежные значения так, чтобы значения дохода основывались на фиксированной стоимости валюты.

В Канаде Лонгитюдная база данных по иммиграции (IMDB) производит несколько регулярных лонгитюдных показателей, связанных с доходом иммигрантов. Они включают процент с, средний и медианный доход по источникам дохода, годы с момента допуска и другие характеристики. В качестве лучшей практики, выходные данные IMDB исключают результаты в нулевом году с момента допуска. Результаты доступны по одному году когорты иммигрантов в Канаду с когортами, начиная с 1980 года, и результатами, начиная с 1982 года. Источники дохода, выбранные для стандартных таблиц, последовательно определены за этот период времени.

Особый интерес представляют следующие параметры оценки:

- Средний и медианный общий доход
- Средний и медианный доход от заработной платы, окладов и комиссионных
- Средний и медианный доход от самозанятости
- Доля получающих пособия по социальному обеспечению
- Доля получающих заработную плату, оклады и комиссионные

Они доступны по году допуска, годам с момента допуска и другим характеристикам. Сравнения могут быть сделаны с общим населением с использованием сопоставимых источников данных, таких как Лонгитюдный административный банк данных.

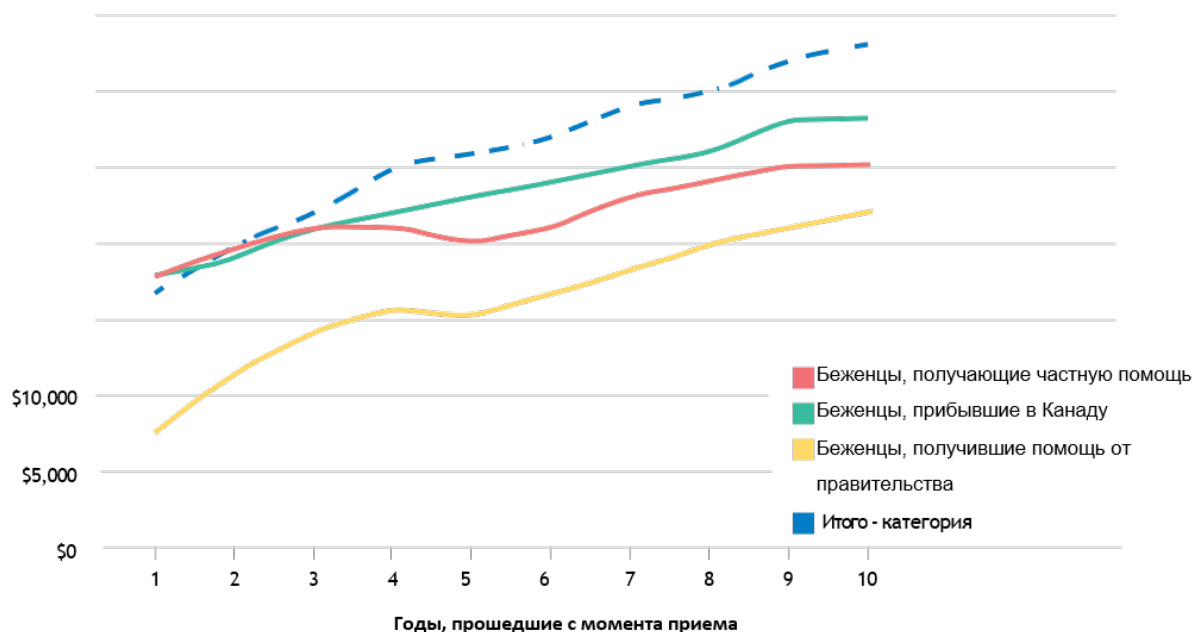
Ниже приведены стандартные таблицы, опубликованные на веб-сайте Статистического управления Канады и обновляемые ежегодно:

[Доход иммигранта в разбивке по годам поступления и годам, прошедшим с момента поступления, Канада и провинции](#)

[Доход иммигранта в разбивке по годам поступления и категориям иммигрантов, Канада и провинции](#)

Ниже приведена диаграмма, показывающая траектории средней заработной платы и окладов для иммигрантов, допущенных в Канаду в 2004 году по выбранным категориям допуска.

РИСУНОК 30: СРЕДНЯЯ ЗАРАБОТНАЯ ПЛАТА И ОКЛАДЫ ИММИГРАНТОВ, ДОПУЩЕННЫХ В КАНАДУ В 2004 ГОДУ, ПО ВЫБРАННЫМ КАТЕГОРИЯМ ДОПУСКА И ГОДАМ С МОМЕНТА ДОПУСКА (ДОЛЛАРЫ 2014 ГОДА)



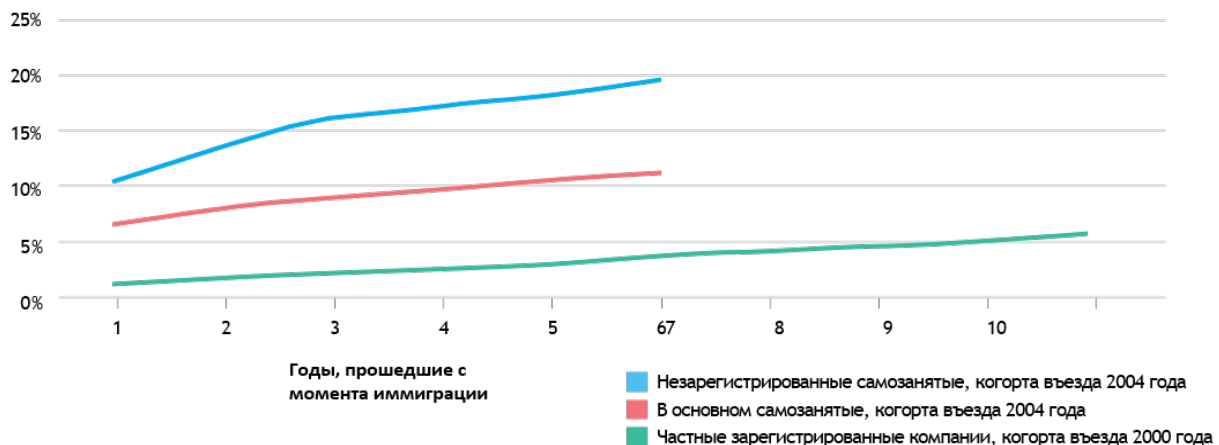
Источник: Статистическое управление Канады, Лонгитюдная база данных по иммиграции, 2014.

4.1.2.6 Владение бизнесом/ предпринимательство среди мигрантов с течением времени

380. Для некоторых мигрантов трудности выхода на рынок труда могут привести к самозанятости; в то время как для других предпринимательство или владение бизнесом могут быть связаны с мотивирующим фактором самого события миграции.
381. Безусловно, не все лица, не говоря уже о мигрантах, стремятся быть самозанятыми или владеть бизнесом. Однако самозанятость или предпринимательство могут быть важным экономическим результатом для мигрантов. Это не только потенциально обходит некоторые барьеры на рынке труда, но и может создавать рабочие места для других.
382. В рамках этой темы предлагаются следующие лонгитюдные показатели:
- Доля мигрантов, которые являются самозанятыми или владеют бизнесом по прошествии определенного периода времени
 - Среднее время до того, как мигранты становятся самозанятыми или владеют бизнесом
 - Число лиц, трудоустроенных мигрантами по прошествии определенного периода времени
383. Как и в случае с показателями статуса рабочей силы, данные о налогообложении могут служить косвенным показателем предпринимательской деятельности, отражая долю мигрантов, получающих доход от самостоятельной занятости или бизнеса. В противном случае можно использовать административные данные о владении бизнесом.
384. Соединение мигрантов-предпринимателей с информацией об их сотрудниках может быть более сложным. Однако, если существуют административные записи, связывающие работников с фирмами и владельцев с фирмами, этот анализ также возможен.

В Канаде канадская база данных динамики отношений работодателя и работника (CEEDD) представляет собой среду для установления связей, основанную на 12 административных источниках, включая базу данных о долгосрочных иммиграционных отношениях (IMDB), и включает характеристики отдельных лиц, фирм и владельцев бизнеса. Поскольку это среда связывания, пользователи должны интегрировать различные компоненты в соответствии со своими аналитическими требованиями, используя анонимизированные уникальные идентификаторы для каждого файла. CEEDD позволяет анализировать фирмы и отдельных лиц с течением времени, отвечая на вопросы, например, о первоначальном размещении фирм и росте доходов, владении бизнесом иммигрантами и траекториях развития фирм, принадлежащих иммигрантам.

РИСУНОК 31: ПРОЦЕНТ ИММИГРАНТОВ-НАЛОГОПЛАТЕЛЬЩИКОВ, КОТОРЫЕ ЯВЛЯЮТСЯ ВЛАДЕЛЬЦАМИ БИЗНЕСА В КАНАДЕ, ПО ГОДАМ С МОМЕНТА ИММИГРАЦИИ, КОГОРТЫ ВЪЕЗДА 2000 И 2004 ГОДОВ



Источник: Статистическое управление Канады, Канадская база данных о динамике работодателей и работников, 2014; Green et al (2016) Business Ownership and Employment in Immigrant-owned Firms in Canada. Economic Insights Catalogue no. 11-626-X -- No. 057.

4.1.2.7 Здоровье мигрантов с течением времени

385. Поддержание и улучшение здоровья среди недавних мигрантов в новой стране является важным признаком успешной интеграции, со связями с другими темами показателей, перечисленными в этой главе, такими как участие в рабочей силе. Без хорошего здоровья может быть трудно участвовать в жизни сообщества, получить работу или удержаться на ней. На общественном уровне, если мигранты прибывают с серьезными проблемами здоровья или быстро их развивают, это может стать нагрузкой на систему здравоохранения. Кроме того, мигранты, выросшие в совершенно разных условиях здоровья, могут принести заразные и передаваемые заболевания, которые могут представлять большой риск для общественного здоровья. Это некоторые из причин, по которым некоторые страны проводят медицинский скрининг и оценку мигрантов перед тем, как разрешить им въехать или поселиться в принимающей стране. Это также может помочь объяснить эффект здорового иммигранта, который можно наблюдать во многих странах, принимающих мигрантов, отражая наблюдение, что большинство мигрантов, за исключением, возможно, беженцев, как правило, здоровы на момент прибытия.

386. На здоровье влияют не только биологические особенности человека, но и множество социальных и экологических факторов, определенных как социальные детерминанты здоровья, которые включают образ жизни, окружающую среду и организацию медицинского обслуживания, а также факторы доступа. Это особенно важно для мигрантов, у которых могут быть факторы риска для здоровья, а также образ жизни, сильно отличающийся от образа жизни других людей в принимающей стране, что затрудняет адаптацию и интеграцию. Мигранты также могут не знать, как ориентироваться в новой системе здравоохранения, поскольку они, как правило, прибывают из стран с другими условиями здравоохранения и языковыми требованиями. Мигранты, которые испытывают трудности с поиском качественной работы или могут испытывать стресс и проблемы с психическим здоровьем. Все эти факторы могут привести к потере первоначального преимущества в плане здоровья.

387. В рамках этой темы предлагаются следующие продольные показатели:

388. Доля мигрантов, которые зарегистрированы у врача в системе здравоохранения принимающей страны по прошествии определенного периода времени (или время до снятия с регистрации)

389. Доля мигрантов, которые были госпитализированы в течение определенного количества лет с момента прибытия в принимающую страну проживания, в целом или по различным выбранным состояниям здоровья, например, хроническим или инфекционным по своей природе (или время до первой госпитализации)

390. Доля мигрантов, умерших с момента прибытия в принимающую страну проживания от различных заболеваний (по прошествии определенного периода времени) (или время до смерти)

391. В прошлом исследования результатов миграции полагались на обследования, с такими показателями, как самооценка состояния здоровья мигрантов, самооценка неудовлетворенных потребностей, самооценка доступа к медицинскому работнику и т.д. Однако истинные продольные показатели потребовали бы от нас отслеживать мигрантов с течением времени, чтобы наблюдать изменения в здоровье. В последнее время, с возрастающей трудностью получения ответов на обследования, вместе с прогрессом методологии связывания данных и снижением стоимости вычислений, связывание административных данных и данных обследований предоставляет исследователям альтернативный источник информации.

В Бельгии исследователи использовали индивидуально связанные данные Переписи, национального регистра населения и свидетельств о смерти за периоды 1991-1997 и 2001-2008 годов для изучения различий в смертности по конкретным причинам между мигрантским и рожденным в Бельгии населением с течением времени (Vanthomme and Vandenneede, 2019). Связывая переписи Бельгии 1991 и 2001 годов с регистровыми данными за 1991-1997 и 2001-2008 годы, позволяя наблюдать всю эмиграцию и смертность в эти периоды. Впоследствии индивидуальная связь со свидетельствами о смерти предоставила смертность по конкретным причинам.

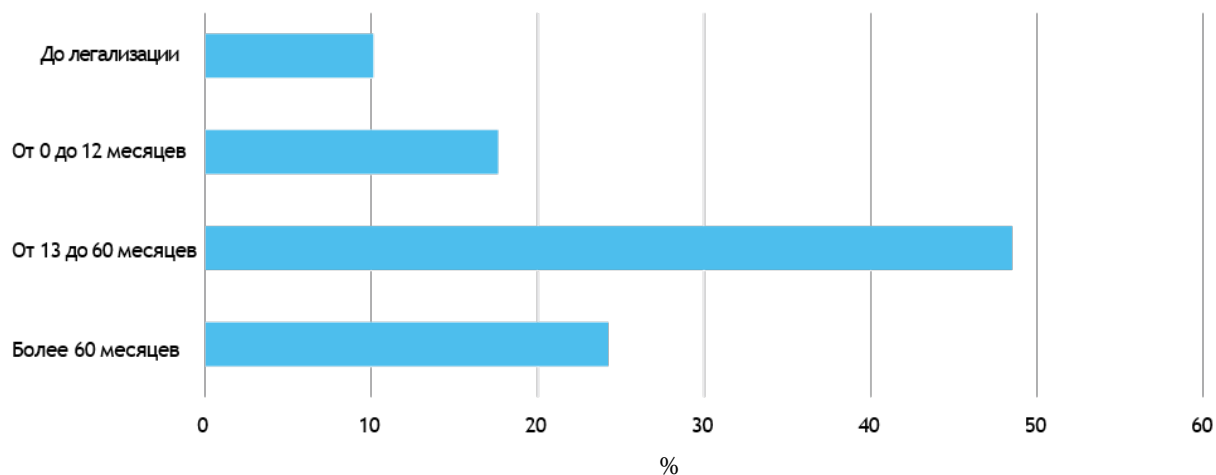
В Канаде некоторые провинции требуют, чтобы лица регистрировались у врачей общей практики. Одно ограничение провинциально администрируемых данных о здоровье заключается в том, что данные доступны только в определенных юрисдикциях, а не централизованно на национальном уровне. База данных регистрации в Онтарио, провинция с наибольшим количеством иммигрантов, была связана с Лонгитюдной базой данных по иммиграции (IMDB) и использовалась исследователями для понимания дифференциальной регистрации. Это особенно важно в ситуациях, когда существует нехватка врачей общей практики, как в Онтарио (Globerman et al, 2018). Возможность зарегистрироваться под наблюдением врача общей практики может быть важным первым шагом к успешной интеграции в систему здравоохранения.

В Канаде записи о выписке из больницы, собираемые провинциальными и территориальными министерствами здравоохранения, передаются и централизуются в Канадском институте информации о здравоохранении в базе данных под названием База данных о выписке из стационара (DAD). Они впоследствии передаются в Статистическое управление Канады для целей связывания записей и исследований. Этот файл больницы был связан с IMDB для изучения моделей и уровней госпитализации, либо в целом, либо для конкретных хронических или инфекционных заболеваний. Эти данные также позволяют оценить изменение моделей госпитализации по всем основным причинам, по полу, регионам происхождения и категории допуска иммигрантов. Признательно, что необходимо находиться в серьезном состоянии здоровья, чтобы быть госпитализированным; таким образом, госпитализация является показателем относительно серьезных состояний здоровья.

В качестве другого примера, IMDB-DAD был использован для изучения распределения случаев первой госпитализации в связи с туберкулезом (ТБ) по времени с момента прибытия в Канаду. ТУБЕРКУЛЕЗ – это очень заразное заболевание, имеющее серьезные последствия для общественного здравоохранения. С уменьшением заболеваемости туберкулезом среди местного населения туберкулез стал более распространенным среди иммигрантов, но редким среди большинства местных жителей, за исключением коренных народов (Vachon et al., 2018).

Анализ данных, связанных с IMDB и DAD, показал, что около 10% этих госпитализаций, связанных с туберкулезом, произошли до начала иммиграционного периода, что означает, что эти иммигранты были госпитализированы, когда они находились в Канаде в качестве временных жителей, таких как иностранные студенты или временные иностранные рабочие (Ng et al., 2018).

РИСУНОК 32: ПРОЦЕНТ ГОСПИТАЛИЗАЦИЙ, СВЯЗАННЫХ С ТУБЕРКУЛЕЗОМ, В КАНАДЕ ПО ВРЕМЕНИ С МОМЕНТА ДОПУСКА В КАНАДУ, ИММИГРАНТЫ, ДОПУЩЕННЫЕ В 2000-2013 ГОДАХ



Источник: Статистическое управление Канады, связанная база данных IMDB-DAD 2000-2013 (Квебек и территории исключены).

Ограничения для этих данных также заключаются в дифференциальной доступности по юрисдикциям. Хотя наборы данных о госпитализации могут быть недоступны для всех провинций и территорий, набор данных все же может использоваться для анализов, касающихся остальной части страны. Хотя информация о посещениях врача или результатах лабораторных анализов может дать указание на ухудшение здоровья за некоторое время до необходимости госпитализации, в Канаде такие наборы данных находятся только на провинциальном уровне.

В Канаде IMDB также была связана с Канадской базой данных статистики естественного движения населения о смертях (CVSD), которая фиксирует все смерти, происходящие в Канаде. Таким образом, другим релевантным показателем здоровья иммигрантов, помимо госпитализации, было бы время до смерти. Эти данные позволяют изучить эволюцию с течением времени в моделях смертности по всем основным причинам смерти, по полу, региону происхождения и другим измерениям, связанным с мигрантами. Также можно оценить изменение с течением времени, сравнивая различия в смертности мигрантов в 2000-х годах с 2010-ми годами.

Одна проблема с этим анализом заключается в том, что, поскольку больше иммигрантов прибывает в принимающую страну с относительно хорошим здоровьем, время до смерти может потребовать длительного периода наблюдения до исхода. В этом смысле время до первой госпитализации может быть дополнительным показателем для обнаружения ухудшения здоровья с течением времени.

Общим фактором, который следует учитывать при расчете этих показателей, является то, что при расчете показателя в знаменателе следует включать только тех лиц, которые входят в сферу охвата, то есть население, подверженное риску возникновения данного результата. Это проблема эмиграции среди мигрантов, которые могут решить вернуться на родину или переехать в другую страну. Эту проблему смещения числителя к знаменателю может быть трудно решить, особенно в странах, где нет записей о выезде. В Канаде административные налоговые файлы интегрированы в базу данных Лонгитюдной иммиграции, и образцы налоговых деклараций могут использоваться в качестве подтверждения места жительства иммигрантов в Канаде.

Также, анализ данных, основанный на связанных данных с участием IMDB, очевидно, не будет иметь соответствующей информации от контрольной группы местного населения. Можно использовать другие связи данных для предоставления сравнения, если набор данных существует (Ng et al, 2016).

В Соединенном Королевстве Лонгитюдное исследование Управления национальной статистики использовалось для изучения опыта смертности мигрантов первого и второго поколения в Англии и Уэльсе. Часто обнаруживается, что мигранты имеют низкую смертность по сравнению с принимающим населением в западных странах. Считается, что «эффект здорового мигранта» исчезает через поколения. Уоллес (2016) анализирует опыт смертности членов Лонгитюдного исследования в возрасте от 20 до 65 лет в 1991 году, в течение периода исследования 1991-2012 годов. После некоторых корректировок на отсутствующую информацию это дает выборку из 555 111 человек с 47 907 смертями. Мигранты и их потомки определяются через комбинацию информации об этнической принадлежности и стране рождения, которая записывается в переписи. Анализ истории событий используется, чтобы обнаружить, что мигранты всех этнических групп имеют более низкую смертность, но есть вариация в опыте смертности потомков мигрантов.

Потомки чернокожих карибских, пакистанских и бангладешских мигрантов имеют высокую смертность, которую нельзя объяснить их социальным происхождением. Потомки индийцев, чернокожих африканцев и других имеют низкую смертность, которая изначально маскируется их социальным происхождением. Автор предполагает, что потеря эффекта отбора для миграции может помочь объяснить повышенную смертность среди потомков. Также рассматриваются межпоколенческие различия в культурных нормах, отношениях и поведении в образе жизни и диете (включая курение), наряду с развитием в раннем возрасте мигрантов и потомков, с точки зрения профилей заболеваний и социального положения. Авторы приходят к выводу, что эффект здорового мигранта сохраняется в некоторых этнических группах после первого поколения миграции, но не в других. Обнаружено, что социальное положение оказывает гораздо большее влияние на смертность потомков.

4.1.3 Семейная миграция

392. Показатели семейной миграции обсуждались, но было определено, что они выходят за рамки данного конкретного отчета.

393. Темы, которые следует рассмотреть в связи с семейной миграцией, включают:

- Среднее время между прибытием в страну первого члена семьи и прибытием последнего члена семьи (т.е., временной лаг в воссоединении семьи)
- Экономические результаты семей мигрантов по прошествии определенного времени пребывания в стране

394. Семейный анализ в лонгитюдном контексте дополнительно сложен. Отдельные лица являются постоянной единицей анализа, которую нельзя разделить или объединить. Семьи, с другой стороны, не являются постоянными единицами анализа, поскольку членство в семье может меняться с течением времени. Это делает определение семьи лонгитюдно сложным и не без ограничений.

Например, семья мигрантов может быть определена на основе одного члена семьи. Но этот подход теряет информацию, когда другие члены покидают семью или новые члены присоединяются.

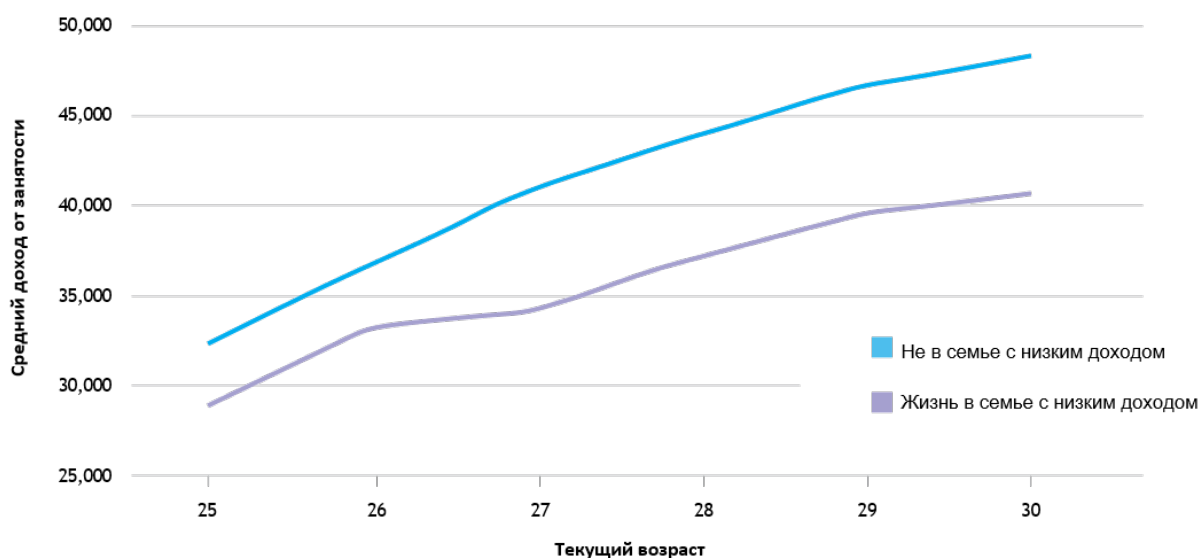
395. Эта изменяющаяся природа семей влияет на любые лонгитюдные показатели, ориентированные на семьи как единицу анализа. Например, доход семьи с течением времени напрямую связан с доходом членов семьи. Если в семью присоединяется кормилец, доход семьи возрастает в результате этого добавления.
396. Определение семей может быть сложным в перекрестном разрезе и при ссылке на общее население, но для мигрантов это может быть еще более сложным. Если семьи определяются теми, кто живет в одном домохозяйстве, как рассматриваются семьи в процессе воссоединения? Становятся ли они семьей только после воссоединения?

Из членов целевой группы Канада производит лонгитюдные данные о мигрантах с семейными переменными как часть Лонгитюдной базы данных по иммиграции (IMDB). IMDB включает индивидуальные и семейные идентификаторы как из административных данных по иммиграции, так и из ежегодных налоговых файлов. Эта база данных позволяет анализировать состав семьи с течением времени со связями с индивидуальными иммигрантами и их результатами в Канаде. В 2019 году Статистическое управление Канады добавило детский модуль в IMDB, позволяющий анализировать семейные экономические результаты детей-иммигрантов. Этот модуль затем связан с экономическими результатами во взрослом возрасте этих иммигрантов, позволяя анализировать межпоколенческую экономическую мобильность.

Следующий анализ был произведен с использованием детского модуля IMDB. Один родитель был выбран в качестве основного родителя. Если основной заявитель был подателем декларации, он выбирался в качестве основного родителя, если нет, то супруг-податель декларации или родитель-неиммигрант. Анализ показывает, что экономические результаты детей-иммигрантов варьируются в зависимости от их социально-экономической ситуации в детстве.

Дети, допущенные до 5 лет и живущие в семье с низким доходом через два года после их допуска, имели более низкие доходы, чем те, кто не жил (Рисунок 33). Разрыв в доходах увеличивался по мере взросления детей-иммигрантов.

РИСУНОК 33: СРЕДНИЙ ТРУДОВОЙ ДОХОД ИММИГРАНТОВ, ПРИБЫВШИХ В КАНАДУ В ВОЗРАСТЕ ДО 5 ЛЕТ В ПЕРИОД С 1980 ПО 1984 ГОД, В РАЗБИВКЕ ПО ТЕКУЩЕМУ ВОЗРАСТУ И МАЛООБЕСПЕЧЕННОМУ СТАТУСУ СЕМЬИ ЧЕРЕЗ 2 ГОДА ПОСЛЕ ВЪЕЗДА (ДОЛЛАРЫ ЗА 2016 ГОД)



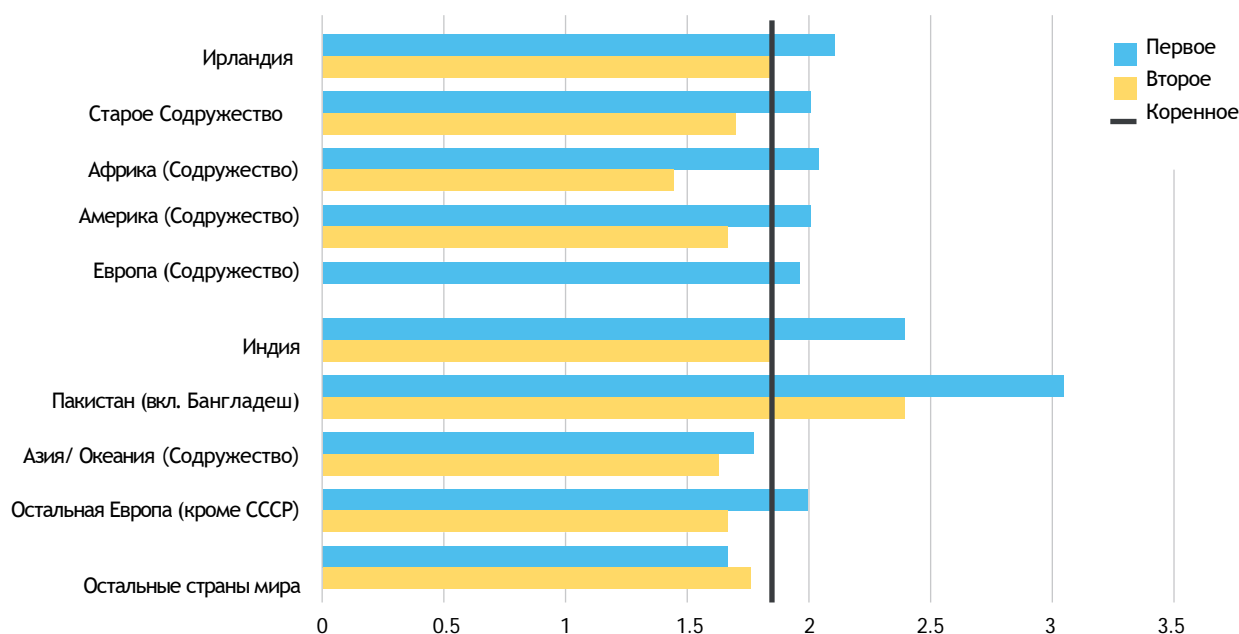
Статус низкого дохода: идентифицирует отдельных лиц и семьи с низким доходом в соответствии с мерой низкого дохода (LIM). LIM составляет половину скорректированного медианного дохода семьи (с учетом размера семьи).

Источник: Статистическое управление Канады, Лонгитюдная база данных по иммиграции (IMDB), 2016.

В Соединенном Королевстве было проведено лонгитюдное исследование, проведенное Управлением национальной статистики, для изучения моделей формирования семей международными мигрантами (Уилсон и Куха, 2017). В этом исследовании использовалась взаимосвязанная информация о составе детского населения в сообществе, чтобы оценить его взаимосвязь с завершённой рождаемостью различных групп мигрантов. В исследовании основное внимание уделяется женщинам иностранного происхождения, которые прибыли в Англию и Уэльс в возрасте до 16 лет, а также женщинам во втором поколении, которые родились в Англии и Уэльсе, но у которых есть по крайней мере один родитель иностранного происхождения. В анализе используется многоуровневое моделирование для учета характеристик сообщества, с конкретными переменными на уровне сообщества и на индивидуальном уровне в качестве контрольных переменных. Использование многоуровневой модели подразумевает сравнение между иммигрантами и их потомками, с одной стороны, и коренными жителями-предками, с другой.

Результаты показывают более высокую среднюю завершённую фертильность (рождаемость) у женщин, которые были детьми-мигрантами, по сравнению с коренными жителями, хотя в зависимости от страны происхождения существуют значительные различия. Уровень завершённой фертильности в пакистанской и бангладешской группах примерно на 50% выше, чем у местных жителей. Этот показатель остается на 30% выше для второго поколения, в то время как полная рождаемость для второго поколения из других групп ниже 1,85 (для коренных жителей). Многоуровневая модель показывает, что дети-мигранты с меньшей вероятностью будут иметь такую же рождаемость, как местные жители, если они будут менее подвержены влиянию местной культуры. Авторы предполагают, что сегрегация по месту жительства отчасти объясняет высокую рождаемость женщин во втором поколении из Пакистана и Бангладеш, что может быть одной из причин, по которой рождаемость некоторых выходцев из Южной Азии в Англии и Уэльсе может оставаться «культурно укоренившейся».

РИСУНОК 34: ПОЛНАЯ РОЖДАЕМОСТЬ РАЗЛИЧНЫХ ГРУПП МИГРАНТОВ ПО ПРОИСХОЖДЕНИЮ И ПОКОЛЕНИЮ ПО СРАВНЕНИЮ С КОРЕННЫМИ ЖИТЕЛЯМИ.



Примечание: на рисунке показана средняя полная рождаемость для мигрантов (по поколению и происхождению) относительно среднего совокупного числа рождений для коренных жителей (равного 1,85).

Источник: Wilson & Kuha (2017), 'Residential segregation and the fertility of immigrants and their descendants', *Population, Space and Place*. 2018; 24:e2098.

Семейный анализ мигрантов — это сложная и комплексная тема. Более тщательный обзор был бы необходим, чтобы полностью учесть различные проблемы и подходы, которые можно было бы использовать для изучения семей мигрантов с течением времени.

4.2 ПЕРЕДОВАЯ ПРАКТИКА РАСПРОСТРАНЕНИЯ ЛОНГИТЮДНОЙ СТАТИСТИКИ

4.2.1 Различные пользователи статистики

397. Существует широкий круг пользователей, которые полагаются на статистику. Их потребности часто расходятся, как и их навыки использования разнообразной доступной статистики (PARIS21 & Statistics Norway, 2009). Поэтому, чтобы обеспечить эффективную коммуникацию, необходимо определить целевую аудиторию (Eurostat, 2014 & 2017b). Широкое различие можно провести между двумя основными группами пользователей:
398. Специалисты, которые обладают статистическими знаниями экспертного уровня и способны обрабатывать и интерпретировать детальные наборы данных (например, статистики, академические круги, специализированные журналисты и аналитики политики);
399. Неспециалисты, т.е. люди без статистических знаний или с ограниченными статистическими знаниями (например, граждане или широкая публика)
400. Эти две широкие группы пользователей имеют разные потребности и модели поведения в отношении использования статистики, которые следует тщательно учитывать, чтобы выбрать соответствующие каналы коммуникации (Eurostat, 2017b). Однако существует общее согласие с тем, что, независимо от целевой аудитории, Интернет должен стать основным вектором коммуникации и распространения. Многие пользователи более квалифицированы или чувствуют себя более уверенно с компьютерами или Интернетом, чем с печатными публикациями. Информация, опубликованная в Интернете, как правило, является самой последней. Интернет — это также самое быстрое средство получения информации, в частности информации из разных источников. Пользователи могут легко и быстро загружать данные или статистику и использовать их непосредственно на своем персональном компьютере. Этот канал коммуникации также позволяет национальным статистическим офисам расширять распространение своих данных без генерации высоких дополнительных затрат.
401. С одной стороны, основной интерес специалистов — получать точную и детальную информацию, включая точные методологические определения, представление статистических тенденций, гармонизированные временные ряды, детальные метаданные и т.д. (Eurostat, 2017b). В этом отношении и учитывая их знакомство с темой и собственную способность анализировать связанные данные, их потребности в основном для информации, а не для объяснительных техник, необходимых для декодирования статистического сообщения неспециалистами.
402. И наоборот, граждане, как правило, не обладают специальными знаниями в области статистики. Таким образом, их главной потребностью является высокий уровень доступа к содержанию показателей. Статистика часто воспринимается как чрезвычайно сложная область, недоступная для неспециалистов. Таким образом, для того чтобы статистика играла надлежащую роль в общественных дискуссиях, важно найти соответствующие инструменты и каналы для доведения соответствующей статистической информации до широкой общественности, состоящей из неспециалистов, также принимая во внимание ее неоднородность. Непрофессиональные пользователи разнообразны. Следовательно, они представляют собой группу, которой требуются самые разнообразные каналы коммуникации. Среди широкой общественности можно выделить такие группы пользователей, как студенты университетов, пенсионеры, семьи, представители гражданского общества и журналисты широкого профиля. СМИ также играют важную роль в том смысле, что они распространяют информацию. Должны соблюдаться некоторые общие критерии в отношении каналов коммуникации, основанной на показателях, ориентированных на широкую общественность. В этих каналах следует ограничить использование технических определений и/или заменить их описаниями, основанными на легком для понимания языке. Чтобы информация была доступна гражданам, ее следует представлять таким образом, чтобы она имела непосредственное отношение к повседневной жизни людей. Непрофессиональным пользователям нужна разнообразная, но не чрезмерно подробная статистическая информация.
403. Пользователи могут быть национальными гражданами, но они также могут быть международными гражданами. Оба пользователя могут иметь разные потребности и интересы.

4.2.2 Распространение лонгитюдной статистики миграции

404. Набор данных является лонгитюдным, если он отслеживает информацию одного и того же типа по одним и тем же темам в разные моменты времени. И наоборот, данные поперечного сечения - это данные, которые собираются от респондентов в один момент времени. Таким образом, хотя в первом случае время является важным измерением, во втором - нет. Основное преимущество лонгитюдных баз данных заключается в том, что они могут измерять постоянство или изменения внутри отдельных людей. Лонгитюдные данные также могут показать, как действия и события могут повлиять на результаты в дальнейшей жизни. Следовательно, данные такого типа особенно подходят для изучения явлений, которые развиваются с течением времени, таких как миграция. Это также может позволить исследователям изучить, как конкретные явления, такие как миграция, могут повлиять на результаты дальнейшей жизни, такие как занятость, условия жизни, здоровье, гражданство и т.д. Поэтому представление лонгитюдных данных должно быть сосредоточено на наиболее подходящем представлении для изучения эволюции.
405. Ниже приведены несколько примеров того, как некоторые страны распространяли результаты, основанные на лонгитюдных данных миграции, которые являются результатом связанных данных.

4.2.2.1 Печатные или онлайн-отчеты и/или публикации

406. Этот канал коммуникации часто содержит подробное описание теоретических основ, показателей и методов оценки (Евростат, 2014). Целевая аудитория заинтересована в углубленном анализе соответствующего вопроса и уделяет ему больше времени (Евростат, 2017b). Они, как правило, обладают базовой информацией по данному вопросу и используют информацию из аналитических публикаций для выработки выводов в своих собственных исследованиях или для принятия политических решений, основанных на фактических данных. Они могут обладать, а могут и не обладать глубокими статистическими знаниями. Большинство НСУ размещают свои публикации на своем веб-сайте, с которого их можно скачать бесплатно. Публикации, предназначенные для непрофессиональных пользователей, должны содержать не только таблицы. Они также должны содержать пояснительные тексты и графики, которые помогут им лучше понять данные.
407. В качестве примера, Канада опубликовала различные статьи по различным областям, таким как виды на жительство и доходы мигрантов, на основе канадской базы данных о долгосрочной иммиграции (IMDB). Статьи варьировались от коротких описательных материалов (например, "Только факты: заявители, ходатайствующие о предоставлении убежища": <https://www150.statcan.gc.ca/n1/pub/89-28-0001/2018001/article/00013-eng.htm>), углубленные аналитические продукты (Бониковская и Хоу, 2015; Хоу и Бониковская, 2015; Прокопенко и Хоу, 2016; Ки и Мориссетт, 2017; Лу и Хоу, 2017), и справочные материалы. IMDB также опубликовал технический отчет, в котором обсуждаются источники данных, концепции и переменные IMDB, взаимосвязь записей, обработка, распространение, оценка данных и показатели качества, сопоставимость с другими наборами данных об иммиграции и анализ, возможный с помощью IMDB (Evra & Prokopenko, 2019). Этот набор материалов предназначен для удовлетворения потребностей широкого круга заинтересованных сторон – от тех, кто ищет базовые истории, до опытных исследователей, непосредственно использующих полученные данные.
408. Италия также опубликовала рабочий документ по видам на жительство и приобретению гражданства (Conti et al, 2014).
409. Швейцария опубликовала рабочий документ, описывающий процесс, использованный для создания Швейцарской лонгитюдной демографической базы данных (SLDD) (Steiner & Wanner, 2015). В документе приводятся два примера возможных применений: модели миграции и интеграция на рынке труда. Ваннер и Хайнигер опубликовали второй рабочий документ в 2017 году о правовых и технических разработках в Швейцарии, которые позволили создать упомянутую базу данных. В нем представлены некоторые результаты двух показателей, которые были построены и рассчитаны с использованием этой базы данных: натурализация и коэффициенты выбытия когорт мигрантов.

4.2.2.2 Таблицы данных

410. Основная функция таблиц данных заключается в упрощении сравнения различных показателей (ПАРИЖ-21 и Статистическое управление Норвегии, 2009). Несмотря на все более широкое

использование графиков, карт и других средств визуализации, таблицы, безусловно, останутся важным инструментом представления и распространения статистических данных в будущем. Если раньше основное внимание уделялось статическим таблицам, то сейчас страны заменяют их более динамичными инструментами, с помощью которых пользователи могут создавать свои собственные табличные представления.

411. В качестве альтернативы, если этих онлайн-таблиц недостаточно для удовлетворения потребностей пользователей, есть также варианты создания пользовательских таблиц на основе пользовательских требований или предоставления прямого доступа к анонимизированным данным в защищенных помещениях для более продвинутых пользователей. Исследователям нужны данные для их работы. Поэтому они должны иметь возможность легко получать доступ к онлайн-данным, которые хорошо документированы и отличаются высоким качеством, чтобы повысить качество своих исследований (Европейские сообщества, 2005).
412. В качестве примера, Статистическое управление Канады распространяет динамические таблицы в режиме онлайн для базы данных о лонгитудной иммиграции, предлагает возможность настраивать таблицы и выводить модели, а также предоставляет анонимные результаты для исследователей в защищенных центрах обработки данных.¹¹
413. Онлайн-показатели — интерактивные агрегированные панели управления
414. Информация предоставляется поэтапно. Использование настраиваемой интерактивной сводной информационной панели позволяет пользователю получить общее представление о ситуации (сводная информация) и в то же время предоставляет доступ к подробной информации (на отдельных уровнях показателей) (Евростат, 2014). Также необходимо предоставить ссылки на дополнительную информацию, касающуюся теоретических основ, методов оценки и т.д. Целевая аудитория заинтересована в получении общего представления о ситуации (прогрессе) и в доступе к исходным данным. Они часто первыми осваивают новые технологии и предпочитают синтезированную текстовую и визуальную информацию традиционным публикациям (Евростат, 2017b). Обычно у них есть некоторая справочная информация по данному вопросу, которая позволяет им оценить интерактивность технологических продуктов, устанавливая свои собственные предпочтения в отношении контента.
415. В Швейцарии Национальный центр компетентности в области исследований (NCCR) на основе SLDD разработал десять долгосрочных показателей в различных областях, таких как модели миграции, виды на жительство, натурализация, (административные) причины миграции, квалификация, предоставление убежища и т.д. С этими долгосрочными показателями можно ознакомиться в интерактивном режиме на их веб-сайте, что позволяет проводить персональную параметризацию¹². Таким образом, они являются инструментом для лучшего понимания процессов миграции и интеграции групп мигрантов в течение более длительного периода времени. Например, на интерактивном графике, показывающем квалификацию мигрантов в Швейцарии¹³, можно выбрать страны с самым высоким уровнем высшего образования, а также уровень высшего образования в разбивке по полу за разные периоды времени. Другой интерактивный график, посвященный административным мотивам въезда в Швейцарию, позволяет пользователям изучить эволюцию административных мотивов въезда на национальном уровне, но также можно выбрать континент и страну¹⁴. Под интерактивными графиками каждого индикатора всегда доступно краткое описание данных. Существует также некоторая информация об источнике, методологии, определениях и т.д.

11. <https://www150.statcan.gc.ca/n1/en/surveys/5057>

12. <https://indicators.nccr-onthemove.ch/>

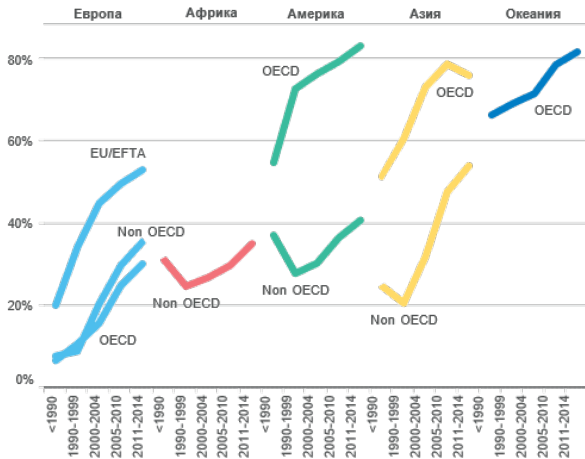
13. <https://indicators.nccr-onthemove.ch/how-qualified-are-migrants-in-switzerland/>

РИСУНОК 35: КВАЛИФИКАЦИЯ МИГРАНТОВ В ШВЕЙЦАРИИ

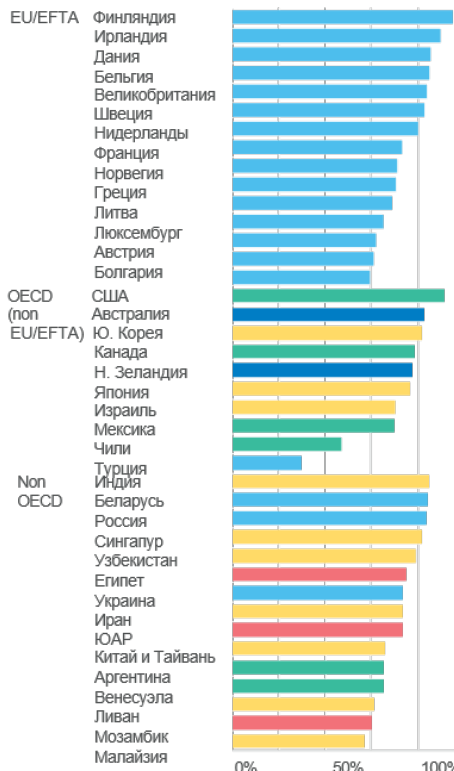
nccr → **on the move**

Доля мигрантов с высшим образованием

Уровень высшего образования в разбивке по группам стран

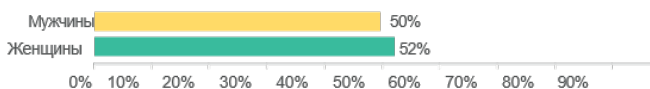


Страны с самым высоким уровнем высшего образования (2011-2014)



Уровень высшего образования в разбивке по полу (2011-2014)

Все страны



© nccr - on the move / Sources: SF50 / SEM (CAR, CEMIS, STATPOP)

Источник: NCCR в движении, Швейцарская лонгитюдная демографическая база данных (SLDD), 2017.

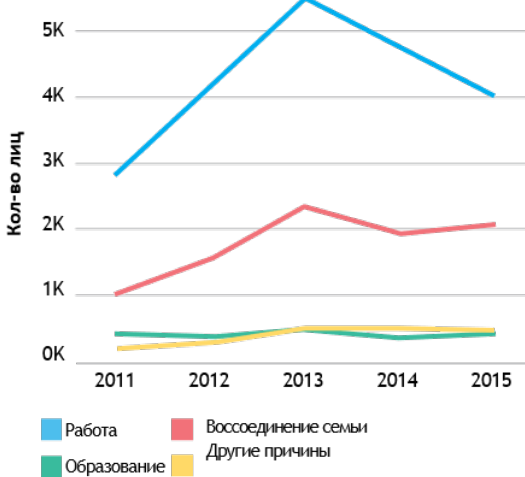
416. Статистическое управление Канады публикует интерактивное приложение, использующее результаты Лонгитюдной базы данных по иммиграции (IMDB).¹⁵ Это интерактивное приложение напрямую подключается к кубам данных, доступных онлайн, и отображает состав (по категории допуска) и тенденции доходов (по годам с момента допуска и по категории допуска) для иммигрантов на основе набора фильтров (год допуска, источник дохода, провинция проживания, пол и предварительный опыт). Оно также позволяет пользователям легко смотреть на различия в траекториях доходов между категориями допуска иммигрантов, контролируя различные характеристики.

РИСУНОК 36: АДМИНИСТРАТИВНЫЙ МОТИВ ВЪЕЗДА В ШВЕЙЦАРИЮ

nccr
on the move

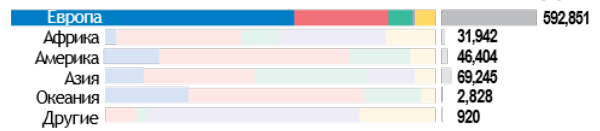
Доля мигрантов с высшим образованием

**Административный мотив въезда:
Европа, Испания**

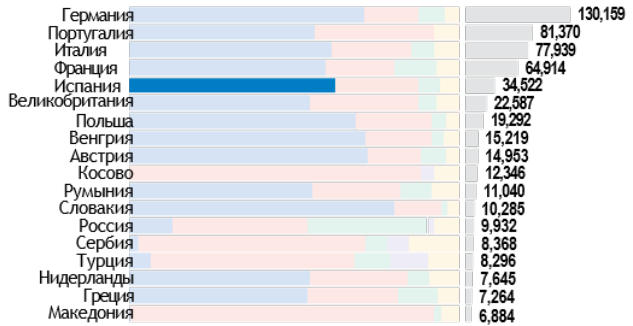


© nccr - on the move / Sources: SFSO / SEM (CAR, CEMIS, STATPOP)

Распределение по континентам 2011-2015



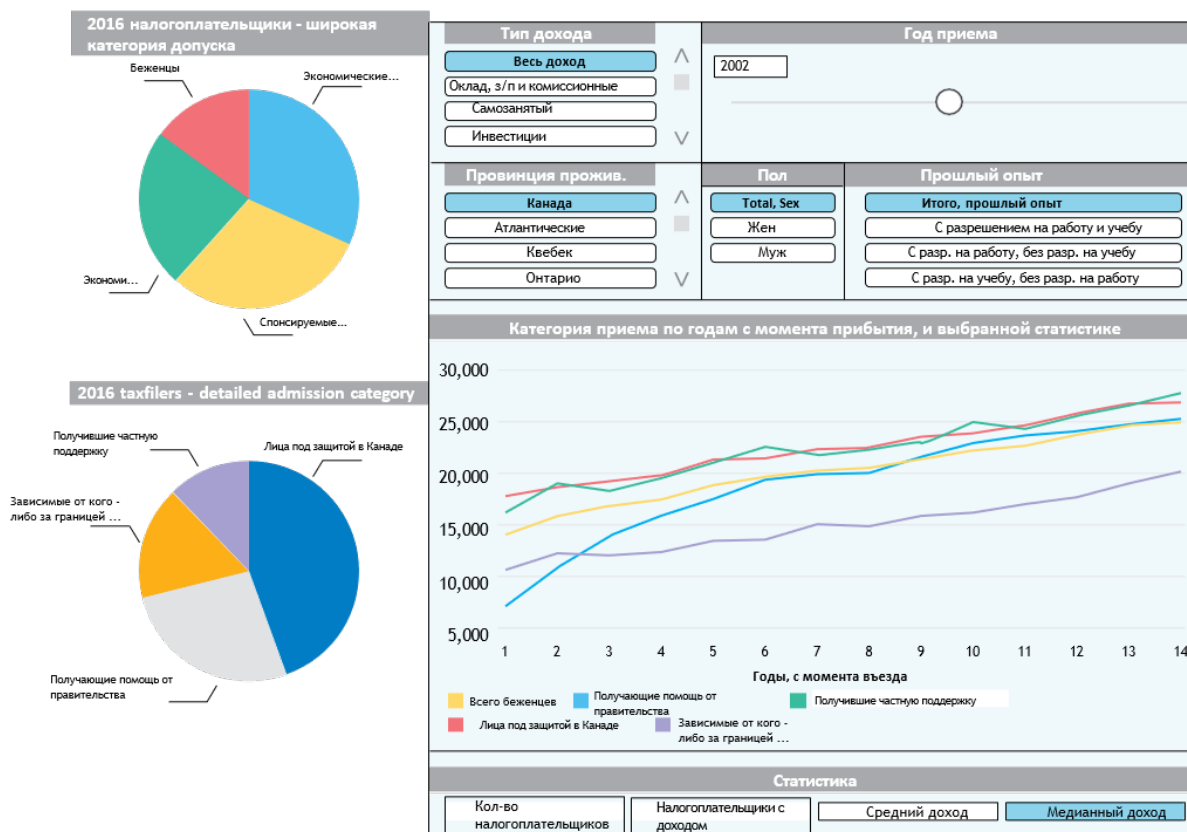
Распределение по странам (основные страны происхождения), Европа (2011-2015)



Источник: NCCR «В движении», Швейцарская лонгитюдная демографическая база данных (SLDD), 2017.

417. Великобритания не использует интерактивные информационные панели для представления результатов, но использует интерактивный инструмент онлайн-опроса, чтобы показать, что происходит с участниками их лонгитюдного исследования (LS) с течением времени.¹⁶ Можно выбрать дату начала и дату окончания. Выход из LS происходит двумя способами: смертью или эмиграцией. Качество данных о смерти в LS очень высокое, поскольку регистрация смерти требуется по закону. Однако данные об эмиграции менее качественные, поскольку они основаны на том, что члены LS информируют Национальную службу здравоохранения (NHS) о своем отъезде, что приводит к пробелам в охвате данных. Это в значительной степени объясняет количество людей, отнесенных к категории потерянных для последующего наблюдения.

РИСУНОК 37: ИНТЕРАКТИВНЫЙ ГРАФИК, ОТРАЖАЮЩИЙ СОСТАВ И ДИНАМИКУ ДОХОДОВ ИММИГРАНТОВ В КАНАДЕ



Источник: Статистическое управление Канады, база данных о долгосрочной иммиграции.

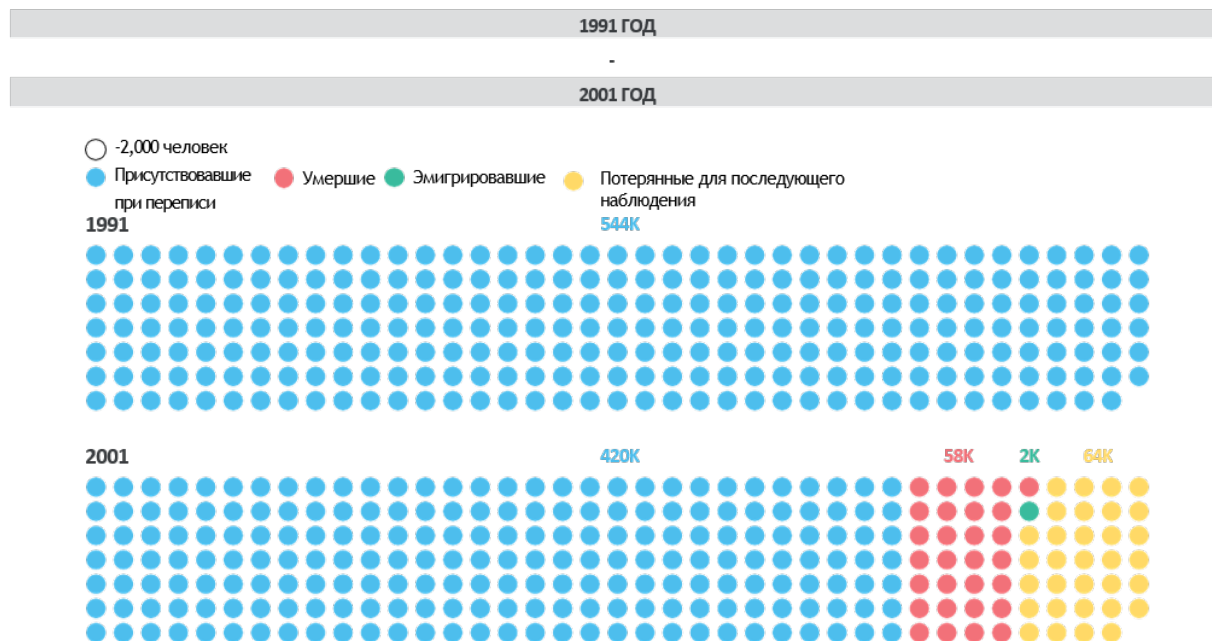
4.2.2.3 Аудиовизуальные продукты

418. Видео ориентированы на пользователей, которые предпочитают видео-коммуникацию, не имеют времени для длительного чтения и/или нуждаются в информировании о последних тенденциях и разработках или основных особенностях соответствующего вопроса (Eurostat, 2017b). Они обычно используют видео как общий источник информации без конкретных целей использования информации в их собственном анализе или исследовании.

419. Например, NCCR «В движении», в Швейцарии создает короткие видеоролики на YouTube и размещает их на своем веб-сайте¹⁷. На момент написания было размещено пять видеороликов на основе лонгитюдных данных с целью решения нескольких предубеждений о мигрантах, таких как:

- Мигранты приезжают, чтобы воспользоваться преимуществами Швейцарии;
- Мигранты приезжают из бедных стран;
- С таким количеством прибытий скоро мигрантов будет больше, чем швейцарцев;
- Мигранты заботятся только о своей стране происхождения, а не о Швейцарии; и
- Все мигранты хотят швейцарский паспорт.

РИСУНОК 38: ИНТЕРАКТИВНЫЙ ИНСТРУМЕНТ КОГОРТЫ ДЛЯ УЧАСТНИКОВ ЛОНГИТЮДНОГО ИССЛЕДОВАНИЯ



Источник: Управление национальной статистики Великобритании, лонгитюдное исследование.

4.2.2.4 Инфографика

420. Согласно Оксфордскому словарю английского языка, инфографика – это «визуальное представление информации или данных». Инфографика представляет собой набор изображений, диаграмм и минимального количества текста, предназначенный для привлечения внимания и предоставления простого для понимания обзора темы. Инфографика отлично подходит для упрощения восприятия сложной информации. Ее можно использовать для обмена статистическими данными и данными переписи. В ней больше внимания уделяется цифрам, диаграммам и другим данным. Как правило, она содержит гораздо меньше текста, чем информационная инфографика, и имеет меньший объем повествования. Вместо этого они делают заявления, используя большие цифры и отдельные факты, которые можно легко донести с помощью новых способов, таких как социальные сети. Целевая аудитория инфографики очень широка: от студентов, профессионалов до широкой общественности. Ни одна страна еще не публиковала инфографику с использованием данных о долгосрочной миграции. Швейцария опубликовала несколько инфографик по миграции и интеграции, в которых использованы данные из разных стран¹⁸.

18 https://www.bfs.admin.ch/bfs/en/home/statistics/catalogues-databases/infographics.html?dyn_prodim=900999&dyn_publishingyearend=2019

ГЛАВА 5

ВЫВОДЫ И РЕКОМЕНДАЦИИ

5.1 ВЫВОДЫ

421. Изучать миграцию — значит изучать изменение: изменение места жительства, изменение правового статуса и изменение социально-экономических результатов. Более того, связанные темы интеграции и обустройства являются процессами, а не состояниями. Результаты для мигрантов могут быть краткосрочными или долгосрочными.
422. Для производства статистики по миграции и связанным темам необходимо использовать данные, измеряющие изменение. Лонгитюдные данные предоставляют средства для понимания изменения на индивидуальном уровне. Помимо простого наблюдения изменений в месте жительства или результатах, отслеживание отдельных лиц с течением времени позволяет лучше понять факторы, которые могут влиять на изменение. Эти факторы могут действовать на макроуровне, такие как изменения политики и экономические сдвиги, или на микроуровне, такие как получение образования и приобретение гражданства.
423. Исследователи и официальные статистики все чаще используют методы, основанные на лонгитюдных подходах, чтобы лучше понять миграционные потоки и динамику обустройства в принимающих странах. Лонгитюдные подходы являются наиболее подходящими для измерения того, как разрыв между иммигрантами и коренным населением изменяется с течением времени, прошедшим с момента иммиграции, например, с точки зрения заработной платы или безработицы.
424. С увеличением доступности административных данных страны теперь могут лучше использовать такие методы, как интеграция данных, чтобы иметь возможность разрабатывать детальные лонгитюдные данные для статистики миграции. Анализы интегрированных административных данных могут отвечать на вопросы, которые требуют больших размеров выборки с богатыми и детальными данными о трудно достижимых популяциях, и генерировать доказательства с высоким уровнем внешней валидности и применимости для выработки политики. Этот подход позволяет лонгитюдно наблюдать результаты миграции или обустройства без увеличения бремени сбора данных через обследования.
425. Настоящее Руководство показывает, как продольные (лонгитюдные) данные могут быть разработаны и использованы для международной статистики миграции. В главе 2 представлен обзор различных типов данных – от панельных опросов до регистров населения. В главе 3 рассказывается о том, как создавать источники лонгитюдных данных на основе интегрированных данных. Наконец, в главе 4 предлагаются ключевые лонгитюдные показатели, связанные с изучением международной миграции, и передовая практика для распространения.
426. Хотя продольные данные предоставляют критическую и уникальную возможность изучать миграцию и обустройство как процессы, настоящее Руководство признает проблемы и ограничения, связанные с такими данными. Эти проблемы варьируются в зависимости от типа источника данных, как описано в главе 2.
427. Традиционные методы, такие как лонгитюдные панельные обследования, обладают преимуществами в плане контроля, предлагаемого статистическим управлением в отношении охвата и измерения. Однако из-за оперативных соображений эти источники, как правило, имеют меньший размер выборки, страдают от истощения и нерегулярности сбора данных. Эти ограничения могут препятствовать их использованию для изучения краткосрочных или долгосрочных результатов для небольших групп населения (например, подгрупп мигрантов).
428. С другой стороны, существуют также ограничения, связанные с использованием интегрированных административных данных. В частности, эти данные не были собраны для статистических целей, и поэтому теряется контроль над содержанием, охватом и измерениями. Кроме того, процесс интеграции может привести к ошибкам, как описано в главе 3. Однако эти методы, как правило, позволяют получить выборку большого объема при более регулярном наблюдении и практически без потерь из-за отсутствия ответов. В результате, они, как правило, подходят для использования в долгосрочном исследовании небольших групп мигрантов.

429. Существуют доступные методы для смягчения, устранения или обхода ограничений источников данных, с тем чтобы распространяемые данные по-прежнему служили целям, изложенным в статистической структуре. В главе 3 представлены различные рекомендации по разработке источников данных, а в главе 4 показано, как можно создавать ценные показатели, несмотря на ограничения, связанные с результирующей базой данных.
430. Важно не игнорировать трудности, но и не допускать, чтобы они создавали препятствия. Важно понимать, соответствует ли какая-либо конкретная статистика поставленной цели. Интегрированные данные для лонгитюдной статистики миграции всегда сопряжены с определенными трудностями, но они также открывают двери для беспрецедентных аналитических возможностей, проливающих свет на подробные лонгитюдные результаты деятельности мигрантов и их влияние на принимающие страны. Эти источники данных не всегда могут дать точные точечные оценки, но они предлагают нечто более содержательное – историю.
431. Эффективность лонгитюдных данных для статистики миграции наглядно проиллюстрирована примерами из главы 4. В конечном счете, преимущества лонгитюдных данных для статистики миграции свидетельствуют о том, что проблемы необходимо преодолевать.

5.2 РЕКОМЕНДАЦИИ

432. Для лучшего понимания процессов миграции и обустройства рекомендуется увеличить использование продольных данных для статистики миграции.
433. При разработке лонгитюдного набора данных рекомендуется следовать каждой из шести фаз, описанных в главе 3, в представленном порядке. В зависимости от результатов может быть необходимо вернуться к одной из более ранних фаз, в этом случае последующие фазы должны быть повторены. Фазы следующие:
1. *Статистический дизайн*
 2. *Оценка и предварительная обработка исходных файлов*
 3. *Интеграция данных для лонгитюдных данных*
 4. *Назначение лонгитюдных индивидуальных идентификаторов*
 5. *Создание окончательной базы данных*
 6. *Распространение результатов*
434. Рекомендуется, чтобы там, где это возможно с учетом наличия данных и других возможных ограничений, страны регулярно разрабатывали лонгитюдные показатели для статистики миграции, охватывающие темы, описанные в главе 4.
435. Эти рекомендации и руководящие указания, приведенные в предыдущих главах, предусматривают конкретные дальнейшие шаги национальных статистических управлений по интеграции лонгитюдных данных в официальную статистику миграции. Однако также признается, что не все страны располагают инфраструктурой, потенциалом или установленными приоритетами для выполнения всех рекомендаций, содержащихся в настоящем руководстве. Хотя было бы полезно дать дополнительные рекомендации, особенно в области обмена данными или установления международных стандартов для долгосрочных показателей, необходимо также учитывать эти ограничения, характерные для конкретной страны. Рекомендуется продолжить международное сотрудничество в этих областях.

5.3 ОБЛАСТИ БУДУЩЕЙ РАБОТЫ

436. Этот отчет охватил широкий спектр тем, связанных с миграцией, которые могут выиграть от лонгитюдных данных. Некоторые темы были рассмотрены более подробно, особенно в главе 4 через предлагаемые показатели. Однако другие были сочтены выходящими за рамки данного Руководства.
437. В частности, дальнейшее исследование было бы необходимо о том, как измерять семейную миграцию. Из-за своей сложности эта тема может потребовать более сфокусированного изучения. Это может включать изучение подтем, таких как:

- Миграционные модели семей
- Семьи мигрантов как единицы анализа (например, результаты обустройства семей мигрантов с течением времени)
- Воссоединение семьи и члены семьи за границей
- Жилищные условия мигрантов в принимающей стране (например, многодетные домохозяйства, многопоколенные домохозяйства)

438. Еще одной темой, заслуживающей дополнительного внимания, является структура расселения на субнациональном уровне, рассмотренная в разделе 4.1.1.2 Продолжительность пребывания в субнациональной географии. Однако страны все чаще проявляют интерес к рассмотрению структуры районов проживания и поощрению мигрантов селиться в нетрадиционных местах (например, по экономическим причинам). Понимание географических особенностей низкого уровня может быть важным для заинтересованных сторон, таких как поставщики услуг и школы. Однако, как уже говорилось в данном руководстве, особенно в главе 4, при проведении такого рода исследований возникают особые проблемы. В частности, границы меняются со временем, и иногда перемещение на небольшое расстояние может рассматриваться как изменение географии, в то время как перемещение на большие расстояния может оставаться в пределах одной и той же географической области.

439. Еще одной развивающейся областью являются способы распространения результатов статистическими службами. Некоторые инновационные подходы представлены в главе 4. С появлением новых инструментов распространения и более сложных данных статистическим службам приходится пересматривать способы наилучшего представления результатов различным аудиториям. Для статистики миграции это особенно важная задача. В связи с недавним ростом международной миграции в странах ведется широкое общественное обсуждение, требующее доказательств о характере миграции и ее влиянии на отдельных лиц, семьи и общество. Поскольку статистика миграции сложна по своей природе, существует риск неправильного толкования статистических результатов. Дополнительные рекомендации о том, как представлять результаты для охвата различных аудиторий, соблюдая при этом ограничения в отношении качества данных, могли бы помочь национальным статистическим управлениям справиться с этой возникающей проблемой.

СПИСОК ЛИТЕРАТУРЫ

Aoki, Y. & Santiago, L. (2015). Education, Health and Fertility of UK migrants: The Role of English Language Skills. IZA Discussion Paper Series No.498 (ONS LS), доступно по адресу: <https://calls.ac.uk/wp-content/uploads/dp9498.pdf>

Berti-Equille, Laure. (2007). Measuring and Modelling Data Quality for Quality-Awareness in Data Mining. 10.1007/978-3-540-44918-8_5.

Blackwell, L. & Rogers, N. (2019). Placing administrative data at the heart of the UK population statistics system; how the total survey error framework is helping to inform the use and design of integrated data solutions. Retrieved from International Total Survey Error Workshop 2019: <https://www.niss.org/news/itsew-2019-international-total-survey-error-workshop>.

Bonikowska, A. & Hou, F. (2015). Which human capital characteristics best predict the earnings of economic immigrants? Analytical Studies Branch Research Paper Series, 368, 1-30.

Brückner, G., (2018). The German Central Register of Foreigners as a longitudinal data source. Note by German Federal Statistical Office, United Nations Economic Commission for Europe, Conference of European Statisticians, Work Session on Migration Statistics. Geneva, Switzerland 24-26 October 2018.

Burke, D. L., Bujkiewicz, S. and Riley, R. D. (2016). Bayesian bivariate meta-analysis of correlated effects: Impact of the prior distributions on the between-study correlation, borrowing of strength, and joint inferences. *Statistical Methods in Medical Research*, 0(0) 1--27.

Chambers, R., Tzavidis, N., and Salvati, N. (2009). Borrowing strength over space in small area estimation: Comparing parametric, semi-parametric and non-parametric random effects and M-quantile small area models. Retrieved from University of Wollongong Research Online: <https://pdfs.semanticscholar.org/eede/9fa86149d5cc312da377dd24d63ab723aa4e.pdf>

Ci, W., & Morissette, R. (2017). Acquisition of permanent residence by temporary foreign workers in Canada: A panel study of labour market outcomes before and after the status transition. Analytical Studies Branch Research Paper Series, 396, 1-31.

Conti, C., Quattrociochi, L., and Rottino, F. M. (2014). New statistics on residence permits and acquisition of citizenship: experiences of integration in Italy. Economic Commission for Europe, Group of Experts on Migration Statistics, Work Session on Migration Statistics, Working paper 14, Geneva, 10-12 September 2014.

Correa-Onel, S., Whitworth, A. and Piller, K. (2016). Assessing the Generalised Structure Preserving Estimator (GSPREE) for Local Authority Population Estimates by Ethnic Group in England. GSS Methodology Series No 42, Office for National Statistics: Newport.

Eberle, J., (2018). Foreigners seeking humanitarian protection. Measuring the stock of foreigners seeking humanitarian protection in Germany. Note by German Federal Statistical Office, United Nations Economic Commission for Europe, Conference of European Statisticians, Work Session on Migration Statistics. Geneva, Switzerland 24-26 October 2018.

EGRIS -- Expert Group on Refugee and Internally Displaced Persons Statistics (2018). International recommendations on refugee statistics. United Nations, Eurostat. Доступно по адресу: <https://www.unhcr.org/refugee-statistics/methodology>.

EGRIS -- Expert Group on Refugee and Internally Displaced Persons Statistics (2020). International recommendations on IDP Statistics. Background document for the 51st session of the United Nations Statistical Commission. Доступно по адресу: <https://unstats.un.org/unsd/statcom/51st-session/documents>.

European Communities (2005). The challenge of communicating statistics. Proceedings of the 91st DGINS Conference, Copenhagen. Получено из: <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/KS-EE-05-001>.

Eurostat (2014). Getting messages across using indicators. A handbook based on experiences from assessing Sustainable Development Indicators. Luxembourg: Publications Office of the European Union.

Eurostat (2017a). A set of tables for circular migration. UNECE Work Session on Migration Statistics, Geneva, 30-31 October 2017. Eurostat (2017b). Towards a harmonized methodology for statistical indicators. Part 2 -- communication through indicators. Luxembourg: Publications Office of the European Union.

Eurostat (2018). Circular migration: new migration topics and revised tables. UNECE Work Session on Migration Statistics, Geneva, 24-26 October 2018.

Evra, R., & Prokopenko, E. (2019). Longitudinal Immigration Database (IMDB) Technical Report. Analytical Studies: Methods and References, Catalogue no. 11-633-X2019005.

Frenette, M., Lu, Y., and Chan, W. (2019). The Postsecondary Experience and Early Labour Market Outcomes of International Study Permit Holders. Analytical Studies Branch Research Paper Series, Statistics Canada.

Globerman, S., Barua, B., and Hasan, S. (2018). The Supply of Physicians in Canada: Projections and Assessment. Fraser Institute. Green, D., Liu, H., Ostrovsky, Y., and Picot, G. (2016). Business Ownership and Employment in Immigrant-owned Firms in Canada.

Economic Insights Catalogue no. 11- X 626- -- No. 057.

Groves, R., Fowler, F., Couper, M., Singer, E., and Tourangeau, R. (2004). Survey Methodology, New York Wiley.

Hou, F., & Bonikowska, A. (2015). The earnings advantage of landed immigrants who were previously temporary residents in Canada.

Analytical Studies Branch Research Paper Series, 370, 1-36.

INE (2018). Duration of migratory episodes in Spain: Procedure of calculation and results. UNECE Work Session on Migration Statistics, Geneva, 24-26 October 2018.

Jackson, D., White, I. R., Price, M., Copas, J., and Riley, R. D. (2017). Borrowing of strength and study weights in multivariate and network meta-analysis. *Statistical Methods in Medical Research*, 26(6), 2853--2868.

Johnston, T., & Weiss, R. (2010). *Managing Time in Relational Databases*. Elsevier
Johnston, T. (2014). *Bitemporal Data: Theory and Practice*. Elsevier

Lu, Y. & Hou F. (2017): Transition from temporary foreign workers to permanent residents, 1990 to 2014. Analytical Studies Branch Research Paper Series, 389, 1-32.

Ng, E., Sanmartin, C., Elie-Massenat, D., Manuel, D. (2016). Vaccine-preventable disease-related hospitalization among immigrants and refugees to Canada: use of two population-based Immigrant linked cohort studies. *Vaccine* 34 (2016):4433-4442.

Ng, E., Elie-Massenat, D., Giovanazzo, G., Ponka, D., and Sanmartin, C. (2018). Tuberculous related hospitalization among recent immigrants; a linkage study. *Health Reports*, Vol 29 no 7, 14-28.

PARIS21 & Statistics Norway (2009). User-friendly presentation of statistics. Guide to creating a dissemination strategy and dissemination guidelines for developing and transition countries. Oslo: Statistics Norway.

Perrin, N. (2006). A Cohort Approach to Acquisition of Citizenship Statistics. In: Poulain, M., Perrin, N., and Singleton, A. (Ed.): *THESIM: Towards Harmonised European Statistics on International Migration*, Presses Universitaires de Louvain.

Poulain, M., Perrin, N., and Singleton, A. (Ed.) (2006). *THESIM: Towards Harmonised European Statistics on International Migration*, Presses Universitaires de Louvain.

Prokopenko, E., & Hou, F. (2016). How temporary were Canada's temporary foreign workers? Analytical Studies Branch Research Paper Series, 402, 1-31.

Reichel, David (2011). Do Legal Regulations Hinder Naturalisation? Citizenship policies and naturalisation rates in Europe. EUDO Citizenship Observatory.

Prins, K. (2016). Population register data, basis for the Netherlands' population statistics. Statistics Netherlands, Bevolkingstrends Januari 2016.

Statistics New Zealand (2016). Guide to reporting on administrative data errors. Statistics New Zealand, Wellington. Получено из: <http://archive.stats.govt.nz/methods/data-integration/guide-to-reporting-on-admin-data-quality/sources-of-error.aspx>

Steiner, I., & Wanner, P. (2015). Towards a new data set for the analysis of migration and integration in Switzerland. Working Papers -- nccr on the move, 1, 1-22.

UK Home Office (2019). Migrant Journey: 2018 Report. Statistical Bulletin 07/19. Получено из: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/803754/migrant-journey-report2018.pdf.

UK Office for National Statistics (2019). International Passenger Survey 3.15. Получено из: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/datasets/internationalpassengersurveyactuallengthofstaybycitizenshiptable315>.

UNECE (2015): Measuring change in the socio-economic conditions of migrants. New York and Geneva: United Nations. Получено из UNECE: <http://www.unece.org/index.php?id=40542>](<http://www.unece.org/index.php?id=40542>

UNECE (2016): Defining and measuring circular migration. New York and Geneva: United Nations. Получено из UNECE: <http://www.unece.org/index.php?id=44717>.

UNECE (2019a). Generic Statistical Business Process Model, version 5.1. Получено из UNECE: <https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1>

UNECE (2019b). Guidance on data integration for measuring migration. New York and Geneva: United Nations.

United Kingdom Home Office (2019). User guide to the Home Office statistics on exit checks. Получено из UK Government Publishing Service: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/826370/user-guide-to-statistics-on-exit-checks.pdf

United Nations (2017). Handbook on measuring international migration through population censuses

United Nations (2018). Handbook on civil registration and vital statistics systems: management, operation and maintenance United Nations (2020). Guidelines on the legislative framework for civil registration, vital statistics and identity management. Vachon, J., Gallant, V., and Siu, W. (2018). Tuberculosis in Canada, 2016.

Canada Communicable Disease Report 2018; 44(3/4): p. 75 -- 81. Vanthomme, K., and Vandenhede, H. (2019). Trends in Belgian cause-specific mortality by migrant origin between the 1990s and the 2000s. BMC Public Health 19:410.

Wallace, M. (2016). Adult mortality among the descendants of immigrants in England and Wales: does a migrant mortality advantage persist beyond the first generation? Journal of Ethnic and Migration Studies, 42:9, 1558-1577, DOI: 10.1080/1369183X.2015.1131973. Доступно по адресу: <https://www.tandfonline.com/doi/full/10.1080/1369183X.2015.1131973>.

Wanner, P., & Heiniger, M. (2017). Migration and structural integration in Switzerland. A longitudinal perspective. Economic Commission for Europe, Group of Experts on Migration Statistics, Work Session on Migration Statistics, Working paper 7, Geneva, 5-6 October 2017.

Wilson, B., & Kuha, J. (2017). Residential segregation and the fertility of immigrants and their descendants. *Popul Space Place*. 2018; 24:e2098. Available at: <https://doi.org/10.1002/psp.2098>

Zhang, L.-C. (2012). Topics of Statistical Theory for Register-Based Statistics and Data Integration. *Statistica Neerlandica* 66: 41 -- 63.

Zuccotti, C. V., & Platt, L. (2016). A reconsideration of ethnic penalties in inactivity and unemployment: a study of second generation men and women in England and Wales. ISA RC28 Summer Meeting, University of Bern, Switzerland, 29 -- 31 August 2016 \[ONS LS\].

Руководство по использованию продольных данных для статистики миграции

По мере роста числа международных мигрантов для общественности и политиков становится все более важным понимать миграционные потоки и влияние миграции на отдельных людей, семьи, общество и экономику. Для этого необходимо учитывать изменения с течением времени, как подчеркивается в Глобальном договоре о безопасной, упорядоченной и законной миграции. Лонгитюдный подход, при котором информация собирается от одних и тех же лиц или домохозяйств с течением времени, особенно полезен для этого, поскольку он может выявить географические и социально-экономические последствия миграционного опыта.

Источники данных, доступные многим национальным статистическим службам, обладают большим потенциалом для составления лонгитюдной статистики миграции, который на сегодняшний день используется недостаточно. Чтобы улучшить ситуацию, целевая группа экспертов из национальных статистических управлений и международных организаций в течение 2018-2020 годов работала над настоящим Руководством.

Руководство состоит из следующих глав:

- Глава 1: Введение
- Глава 2: Обзор источников лонгитюдных данных для статистики миграции
- Глава 3: Как разработать лонгитюдный набор данных для статистики миграции с использованием интегрированных данных
- Глава 4: Регулярное распространение статистики миграции из лонгитюдных источников данных
- Глава 5: Выводы и рекомендации

Главы организованы как самостоятельные части, и их нет необходимости читать в сочетании с другими частями публикации, что позволяет пользователям легко сосредоточиться на тех частях, которые наиболее важны для их работы.

Руководство было одобрено на 68-й пленарной сессии Конференции европейских статистиков в 2020 году.

Информационная служба
Европейская экономическая комиссия Организации Объединенных Наций

Дворец Наций
СН - 1211 Женева 10, Швейцария
Телефон: +41(0)22 917 12 34
Факс: +41(0)22 917 05 05
E-mail: unece_info@un.org
Веб-сайт:
<http://www.unece.org>

Составлено и напечатано в Организации Объединенных Наций, Женева -- 2015311 (E) -- Декабрь 2020 -- 754 --
ECE/CES/STAT/2020/6

ISBN 978-92-1-117263-8