

**Европейская экономическая комиссия****Конференция европейских статистиков****Шестьдесят восьмая пленарная сессия**

Женева, 22–24 июня 2020 года

Пункт 4 f) предварительной повестки дня

**Доклады, руководства и рекомендации,  
подготовленные под эгидой Конференции:****Продольные данные по миграции****Руководство по использованию продольных данных  
в ежегодной статистике миграции<sup>1</sup>****Записка Целевой группы по использованию продольных данных  
для статистики миграции***Резюме*

Настоящее Руководство было подготовлено Целевой группой по использованию продольных данных для статистики миграции в составе представителей Австрии, Бельгии, Германии, Испании, Италии, Казахстана, Канады (Председатель), Мексики, Нидерландов, Российской Федерации, Соединенного Королевства, Турции, Швейцарии, Евростата, ОЭСР и ЕЭК ООН.

Настоящий краткий вариант Руководства подготовлен для целей перевода. Он содержит введение, вводные разделы всех основных глав и полную главу с выводами, рекомендациями и направлениями дальнейшей работой.

Полный текст Руководства был разослан всем членам Конференции европейских статистиков для электронных консультаций. С ним можно ознакомиться на веб-странице Конференции <http://www.unecce.org/index.php?id=53381>.

При условии получения положительного отклика по итогам консультаций пленарной сессии КЕС будет предложено одобрить данное Руководство.

<sup>1</sup> Настоящий документ было решено издать позднее установленной даты его опубликования в связи с обстоятельствами, не зависящими от стороны, представившей документ.



## I. Глава 1: Введение

1. Поскольку число международных мигрантов продолжает расти, для общественности и директивных органов становится все более важным понимать миграционные потоки и воздействие миграции на отдельных лиц, семьи, общество и экономику. Во многих случаях ключевые вопросы касаются процесса обустройства мигрантов – как долго мигранты находятся в принимающей стране, как они интегрируются в принимающее общество и как со временем меняются их социально-экономические результаты.
2. В конечном счете, изучение миграции – это изучение изменений. Это начинается с изменения места проживания, но может распространяться и на изменения в правовом статусе или статусе проживания, а также на изменения в социально-экономических результатах. Миграция, интеграция и обустройство – это процессы, а не состояния, и результаты могут быть краткосрочными или долгосрочными. По причине их меняющегося характера эти темы хорошо подходят для изучения с использованием продольного подхода. Необходимость временной основы для статистики миграции была подчеркнута в Глобальном договоре о безопасной, упорядоченной и легальной миграции, в котором содержится призыв к сбору данных, «позволяющих осуществлять эффективный мониторинг и оценку выполнения обязательств во времени».
3. Применение традиционных методов сбора продольных данных (например, когортные исследования и панельные обследования) становится все более затруднительным по причине высоких затрат и истощения выборки. Соответственно, страны все чаще обращаются к альтернативным источникам данных для расчета продольных результатов.
4. Формирование статистики во многих странах эволюционировало под воздействием возросшей доступности и удобства использования административных данных. Массивы административных данных постоянно совершенствуются с точки зрения полноты, периодичности и качества собираемой информации. Благодаря все более широкому использованию административных данных и интеграции данных для формирования статистики миграции все больше стран способны создавать наборы продольных данных без чрезмерных затрат.
5. Если в прошлом продольные исследования миграции часто проводились обособленно или на разовой основе, то сегодня национальные статистические управления (НСУ) начинают все чаще включать их в свою регулярную программу работ по статистике миграции. Однако в настоящее время не существует международных руководящих принципов в отношении того, как разрабатывать источники продольных данных на основе интегрированных данных, как решать различные проблемы, связанные с такими проектами, и как распространять ключевые показатели и другие выводы.
6. Использование продольного подхода к измерению интеграции мигрантов неоднократно обсуждалось на совместных рабочих сессиях ЕЭК ООН и Евростата по статистике миграции (например, в 2012, 2014 и 2017 годах). В октябре 2017 года Рабочая сессия рекомендовала продолжить методологическую работу по этой теме, провести обзор национальной практики и разработать рекомендации, которые будут способствовать международной сопоставимости продольных данных.
7. В феврале 2018 года Бюро Конференции европейских статистиков (КЕС) учредило Целевую группу ЕЭК ООН по использованию продольных данных для статистики миграции. Задача Целевой группы заключалась в подготовке руководства по включению продольных данных в годовую статистику миграции и дополнению имеющихся перекрестных показателей. Настоящее Руководство отражает результаты работы этой Целевой группы.
8. Руководство основано на результатах проведенной в последнее время целевыми группами ЕЭК ООН методологической работы. В публикации «Измерение изменений в социально-экономических характеристиках мигрантов» (ЕЭК ООН, 2015)

пояснились преимущества использования продольных данных и странам было рекомендовано разработать методологии связывания данных для получения наборов продольных данных. В последовавшем за ней «Руководстве по интеграции данных для измерения миграции» (ЕЭК ООН, 2019b) было приведено несколько примеров того, как интеграция различных наборов данных ведет к компиляции продольных данных.

9. Руководство состоит из трех основных тематических 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 организованы как самостоятельные части, и читать их в сочетании с другими частями настоящего доклада не требуется. Важно то, чтобы пользователи могли извлечь выгоды из прочтения наиболее актуальных для их работы частей.

11. Глава 2 содержит обзор источников продольных данных для статистики миграции. В ней содержатся рекомендации по различным типам данных для статистики миграции и приводятся конкретные примеры из опыта разных стран. Отдельный раздел посвящен руководящим указаниям, касающимся, в частности, использования регистров населения для целей статистики миграции, включая конкретные проблемы и средства их решения.

12. Глава 3 содержит рекомендации по разработке наборов продольных данных для статистики миграции с использованием интегрированных данных со ссылкой на работу, проделанную предыдущими целевыми группами, в которых излагается методика, которой необходимо следовать начиная с разработки статистического плана и заканчивая распространением данных, включая описание общих проблем и конкретные примеры передовой практики различных стран. Охватываемые этапы включают в себя:

- разработку статистического плана;
- оценку и предварительную обработку исходных файлов;
- интеграцию данных для получения продольных данных;
- присвоение продольных индивидуальных идентификаторов;
- создание окончательной базы данных;
- распространение результатов.

13. Глава 4 содержит предлагаемый набор продольных показателей миграции наряду с описанием передовой практики распространения продольных результатов. По каждому представленному показателю приводится краткое описание проблем и передовой практики с наглядными примерами из различных стран.

14. Одна из сквозных тем настоящего Руководства касается того, что разработка и использование продольных данных для целей статистики миграции является сложным и трудным делом. Однако индивидуальные главы показывают, как эти ограничения могут быть устранены. В главе 4 рассказывается о том, как даже при ограничениях результирующей базы данных можно рассчитывать полезные показатели. Для удовлетворения новых и растущих потребностей в более глубоком понимании моделей миграции, интеграции и обустройства международных мигрантов важно, чтобы НСУ понимали имеющиеся источники данных, искали способы смягчить или устранить ограничения и предлагали подходы к распространению результатов, которые информируют об этих моделях.

15. Благодаря более широкому использованию интегрированных данных и статистических регистров многие национальные статистические управления оказались сидящими на горе неиспользованного потенциала в области продольной статистики миграции. В главе 3 странам предлагается методика разработки новых

источников продольных данных. В главе 4 приводятся примеры того, что можно делать с результатами.

16. В конечном счете, изучение миграции является естественным видом применения продольного анализа. Более широкое использование интеграции данных и статистических регистров открывает новые возможности для разработки и распространения продольных данных в целях статистики миграции. Несмотря на наличие проблем, выгоды от использования этих новых возможностей являются несоизмеримо большими. Изучение миграции – это изучение изменений, а для изучения изменений необходимы продольные данные.

## **II. Глава 2: Обзор источников продольных данных для статистики миграции**

### **A. Определение**

17. Под продольными данными понимается информация, которая собирается от одних и тех же единиц анализа, таких как физические лица или домашние хозяйства, с ходом времени. Продольный анализ способен уникальным и точным образом описывать индивидуальные траектории во времени. Продольные данные позволяют изучать и понимать жизненные события и переходы, как на протяжении жизни, так и между поколениями. Это особенно полезно для изучения международной миграции, поскольку обустройство в новой стране является длительным процессом. Продольные данные способны пролить свет на географические и социально-экономические результаты миграционного опыта. Продольные данные, собранные за длительный период, позволяют понять не только опыт мигрантов, но и опыт их детей.

18. В этой главе описаны различные типы продольных данных для изучения международной миграции. Кроме того, в ней рассматривается вопрос о том, как можно использовать перекрестные данные для получения продольной перспективы миграции и опыта групп мигрантов. Рассматриваются также вопросы взаимодополняемости и сопоставимости продольных и перекрестных показателей, и обсуждаются наблюдения, сочетающие оба подхода.

19. Примерами могут служить панельные обследования, встроенные в перекрестные обследования. Еще одним примером являются ретроспективные вопросы для получения продольных данных в рамках перекрестного обследования. Синтетические когорты в последовательных переписях или обследованиях имитируют продольные исследования в некоторых отношениях, но не отслеживают одни и те же единицы во времени.

20. И наконец, поскольку многие страны используют регистры населения для расчета показателей миграции, характеристики этих регистров, которые делают их более или менее пригодными для оценки продольных аспектов миграции, также подробно рассматриваются в ней.

### **B. Типы данных**

21. Приводимый ниже рисунок иллюстрирует концептуальную типологию продольной информации для статистики международной миграции.

Рис. 1

**Типология продольных данных и перспективы их использования для статистики миграции**



22. Как видно из рис. 1, данные можно классифицировать в зависимости от того, связывают ли они информацию по одним и тем же единицам во времени и, следовательно, являются продольными данными, или от того, обеспечивают ли они некоторую продольную перспективу. Под «интеграцией микроданных» понимаются данные, которые связаны на уровне записей во времени. «Сопоставления макроуровня» означают продольные *перспективы*, полученные путем сопоставления рядов динамики или использования синтетических когорт. В отличие от интеграции микроданных, сопоставления макроуровня не требуют, чтобы в выборку каждой волны сбора данных включались одни и те же лица.

23. Методологические подходы варьируются по континууму с различным балансом между использованием обследования (обозначено желтым цветом на рисунке) и административных данных (синий цвет). Методы, использующие как данные обследований, так и административные данные, обозначены зеленым цветом. К ним относятся варианты 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 меньше контроля над сбором данных и должны стремиться к максимальной по возможности синхронизации, чтобы данные соответствовали желаемым исследовательским или аналитическим целям.

24. Далее в этой главе будут рассмотрены два измерения времени, которые требуют тщательного учета при использовании административных данных. Первым из них является *эндогенная* операционализация времени при разработке плана исследования, которая подразумевает, что результаты наблюдения относятся к моменту фактического измерения. Второе – это *экзогенная* манифестация времени в наборе административных данных, которое подразумевает, что регистрация осуществляется в соответствии с административной практикой, регулированием, инфраструктурой и

качеством администрирования. Расхождения между эндогенной операционализацией и экзогенной манифестацией времени необходимо понимать, поскольку они являются источником потенциальной ошибки в данных. Ярким примером этого является проблема снятия с учета. Именно по причине несоблюдения этого требования эмиграция не регистрируется в административных данных, и многие источники, в частности регистры населения, страдают этой проблемой.

25. Новые перспективные панельные исследования предусматривают формирование выборки (отдельных лиц или домохозяйств, но могут также охватывать, например, предприятия или события). Наблюдаемая выборка обследуется через регулярные промежутки времени. Продольные данные панельных обследований также могут иметь форму новых ретроспективных данных. В случае такого типа данных при опросе респондентов задаются вопросы о событиях, произошедших в прошлом. Перспективное обследование с ретроспективным сравнением сочетает в себе эти два типа – ретроспективное сравнительное исследование в начале, за которым следует перспективное последующее по времени исследование.

26. Национальные статистические управления все чаще используют административные источники данных для статистического анализа. Анализ результатов мигрантов может полностью основываться на административных данных путем связывания этих данных для создания наборов продольных данных и сопоставления рядов динамики административных данных.

27. Комбинированный подход может принимать различные формы. Он может брать за основу информацию обследований и обновлять ее с помощью административных записей. В таблице 1 ниже описывается создание продольных данных путем связывания данных переписи и административных записей в рамках программы Продольные исследования Соединенного Королевства. Другой альтернативой является комбинация перекрестного и продольного подходов, используемая в Обследовании рабочей силы ЕС. Ретроспективные вопросы в рамках перекрестных обследований также дают продольную картину.

28. В отсутствие продольной информации, полученной в результате многократных измерений одних и тех же лиц, продольная перспектива может быть получена из синтетических когорт. Когорты можно определить как «группы людей, идущих вместе по времени»; они могут быть определены, например, на основе даты их рождения или периода их прибытия в страну.

### **Текущая практика**

29. Приводимый ниже рисунок иллюстрирует группировку по категориям текущей международной практики сбора продольных данных о мигрантах.

Рис. 2

**Типология текущей международной практики использования продольных данных и перспектив для статистики миграции**

Тип данных	Интеграция микроданных	Сопоставления макроуровня
	<b>Административные данные</b> (встроенный временной компонент, зависящий от внешних факторов)	Центральный регистр населения (ЦРН) Бельгии Связанные административные данные об иммиграции, беженцах и гражданстве Канады (IRCC) Канадская база продольных данных об иммиграции (IMDB) Центральный регистр иностранцев Германии АНАГРАФЕ (регистры населения) Италии Разрешения на проживание в Италии Регистр населения Испании (Падрон) Швейцарская база продольных данных (SLDD)
<b>Данные обследований</b> (встроенный временной компонент)	Обследование интеграции второго поколения Италии Продольные исследования Соединенного Королевства <div style="border: 1px dashed black; padding: 2px; margin: 5px 0;">             Микроперепись Германии              Обследование рабочей силы Соединенного Королевства           </div> Продольное обследование иммигрантов в Канаду	Сопоставления данных обследований с административными данными по международной миграции Соединенного Королевства Скользящие годовые оценки долгосрочной международной миграции на основе данных опроса международных пассажиров Соединенного Королевства

### С. Резюме

30. Продольные данные позволяют 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 интеграции мигрантов, поскольку продольные данные способны пролить свет на географические и социально-экономические результаты миграции. В этой главе приводится обзор различных источников продольных данных и конкретные примеры из опыта стран, включая использование продольных панельных обследований, административных данных, включая регистры населения, и комбинирование обследований и административных данных.

31. Не все НСУ будут иметь доступ к административным данным или финансирование для проведения продольных исследований по мигрантам, поэтому, возможно, потребуется использовать более прагматичный подход. Например, перекрестные обследования способны обеспечить получение продольной информации и перспективы во времени, а к существующим обследованиям могут быть добавлены ретроспективные вопросы. В отсутствие продольных данных можно также рассмотреть следующие наблюдения синтетических когорт во времени, все из которых подробно описаны в данной главе.

32. При наличии продольных данных следует учитывать проблемы и ограничения, связанные с этими данными, и тот факт, что они будут варьироваться в зависимости от каждого источника данных. Традиционные методы, такие как продольные панельные обследования, имеют преимущества в плане контроля, обеспечиваемого статистическим управлением, такие как выбор времени, охвата и метода измерения. Однако, поскольку мигранты, в целом, более мобильны в географическом отношении, чем немигранты, и с ними, как правило, труднее связаться для последующего опроса, эти источники данных, обычно, имеют меньшие размеры выборки, страдают истощением и могут не регулярно собирать данные. Все это может препятствовать их использованию для изучения краткосрочных или долгосрочных результатов групп мигрантов.

33. Существуют также ограничения, связанные с использованием административных данных, таких как регистры населения, или с процессом интеграции административных данных, в плане создания набора продольных данных. Интеграция микроданных, предполагающая связывание и последующие наблюдения лиц, представляет собой сложную задачу. Связывание индивидуальных записей во времени, будь то между обновлениями регистра населения или между регистром населения и другими административными данными, часто сопряжено с проблемами, даже когда используется личный идентификационный номер. Эти вопросы подробно описаны в данной главе, но ключевым ограничительным фактором является то, что эти данные не были собраны для статистических целей. Таким образом, теряется контроль над выбором времени, охватом населения и согласованием с концепциями и определениями миграции. Однако административные источники, как правило, позволяют формировать выборки большего размера по сравнению с традиционными продольными обследованиями с более регулярным последующим наблюдением и незначительным истощением из-за непредоставления ответов. В результате они, как правило, хорошо подходят для продольного изучения небольших совокупностей мигрантов.

34. Эти проблемы и ограничения рассматриваются в главе 3, в которой описывается передовая практика разработки набора продольных данных для статистики миграции и предпринимается попытка устранить ограничения, присущие различным источникам данных.

### **III. Глава 3: Как разработать набор продольных данных для статистики миграции с использованием интегрированных данных**

#### **A. Введение**

35. В настоящей главе содержатся рекомендации по разработке набора продольных данных для статистики миграции на основе интегрированных данных с учетом [Типовой модели производства статистической информации \(ТМПСИ\)](#) (ЕЭК ООН, 2019). Основное внимание в этой главе уделяется конкретным вопросам, связанным с интеграцией данных, продольными данными и статистикой миграции. В качестве дополнения следует использовать [Руководство ЕЭК ООН по интеграции данных для измерения миграции](#) (ЕЭК ООН, 2019), особенно в отношении интеграции данных.



36. Каждый раздел данной главы будет описывать один из этапов процесса, поясняя порядок осуществления каждого шага на конкретных примерах из практики различных национальных статистических управлений. Будут определены конкретные проблемы и предложены их решения.

37. Этапы разработки набора продольных данных для статистики миграции на основе интегрированных данных включают в себя:

- разработку статистического плана;
- оценку и предварительную обработку исходных файлов;
- интеграцию данных для получения продольных данных;
- присвоение продольных индивидуальных идентификаторов;
- создание окончательной базы данных;
- распространение результатов.

38. В данной главе изложены рекомендации по каждому из этих этапов. Они должны учитываться как в контексте разработки новой базы данных, так и в контексте обновления существующей базы данных (например, для добавления новых результатов путем интеграции данных).

39. Последовательность этих этапов имеет важное значение. Однако в зависимости от результатов более поздних этапов может потребоваться повторение этих этапов. Например, оценка исходных файлов (этап 2) может потребовать пересмотра статистического плана (этап 1). Также возможно, что вопросы качества данных, связанные с процессом интеграции, могут быть выявлены только в ходе оценки окончательной базы данных (этап 5). Это может потребовать проведения новой работы по интеграции данных (этап 3) и повторения последующих этапов.

40. Хотя необходимо учитывать все этапы, не во всех случаях будет необходим повтор. Например, обновление существующего проекта может не потребовать пересмотра этапа разработки статистического плана, а в случае формирования продольных данных из системы регистров этап интеграции данных может быть простым.

Рис. 3  
Этапы разработки набора продольных данных



## В. Система оценки ошибок для продольных данных, получаемых из интегрированных источников данных

41. Качество данных является постоянным предметом изучения на всех этапах. Исходные файлы должны выбираться так, чтобы они наилучшим образом соответствовали целям статистического плана, и качество потенциальных исходных файлов имеет важнейшее значение для принятия этого решения. Качество исходных файлов дополнительно оценивается на втором этапе и, по возможности, определяется с помощью различных методов. Сама интеграция данных требует оценки ошибок интеграции, а присвоение продольных индивидуальных идентификаторов требует определенного согласования результатов интеграции данных. После создания окончательной базы данных используются определенные методы для гармонизации концепций во времени и подготовки базы данных к анализу. И наконец, все соображения, касающиеся качества данных, должны документироваться, а пользователям до распространения результатов должны предоставляться соответствующие рекомендации.

42. Для этой цели полезной основой служит Базовая система оценки суммарной ошибки наблюдения, которая была определена для поддержки разработки статистического плана (Groves, 2004), (Zhang, 2012). Она содержит полезную таксономию различных потенциальных источников ошибки, относящихся, главным образом, к административным данным. Не все виды ошибок поддаются измерению. Но их необходимо учитывать при проектировании продольно-интегрированного

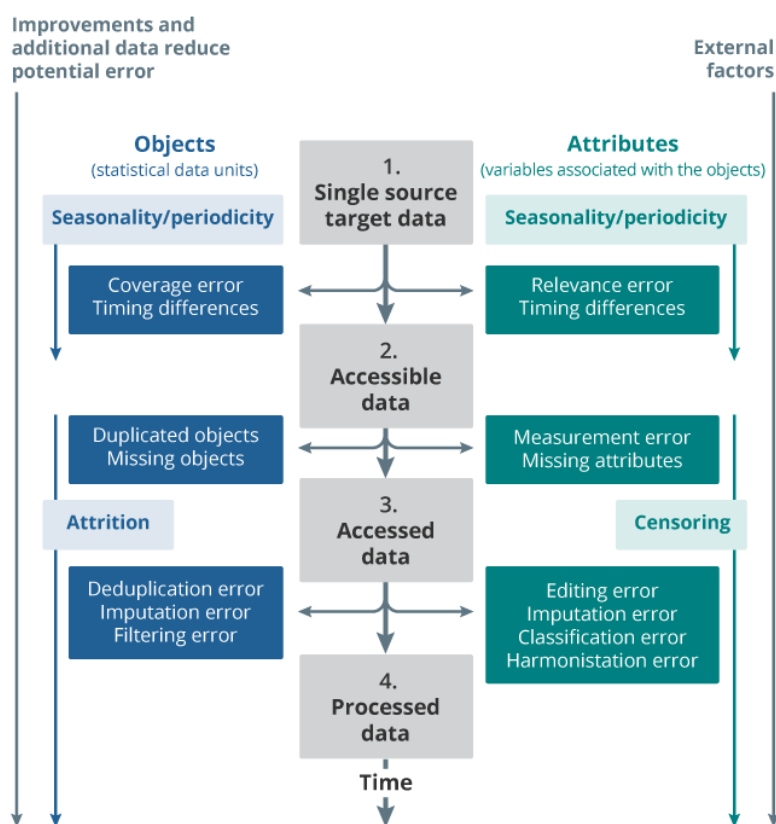
набора данных. Приводимый ниже текст является расширением системы, описанной в работе Zhang (2012), для более полного обсуждения источников ошибок при связывании административных источников с целью получения наборов продольных данных.

43. Эта базовая система, разработанная Соединенным Королевством (Blackwell & Rogers, 2019), предлагает поэтапный подход к пониманию качества данных, предусматривающий изучение отдельных наборов данных перед тем, как приступить к оценке ошибки при формировании наборов данных, созданных путем интеграции нескольких источников. Как и в системе, предложенной в работе Zhang (2012) и доработанной Статистическим управлением Новой Зеландии (Statistics New Zealand, 2016), объектом оценки являются ошибки, связанные как с объектами (записями) данных, так и с их признаками (переменными). Однако в административных данных эти ошибки обусловлены исходными наборами данных; признаки не проектируются как в случае обследования, а лишь меняется цель использования того, что уже было собрано. В основе системы лежат наборы данных, которые не были спроектированы – они являются тем, что они есть.

44. Рис. 4 и 5 иллюстрируют системы оценки ошибок для одного и множественных источников. В таблице 3, таблице 4 и таблице 5 указаны ошибки каждого этапа. В системе оценки для одного источника выделяются четыре различных состояния данных – целевые данные, доступные данные, отобранные данные и обработанные данные. Ошибки представлены в виде концептуального расхождения между данными в каждом из этих состояний. Целевые данные носят концептуальный характер, они представляют собой идеальные данные для сбора, в связи с чем расхождения между целевыми данными и доступными данными носят концептуальный характер.

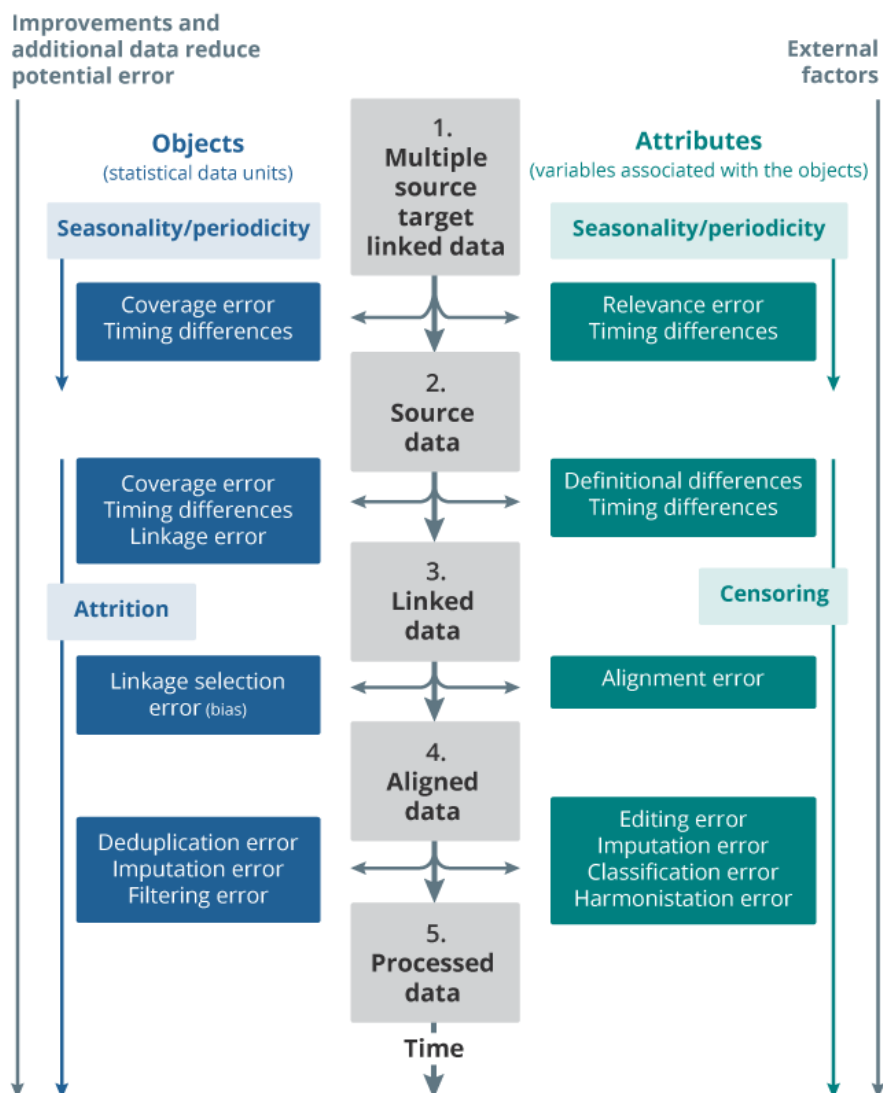
Рис. 4

#### Система оценки ошибок для одного источника



*Примечание:* Ось времени проходит по вертикали. Целевые данные представляют собой идеальные данные для сбора. Объекты (единицы статистических данных) относятся к строкам в источнике данных, а также к тому, что описывают эти строки данных. Признаки (переменные, описывающие объекты) относятся к колонкам в источнике данных, и к тому, что описывают эти колонки.

Рис. 5  
Система оценки ошибок для множественных источников



45. Ошибки распределяются между объектами и признаками: ошибки, связанные с объектами, относятся к единице, которую призваны описывать данные, будь то люди, события или предприятия. Ошибки, связанные с признаками, относятся к тому, что вы измеряете в отношении объектов. Ошибки, связанные как с объектами, так и с признаками, могут влиять друг на друга (что указано двойной стрелкой между ними). Не все типы ошибок будут применимы к каждому источнику данных, и могут быть выявлены другие источники ошибок, в частности в случае обработанных данных. В контексте настоящего доклада объектами, как правило, являются физические лица, определенные файлами когорт (или когорты, определенными файлами результатов), в то время под признаками понимаются результаты, определенные файлами результатов.

46. Ошибка распространяется через всю систему, хотя и не обязательно определяется на каждом этапе. Эта система используется для того, чтобы помочь исследователям Управления национальной статистики Соединенного Королевства в понимании продольно связанных административных данных, в том числе данных выездного контроля Министерства внутренних дел. Цель состоит в том, чтобы оценить возможности их использования для статистики международной миграции. Формирование данных выездного контроля начинается с передачи Министерству внутренних дел данных о пассажирах, предоставляемых коммерческими перевозчиками в портах въезда/выезда Соединенного Королевства (United Kingdom Home Office, 2019). Министерство внутренних дел обрабатывает эти данные и

контролирует их качество, связывая их с другими оперативными данными о гражданах тех стран, для которых предусмотрен визовый режим при въезде в Соединенное Королевство.

47. Этот анализ продемонстрировал, что в связанных административных данных ошибка не обязательно накапливается. Это связано с тем, что эксперты министерства внутренних дел пытаются устранять ошибки в данных по мере их прохождения через различные процессы. Хотя ошибка может накапливаться с течением времени, существует также возможность того, что компенсирующие ошибки могут усугубить проблему. Компенсирующие ошибки (например, импутация отсутствующих значений) могут быть не видны в поперечных распределениях данных, но могут объединяться в продольном направлении. Это поясняется в контексте системы оценки ошибок для множественных источников (рис. 6).

48. Эта система также охватывает продольную ошибку, создаваемую сбором данных с течением времени – сезонность/периодичность, истощение и цензурирование (таблица 4).

Таблица 1

**Система оценки ошибок для одного источника**

<b>Объекты<sup>1</sup></b>	<b>Признаки<sup>2</sup></b>
<p><i>Ошибка совокупности:</i></p> <p><i>Ошибка охвата</i></p> <p>Оценка объектов, которые не находятся в целевых данных, или отсутствие возможности доступа к объектам, находящимся в целевых данных.</p> <p><i>Расхождения во времени</i></p> <p>Объекты в идеальных целевых данных, которые недоступны из-за расхождения во временном интервале получения наблюдений.</p>	<p><i>Ошибка релевантности:</i></p> <p><i>Ошибка валидации</i></p> <p>Расхождение между идеальным измерением признаков, искомым объекте, и оперативным измерением, используемым для сбора данных о них.</p> <p><i>Расхождения во времени</i></p> <p>Концептуальное расхождение во времени измерения признаков между идеальными целевыми данными и доступными данными.</p>
<p><i>Ошибка отбора:</i></p> <p><i>Дублированные объекты</i></p> <p>Объекты, которые представлены более одного раза в отобранных данных.</p> <p><i>Отсутствующие объекты</i></p> <p>Объекты, которые теоретически доступны, но не присутствуют в отобранных данных.</p>	<p><i>Ошибка измерения:</i></p> <p><i>Ошибка измерения</i></p> <p>Ошибки, возникающие в результате неточной регистрации признаков.</p> <p><i>Отсутствующие признаки</i></p> <p>Признаки, отсутствующие в доступных данных (возможно по конкретным объектам или по всем объектам).</p>
<p><i>Ошибка обработки:</i></p> <p><i>Ошибка дедупликации</i></p> <p>Ошибки, возникающие в результате дедупликации объектов в отобранном наборе данных. Это может включать как дедупликацию объектов, которые на самом деле являются разными (ложноположительная ошибка), так и невозможность дедупликации одинаковых объектов (ложноотрицательная ошибка).</p>	<p><i>Ошибка обработки:</i></p> <p><i>Ошибка редактирования</i></p> <p>Ошибки, возникающие в результате редактирования значения признака. Это может включать в себя редактирование в результате валидации или проверки качества.</p> <p><i>Ошибка импутации</i></p> <p>Ошибки, возникающие в результате импутации отсутствующих значений признаков.</p>

Объекты <sup>1</sup>	Признаки <sup>2</sup>
<p><i>Ошибка импутации</i></p> <p>Ошибки, возникающие в результате импутации отсутствующих объектов.</p> <p><i>Ошибка фильтрации</i></p> <p>Ошибки, возникающие в результате выбора или отмены выбора отобранных объектов для идеального целевого набора.</p>	<p><i>Ошибка классификации</i></p> <p>Ошибки, возникающие в результате классификации значений по группам или расчета новых признаков.</p> <p><i>Ошибка гармонизации</i></p> <p>Ошибки, возникающие в результате гармонизации значений признаков с идеальной или целевой концепцией.</p>

<sup>1</sup> Они представляют собой единицы данных и могут быть событиями, операциями, лицами, домашними хозяйствами, фирмами или другими записями в административном наборе данных.

<sup>2</sup> Они представляют собой показатели или переменные, относящиеся к объектам/единицам данных.

Таблица 2

**Продольная ошибка – применяется как к событиям в одном источнике, так и к наборам продольных данных из множественных источников**

Объекты <sup>1</sup>	Признаки <sup>2</sup>
<p><i>Истощение:</i></p> <p>Утрата объектов или единиц наблюдения с течением времени. Происходит естественным путем в результате смерти (или незарегистрированной миграции). Происходит также в результате непроведения последующих наблюдений, отказа от участия в случае данных обследований, или в результате отсутствия информации или неустранения связи в административных источниках.</p> <p><i>Ошибка периодичности/сезонности:</i></p> <p>Объекты не имеют наблюдений, потому что сбор данных проводится недостаточно часто (периодичность) или не адекватно для учета сезонных колебаний в данных (сезонность).</p>	<p><i>Цензурирование:</i></p> <p>Проводится в случае, когда значение измерения или наблюдения известно лишь частично. Цензурирование справа – это когда объект удаляется из данных до конца периода наблюдения или не имеет события, представляющего интерес в период наблюдения. Цензурирование слева – это когда событие, представляющее интерес, уже произошло до начала периода наблюдения.</p> <p><i>Ошибка периодичности/сезонности:</i></p> <p>Измерение признаков во времени не является достаточно частым (периодичность) или адекватным для учета сезонных колебаний в данных (сезонность).</p>

<sup>1</sup> Они представляют собой единицы данных и могут быть событиями, операциями, лицами, домашними хозяйствами, фирмами или другими записями в административном наборе данных.

<sup>2</sup> Они представляют собой показатели или переменные, относящиеся к объектам/единицам данных.

49. Система оценки для множественных источников (таблица 3) предназначена для оценки ошибок при интеграции множественных источников данных. Она состоит из пяти этапов – целевые связанные данные, исходные данные, связанные данные, согласованные данные и обработанные данные. В контексте настоящего доклада эти этапы согласуются с этапами, охватываемыми данной главой, как показано на рис. 6.

Рис. 6

**Этапы интеграции множественных источников данных**

50. Ошибки представлены в виде концептуального расхождения между данными на каждом из этих этапов. Многие ошибки в системе оценки для множественных источников концептуально схожи с системой оценки для одного источника. Основное отличие заключается в том, что измеряются ошибки между исходными наборами данных и идеальными целевыми связанными данными (а не с целевыми данными), а также ошибки между исходными наборами данных, подлежащими связыванию. Эти концептуально схожие ошибки имеют одно и то же название в системе оценки для одного источника и системе оценки для множественных источников. Например, расхождения во времени, ошибка охвата, ошибка релевантности, ошибка импутации, ошибка выборки и ошибка обработки.

51. Целевые связанные данные отличаются от целевых данных для каждого отдельного источника. Целевые связанные данные, скорее всего, будут специфическими для группы объектов, которые будут измеряться путем связывания нескольких источников. На заключительном этапе (обработанные данные) обработка, возможно, происходила в одном источнике или в рамках множественных источников. Например, импутация могла уже быть проведена в одном источнике или была проведена после связывания.

Таблица 3

**Система оценки ошибок для множественных источников**

Объекты <sup>1</sup>	Признаки <sup>2</sup>
<i>Ошибка совокупности:</i> <i>Ошибка охвата</i> Наблюдение объектов, которые отсутствуют в целевых связанных данных, или отсутствие возможности доступа к объектам, находящимся в целевых связанных данных.	<i>Ошибка релевантности:</i> <i>Ошибка релевантности</i> Расхождения между идеальным измерением признаков, искомым объекте, и оперативным измерением, используемым для сбора данных о них.

Объекты <sup>1</sup>	Признаки <sup>2</sup>
<p><i>Расхождения во времени</i></p> <p>Объекты не имеют результатов наблюдения из-за концептуальных расхождений во времени регистрации между целевыми связанными данными и исходными данными.</p> <p><i>Ошибка охвата:</i></p> <p><i>Ошибка охвата</i></p> <p>Объекты не связаны между собой из-за расхождений в охвате объектов между источниками данных.</p> <p><i>Расхождения во времени</i></p> <p>Различия в наблюдаемых объектах в исходных наборах данных из-за того, что данные регистрировались в разное время.</p> <p><i>Ошибка связывания</i></p> <p>Ошибки, возникающие при ошибочном связывании объектов (ложноположительная ошибка) и невозможности связать объекты, которые должны были быть связаны (ложноотрицательная ошибка).</p> <p><i>Ошибка идентификации:</i></p> <p><i>Ошибка выбора связи (систематическая)</i></p> <p>Погрешности, возникающие в результате выбора связанных объектов (или отмены выбора несвязанных объектов) вследствие систематических ошибок, в связывании или из-за ошибки в разрешении конфликтующих связей.</p> <p><i>Ошибка обработки:</i></p> <p><i>Ошибка импутации</i></p> <p>Ошибки, возникающие в результате импутации отсутствующих объектов.</p> <p><i>Ошибка фильтрации</i></p> <p>Ошибки, возникающие в результате выбора или отмены выбора отобранных объектов для идеального целевого набора.</p>	<p><i>Расхождения во времени</i></p> <p>Концептуальное расхождение во времени измерения признаков между идеальными целевыми связанными данными и исходными данными.</p> <p><i>Ошибка соотнесения:</i></p> <p><i>Различия в определениях</i></p> <p>Различия в том, как признаки практически измерены в каждом из исходных наборов данных.</p> <p><i>Расхождения во времени</i></p> <p>Различия в значениях признаков связанного объекта между исходными наборами данных, вызванные регистрацией данных в разное время.</p> <p><i>Ошибка сопоставимости:</i></p> <p><i>Ошибка согласования</i></p> <p>Ошибки, возникающие в результате согласования конфликтующих значений признаков между источниками.</p> <p><i>Ошибка обработки:</i></p> <p><i>Ошибка редактирования</i></p> <p>Ошибки, возникающие в результате редактирования значения признака. Это может включать в себя редактирование в результате валидации или проверки качества.</p> <p><i>Ошибка импутации</i></p> <p>Ошибки, возникающие в результате импутации отсутствующих значений признаков.</p> <p><i>Ошибка классификации</i></p> <p>Ошибки, возникающие в результате классификации значений по группам или расчета новых признаков.</p>



Объекты <sup>1</sup>	Признаки <sup>2</sup>
<p><i>Истощение:</i></p> <p>Утрата объектов или единиц наблюдения с течением времени. Происходит естественным путем в результате смерти (или незарегистрированной миграции). Происходит также в результате непроведения последующих наблюдений, отказа от участия в случае данных обследований, или в результате отсутствия информации или неустановления связи в административных источниках.</p> <p><i>Ошибка периодичности/сезонности:</i></p> <p>Объекты не наблюдаются, потому что сбор данных производится недостаточно часто (периодичность) или неадекватно для учета сезонных колебаний в данных (сезонность).</p>	<p><i>Ошибка гармонизации</i></p> <p>Ошибки, возникающие в результате гармонизации значений признаков с идеальной или целевой концепцией.</p> <p><i>Цензурирование:</i></p> <p>Проводится в случае, когда значение измерения или наблюдения известно лишь частично. Цензурирование справа – это когда объект удаляется из данных до конца периода наблюдения или не имеет события, представляющего интерес в период наблюдения. Цензурирование слева – это когда событие, представляющее интерес, уже произошло до начала периода наблюдения.</p> <p><i>Ошибка периодичности/сезонности:</i></p> <p>Измерение признаков во времени не является достаточно частым (периодичность) или адекватным для учета сезонных колебаний в данных (сезонность).</p>

<sup>1</sup> Они представляют собой единицы данных и могут быть событиями, операциями, лицами, домашними хозяйствами, фирмами или другими записями в административном наборе данных.

<sup>2</sup> Они представляют собой показатели или переменные, относящиеся к объектам/единицам данных.

52. Цель применения системы оценки ошибок в административных данных заключается в том, чтобы обеспечить оптимизацию качества данных на всех этапах статистического проекта. Первоначально необходимо выявить и изучить источники ошибок для принятия статистических проектных решений при проведении дальнейшего связывания. Однако важно вновь изучать различные источники ошибок также на всех последующих этапах. В перечнях, приведенных в таблицах 4 и 5, предложен ряд показателей качества. Для НСУ может оказаться целесообразным измерять не каждый показатель, а определить основные показатели для включения в отчет о качестве (см. раздел 3.4.3 в полной версии Руководства).

53. Существует взаимное влияние между различными источниками ошибок. Например, могут существовать зависимости между ошибками связывания, охвата и импутации. Записи с низким качеством данных, возможно, из-за ошибки измерения, также могут с большим трудом поддаваться связыванию. Это может быть обусловлено качеством идентификаторов, используемых для связывания. Одним из вариантов является разработка сложных методов связывания записей, чтобы минимизировать количество ложных отрицательных совпадений (и, следовательно, увеличить допуск ложно положительных) и максимизировать охват набора связанных данных. Но есть вероятность того, что признаки, которые относятся к этим объектам, также страдают низким качеством и будут генерировать либо пропуски в полях признаков, либо требовать импутации. Импутация часто нежелательна в продольных данных, так как она может привести к ложным результатам. Избежание пропусков в данных и импутации может быть главной задачей.

54. Существует также взаимное влияние между ошибками в объектах (или когортах) и в признаках (или результатах), а также между одним и множественными наборами исходных данных. Отсутствующие или неправильно зарегистрированные

данные признаков могут влиять на возможность дедупликации или связывания записей на этапе одного источника. Это, в свою очередь, влияет на возможность связывания объектов и, следовательно, создает ошибки в наборе данных из множественных источников.

#### **IV. Глава 4: Распространение регулярной статистики миграции из источников продольных данных**

55. В этой главе приводятся рекомендации по распространению регулярной статистики миграции с использованием продольных источников данных. В ней будут описаны различные продольные показатели, которые могут быть рассчитаны для статистики миграции. Эти показатели охватывают темы, связанные с самим процессом миграции, а также с социально-экономическими результатами мигрантов во времени. В ней излагаются конкретные проблемы, связанные с каждым тематическим показателем, и предлагаются возможные подходы к их решению на примерах из опыта разных стран. Проблемы и примеры касаются использования административных данных или интегрированных источников данных.

56. С опорой на публикацию ЕЭК ООН «Измерение изменений в социально-экономических условиях мигрантов» (ЕЭК ООН, 2015) в этой главе предлагаются показатели продольных результатов, т. е. результатов, достигнутых на протяжении определенного периода времени.

57. В контексте настоящего доклада «показатели» относятся к статистике, представляющей общий интерес, которая может формироваться на регулярной основе. Они отличаются от статистических данных, генерируемых для конкретных исследований. Показатели могут использоваться для информирования процесса принятия решений правительством по политике или программам, но могут также обслуживать потребности и других заинтересованных сторон, включая общинные организации и самих мигрантов.

58. В качестве отправной точки перечислены ключевые тематические показатели для основных областей, представляющих интерес в плане продольных данных. В рамках каждой темы может быть предложено несколько конкретных показателей. В то время как ключевые тематические показатели могут относиться к общим терминам, таким как «продолжительность времени до события», конкретные показатели могут описывать долю мигрантов, которые пережили это событие после X лет, среднее время до события или другие временные показатели. Отдельные наборы показателей могут быть определены для подмножеств мигрантов (например, просителей убежища, временных резидентов и т. д.) и могут быть подвергнуты перекрестной табуляции по демографическим переменным или переменным социально-экономического статуса (например, возраст, пол, страна рождения, образование, условия приема и т. д.).

59. Для целей настоящей главы термин «мигрант» будет использоваться, как правило, в отношении набора тематических показателей и видов анализа. Однако некоторые показатели могут применяться только к определенным подгруппам населения. Например, не все мигранты будут охвачены показателями, связанными с изменением правового статуса или гражданства. Важно учитывать группы населения, представляющие интерес для каждого набора показателей, в зависимости от обстоятельств. В примерах четко определены конкретные группы мигрантов.

60. Будет дано описание каждого показателя и рассмотрены вопросы, связанные с его разработкой. Будут приведены практические примеры, в том числе случаи, когда практический показатель отклоняется от желаемого из-за ограничений источников данных. Цель данной главы – привести примеры описательной статистики, которая может использоваться в качестве регулярных продольных показателей. Более сложные продольные аналитические методы, такие как анализ долговечности или обобщенные оценочные уравнения, в данной главе не рассматриваются, но могут быть использованы для получения более глубокого представления о продольных

результатах мигрантов. Наборы показателей, определенные в этой главе, могут быть использованы в качестве переменных результатов, опирающихся на эти методы.

61. При рассмотрении вопроса о том, какие показатели могут быть рассчитаны, важно учитывать общие ограничения доступных источников данных. В главе 2 описываются некоторые сильные стороны и ограничения обследований, административных данных и регистров, а в главе 3 более подробно рассматриваются ограничения, которые могут существовать в случае интегрированных источников данных.

62. Одной из главных проблем является определение единицы времени для расчета продольных показателей. Если данные наблюдаются не в реальном времени, а в периодических (например, годовых) сегментах, то время не может быть измерено с достаточной точностью. Может возникнуть еще одна сложность при определении того, когда начать отсчет – с первого приезда мигранта, с момента, когда он станет постоянным жителем и т. д.? Эти вопросы должны быть рассмотрены в отношении всех этих показателей.

63. Продольные показатели можно разделить на три основные категории:

- типы миграции;
- социально-экономические результаты;
- семейная миграция.

64. По каждому показателю в таблице указываются страны, которые предоставили конкретные примеры этой статистики вместе с имеющимися у них данными на продольной или перекрестной основе. Примечательно, что наличие показателей не обязательно означает регулярную публикацию этой статистики. Хотя многие страны могут располагать доступными данными, эти показатели могут составляться нерегулярно, а, например, в рамках отдельной или одноразовой публикации.

Таблица 4

**Ключевые продольные показатели**

<i>Категория</i>	<i>Тема</i>	<i>Показатель</i>
Типы миграции	Продолжительность пребывания в стране	Доля мигрантов, все еще присутствующих в стране по истечении X времени
		Среднее время пребывания до выезда мигрантов из страны
		(Кумулятивная) продолжительность пребывания с момента первого прибытия
		Продолжительность пребывания с момента последнего прибытия
		Средняя продолжительность пребывания (в случае более одного пребывания из расчета на душу)
		Доля мигрантов, все еще присутствующих на субнациональном географическом уровне по истечении X времени
		Среднее время пребывания до выезда мигрантов с субнационального географического уровня

<i>Категория</i>	<i>Тема</i>	<i>Показатель</i>
	Продолжительность пребывания на субнациональном географическом уровне	Из общего числа остающихся в стране доля мигрантов, все еще присутствующих на субнациональном географическом уровне по истечении X времени  В отношении остающихся в стране среднее время пребывания до выбытия мигрантов с субнационального географического уровня
	Время до изменения места жительства или правового статуса, включая гражданство	Доля временных или краткосрочных резидентов, которые становятся постоянными или долгосрочными резидентами по истечении X времени  Среднее время до момента, когда временные или краткосрочные резиденты становятся постоянными или долгосрочными резидентами  Доля просителей убежища, которым по истечении X времени разрешено постоянное проживание  Среднее время до того, как просителям убежища разрешают постоянное проживание  Доля мигрантов, родившихся за рубежом, которые приобретают национальное гражданство по истечении X времени  Среднее время до приобретения мигрантами, родившимися за границей, национального гражданства
	Показатели круговой миграции	Среднее количество пребываний из расчета на одного «кругового мигранта» (требует дальнейшего определения в плане цензурированного наблюдения, периода времени, требуемой минимальной продолжительности пребывания и т. д.)  Средняя продолжительность пребывания из расчета на одного «кругового мигранта»
Социально-экономические результаты	Владение языком	Доля мигрантов, способных говорить на официальных языках страны по истечении X времени  Среднее время до момента, когда мигранты смогут говорить на официальных языках страны  Доля мигрантов, говорящих на официальных языках страны дома по истечении X времени  Среднее время до момента, когда мигранты начинают говорить на официальных языках страны дома

<i>Категория</i>	<i>Тема</i>	<i>Показатель</i>
Социально-экономические результаты	Владение жильем	Доля мигрантов, ставших собственниками своего жилья по истечении X времени
		Среднее время до момента, когда мигранты впервые становятся собственниками своего жилья в новой стране проживания
		Тип владения жильем (например, аренда или собственность) мигрантов во времени
		Доля мигрантов, проживающих в субсидируемом или финансируемом государством жилье по истечении X времени
	Занятость	Тип владения в разбивке по типу жилой недвижимости (например, квартира, двухквартирный дом и т. д.) по истечении X времени
		Доля мигрантов, являющихся занятыми по истечении X времени
		Среднее время до момента, когда мигранты впервые устроятся на работу в новой стране проживания
		Полная или частичная занятость мигрантов во времени
		Статус рабочей силы мигрантов во времени
		Высшее образование
Доход	Посещение школы мигрантами во времени, в разбивке по возрасту	
	Средний и медианный совокупный доход по истечении X времени	
	Средний и медианный общий доход от занятости после X времени	
	Доля мигрантов с доходом от занятости по истечении X времени	
	Доля мигрантов с доходом от самозанятости и доходом от занятия предпринимательской деятельностью по истечении X времени	
		Доля мигрантов, получающих социальную или государственную помощь по истечении X времени

<i>Категория</i>	<i>Тема</i>	<i>Показатель</i>
		<p>Время до момента, когда средний или медианный совокупный доход мигрантов сравнивается со средним или медианным совокупным доходом населения в целом</p> <p>Время до момента, когда средний или медианный доход мигрантов сравнивается от занятости сравнивается со средним или медианным совокупным доходом от занятости населения в целом</p> <p>Доля мигрантов с низкими доходами по истечении X времени</p>
Владение предприятием/ предпринимательская деятельность		<p>Доля мигрантов, являющихся самозанятыми или владеющими предприятием по истечении X времени</p> <p>Среднее время до момента, когда мигранты становятся самозанятыми или владельцами предприятия</p> <p>Количество лиц, работодателями которых являются мигранты, по истечении X времени</p> <p>По аналогии с показателями, касающимися статуса рабочей силы, данные налогового учета могут служить косвенным показателем предпринимательской деятельности, показывая долю мигрантов, имеющих доход от самозанятости или предпринимательской деятельности. В противном случае могут использоваться административные записи о владении предприятием</p> <p>Увязка мигрантов-предпринимателей с информацией об их работниках может быть более сложным делом. Однако при наличии административных записей, связывающих работников с фирмами и владельцев – с фирмами, такой анализ также возможен</p> <p>Канадская база данных о динамике работодатели–работники (КБДРР) представляет собой среду связывания, основанную на 12 административных файлах-источниках, включая Продольную базу данных об иммиграции (ПБДМ), и содержащую характеристики отдельных лиц, фирм и владельцев предприятий. В среде связывания пользователи должны интегрировать различные компоненты в соответствии со своими аналитическими потребностями, используя анонимизированные уникальные идентификаторы в каждом файле.</p>

<i>Категория</i>	<i>Тема</i>	<i>Показатель</i>
		КБДРР позволяет проводить анализ фирм и физических лиц во времени, предоставляя ответы на вопросы, например, о первоначальной принадлежности фирмы и росте доходов, о принадлежащих иммигрантам предприятиях, а также о траекториях фирм, принадлежащих иммигрантам
	Здравоохранение	<p>Доля мигрантов, обратившихся за медицинской помощью в систему здравоохранения принимающей страны по истечении X времени (или времени до обращения)</p> <p>Доля мигрантов, подвергшихся госпитализации в течение X лет с момента прибытия в принимающую страну, в целом или в связи с избранными заболеваниями, например, хронического или инфекционного характера (или время до первой госпитализации)</p> <p>Доля мигрантов, умерших с момента прибытия в принимающую страну по причине различных заболеваний (по истечении X времени) (или время 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 X времени пребывания в стране</p>

## V. Глава 5: Выводы и рекомендации

### A. Выводы

65. Изучать миграцию – значит изучать изменения: изменение места жительства, изменение правового статуса, изменение социально-экономических результатов. Кроме того, связанными с интеграцией и обустройством вопросы касаются процессов, а не состояний. Результаты мигрантов могут быть краткосрочными или долгосрочными.

66. Для формирования статистики миграции и по смежным вопросам необходимо использовать данные, измеряющие изменения. Продольные данные обеспечивают средства для понимания изменений на индивидуальном уровне. Помимо простого наблюдения за изменениями в месте жительства или результатах, отслеживание индивидуумов во времени позволяет лучше понять факторы, которые могут влиять на изменения. Эти факторы могут действовать на макроуровне, например изменения в политике и экономические сдвиги, или на микроуровне, например получение образования или профессиональной подготовки и приобретение гражданства.

67. Исследователи и официальные статистики все чаще используют методы, основанные на продольных подходах, для лучшего понимания миграционных потоков и динамики интеграции в принимающих странах. Продольные подходы лучше всего подходят для измерения того, как изменяется разрыв между иммигрантами и коренным населением со временем, прошедшим с момента иммиграции, например с точки зрения заработной платы или безработицы.

68. Благодаря увеличению объема доступных административных данных страны сегодня могут более эффективно использовать такие методы, как интеграция данных, с тем чтобы иметь возможность разрабатывать подробные продольные данные для статистики миграции. Анализ интегрированных административных данных может позволить дать ответы на вопросы, требующих наблюдения крупных выборочных совокупностей для получения большого числа подробных данных о труднодоступных группах населения, а также получить фактические данные с высокой степенью внешней общезначимости и применимости для целей разработки политики. Такой подход позволяет вести продольное наблюдение за результатами миграции, интеграции или обустройства, не увеличивая при этом нагрузку по сбору данных с помощью обследований.

69. В настоящем Руководстве показано, каким образом можно разрабатывать и использовать продольные данные в целях статистики международной миграции. Глава 2 содержит обзор типов данных, начиная с данных панельных обследований и заканчивая регистрами населения. Глава 3 содержит рекомендации в отношении разработки источников продольных данных на основе интегрированных данных. И наконец, в главе 4 предлагаются основные продольные показатели, связанные с изучением международной миграции, а также описывается передовая практика в области распространения.

70. Хотя продольные данные обеспечивают важнейшую и уникальную возможность для изучения миграции, интеграции и обустройства в качестве процессов, в настоящем Руководстве признаются проблемы и ограничения, связанные с такими данными. Эти проблемы варьируются в зависимости от типа источника данных, как показано в главе 2.

71. Традиционные методы, такие как продольные панельные обследования, имеют преимущества в плане контроля, обеспечиваемого статистическим управлением в отношении охвата и измерения. Однако по практическим соображениям эти источники, как правило, имеют меньшие размеры выборки, страдают истощением и нерегулярностью сбора. Эти ограничения могут препятствовать их использованию для изучения краткосрочных или долгосрочных результатов малых совокупностей (например, подгрупп мигрантов).

72. С другой стороны, существуют также ограничения, связанные с использованием интегрированных административных данных. В частности, эти данные не были собраны для статистических целей, что ведет к утрате контроля над охватом и измерением. Кроме того, процесс интеграции может привести к появлению ошибок, как это обсуждалось в главе 3. Однако эти методы, как правило, позволяют получать данные по выборкам большого размера с более регулярной периодичностью и незначительным истощением по причине непредоставления ответов. В результате они, как правило, хорошо подходят для продольного изучения небольших совокупностей мигрантов.

73. Существуют методы, позволяющие смягчить, устранить или обойти ограничения источников данных, с тем чтобы распространяемые данные по-прежнему служили целям, изложенным в статистическом плане. В главе 3 приводятся различные примеры передовой практики разработки источников данных, а в главе 4 показано, как можно получить полезные показатели, несмотря на ограничения, связанные с результирующей базой данных.

74. Важно не игнорировать эти проблемы, но также и не страшиться их. Главное понять, подходит ли та или иная статистика для конкретной цели. Использование интегрированных данных для продольной статистики миграции всегда будет сопровождаться проблемами, но они также открывают беспрецедентные



аналитические возможности, проливая свет на подробные продольные результаты мигрантов и их влияние на принимающие страны. Эти источники данных не всегда дают идеальные оценки перекрестных точек, но они предлагают нечто более богатое – историю.

75. Потенциал продольных данных для статистики миграции четко иллюстрируют примеры, приводимые в главе 4. В конце концов, преимущества продольных данных для статистики миграции говорят о том, что эти проблемы должны быть преодолены.

## **В. Рекомендации**

76. Для лучшего понимания миграции и, в частности, процессов интеграции и обустройства рекомендуется расширить использование продольных данных в целях статистики миграции.

77. При разработке продольного набора данных рекомендуется следовать каждому из шести этапов, описанных в главе 3, в предложенном порядке. В зависимости от результатов, возможно, придется вернуться к одному из более ранних этапов, и в этом случае последующие этапы придется повторить. Эти этапы включают в себя:

- разработку статистического плана;
- оценку и предварительную обработку исходных файлов;
- интеграцию данных для получения продольных данных;
- присвоение продольных индивидуальных идентификаторов;
- создание окончательной базы данных;
- распространение результатов.

78. Рекомендуется, чтобы страны регулярно составляли продольные показатели для миграционной статистики, описанные в главе 4.

## **С. Области будущей работы**

79. В этом докладе был охвачен широкий круг тем, связанных с миграцией, для которых могут быть полезны продольные данные. Некоторые темы были рассмотрены более подробно, особенно в главе 4 при обсуждении предлагаемых показателей. Однако было сочтено, что другие выходят за рамки настоящего Руководства.

80. В частности, необходимо будет провести дальнейшее исследование вопроса о том, как измерять миграцию, связанную с семьей. В силу своей сложности эта тема может потребовать более целенаправленного изучения. Это может включать в себя изучение таких подтем, как:

- модели миграции семей;
- семьи мигрантов как единицы анализа (например, результаты обустройства и интеграции семей мигрантов с течением времени);
- воссоединение семей и члены семьи за рубежом;
- условия проживания мигрантов в принимающей стране (например, многосемейные домашние хозяйства, домашние хозяйства, состоящие из нескольких поколений).

81. Еще одной темой, заслуживающей дополнительно внимания, является вопрос о субнациональных моделях обустройства, рассмотренный в разделе 4.1.1.2 полного варианта Руководства, а именно – продолжительность присутствия на субнациональном географическом уровне. Однако страны проявляют все больший интерес к изучению состава жилых кварталов и поощрению мигрантов к тому, чтобы они селились в нетрадиционных местах (например, по экономическим причинам). Понимание низкоуровневых географических моделей может иметь важное значение

для заинтересованных сторон, таких как поставщики услуг и школы. Однако, как обсуждается в настоящем Руководстве, в частности в главе 4, существуют уникальные проблемы, связанные с этим типом исследований. В частности, границы изменяются во времени, и иногда их перенос в краткосрочной перспективе может рассматриваться как изменение в географии, в то время перенос в долгосрочной перспективе может не менять географию.

82. Еще одной новой областью являются способы распространения информации статистическими управлениями. Некоторые инновационные подходы описываются в главе 4. В связи с появлением новых инструментов распространения и более сложных для распространения данных статистическим управлениям следует вновь рассмотреть вопрос о том, как наилучшим образом представлять выводы различным аудиториям. В случае статистики миграции это является особенно важной задачей. По причине наблюдающегося в последнее время роста международной миграции в ряде стран ведется широкая общественная дискуссия, которая требует фактических данных о моделях миграции и ее воздействии на отдельных лиц, семьи и общество. Поскольку статистика миграции является сложной по своему характеру, существует также риск ошибочной интерпретации статистических результатов. Более подробные рекомендации в отношении того, как представлять результаты различным аудиториям при одновременном уважении и признании ограничений, связанных с качеством данных, могли бы помочь национальным статистическим управлениям в решении этой возникающей проблемы.

---