

Distr.: General
19 января 2023 г. 15:11:01

English

Европейская экономическая комиссия

Конференция европейских статистиков

Группа экспертов по статистике миграции

Женева, Швейцария, 26-28 октября 2022 года

Пункт А предварительной повестки дня

Положительные изменения в использовании административных данных для статистики миграции

Оценка охвата для Интегрированной базы данных по международной миграции (ИБДММ) Бюро переписи США

Записка Бюро переписи США

Аннотация

В последние годы Отдел международной миграции Бюро переписи населения США изучал использование административных данных для улучшения оценок потоков международной миграции иностранцев, в результате чего была создана связанная база данных под названием Интегрированная база данных по международной миграции (ИБДММ). Недостаток ИБДММ заключается в том, что она ограничена лицами, присутствующими в наборах данных федеральных административных записей, полученных Бюро переписи населения, что приводит к пробелам в учете лиц иностранного происхождения. Мы предполагаем, что ИБДММ занижает показатели или упускает определенные группы иностранцев, не присутствующие в связанных источниках данных, в том числе работающих мигрантов, которые не подавали налоговые декларации, неработающих иждивенцев, не заявленных в качестве освобожденных от уплаты налогов, иностранных студентов, посетителей по обмену и нелегальных мигрантов. Обследование американского сообщества (ОАС) — это крупное ежегодное обследование домохозяйств, проводимое Бюро переписи населения США, которое теоретически должно охватывать группы населения, не учтенные в ИБДММ. Чтобы изучить ограничения охвата ИБДММ, мы сопоставляем отдельные записи ОАС с ИБДММ.

В настоящем документе представлены результаты зондирующего исследования для оценки охвата ИБДММ. Здесь также оцениваются численность и характеристики респондентов ОАС, не включенных в ИБДММ, в отличие от тех, кто присутствует в ИБДММ. Кроме того, это исследование призвано оценить качество данных ОАС и ИБДММ, поскольку мы можем сравнить результаты для связанных лиц и домохозяйств в каждом источнике данных. Результаты сравнения дают нам информацию о потенциале использования ОАС для коррекции ограниченного охвата ИБДММ.

*Подготовили Джейсон Шахтер, Эстер Миллер и Анджелика Менчака

ПРИМЕЧАНИЕ: Обозначения в настоящем документе не подразумевают выражения какого-либо мнения Секретариата Организации Объединенных Наций в отношении юридического положения любой страны, территории, города или края или их властей или в отношении делимитации ее границ.

I. Введение

1. Отделение международной миграции Бюро переписи населения США ежегодно производит оценки иммиграционных потоков иностранцев в США с указанием демографических данных (возраст, пол, раса и латиноамериканское происхождение) в географическом разрезе страны, штата и округа. Эти оценки формируются с использованием Обследования американского сообщества (ОАС) в качестве основного источника данных, хотя оценки по итогам обследования имеют ограничения, такие как повышенная дисперсия, особенно в субнациональных регионах, и запаздывающее измерение миграционных событий. Чтобы помочь преодолеть эти ограничения, Бюро переписи населения разработало альтернативный источник данных под названием «Интегрированная база данных по международной миграции» (ИБДММ), который включает доступные административные данные из системы социального обеспечения и налоговой отчетности для оценки международной миграции.
2. Одним из недостатков ИБДММ является то, что она ограничивается лицами, присутствующими в наборах данных федеральных административных записей, из-за чего лица, родившиеся за границей и не включенные в эти записи, не учитываются. Это означает, что ИБДММ, вероятно, занижает количество или упускает определенные группы иностранцев, например работающих мигрантов, которые не подавали налоговые декларации, неработающих иждивенцев, не заявленных в качестве освобожденных от уплаты налогов, иностранных студентов, посетителей по обмену и нелегальных мигрантов. ОАС представляет собой крупное ежегодное обследование домохозяйств, проводимое Бюро переписи населения США. Обследование репрезентативно для всего постоянного населения и, таким образом, теоретически должно охватывать эти недостающие в ИБДММ группы населения. Хотя вероятно, что в ОАС тоже недостаточно представлены некоторые из этих отсутствующих групп иностранного происхождения, мы все же можем проанализировать ограничения охвата ИБДММ, сопоставив отдельные записи ОАС с ИБДММ. Это позволяет нам сравнивать тех, кто присутствует и в ОАС, и в ИБДММ («совпадение ОАС» или «совпадение ОАС-ИБДММ»), с теми, кто присутствует в ОАС, но не в ИБДММ («только ОАС» или «несовпадение ОАС-ИБДММ»). В дополнение к этому сравнению для лиц, представленных как в ОАС, так и в ИБДММ, мы можем оценить качество данных по ряду переменных, общих для обоих наборов данных, включая возраст, пол, статус гражданства, год въезда и текущее место жительства. Наконец, на основе этих выводов в настоящем документе обсуждаются возможности использования ОАС для коррекции результатов ИБДММ для улучшения соответствующих оценок.

II. Обзор ИБДММ

3. ИБДММ создается путем установления связей между источниками административных данных, которые можно использовать для оценки показателей иммиграции лиц иностранного происхождения. Хотя существует много административных источников данных, которые потенциально могут быть использованы в ИБДММ, например, те, которые ведет Министерство национальной безопасности (МНБ), в настоящее время мы ограничены источниками данных, уже доступными Бюро переписи населения: система Numident Службы социального обеспечения (ССО) и информация о налоговых декларациях от Налоговой службы (НС). Привязка данных осуществляется путем сопоставления

уникальных персональных ключей идентификации (ПКИ), которые присваиваются каждому отдельному человеку в наборах данных. ПКИ проще всего создать с помощью прямого сопоставления зашифрованных номеров социального страхования (НСС), но они также создаются путем вероятностного сопоставления информации об имени, поле, возрасте и адресе.

A. Numident (Цифровая система опознавания)

4. Numident — это набор данных микрозаписей, который объединяет записи НСС ССО с записями о смертях Бюро переписи населения. Он включает данные о демографических характеристиках, месте рождения и статусе гражданства. Однако он не включает адресных данных. НСС можно легко анонимизировать с помощью ПКИ, которые позволяют связывать наборы данных. Поскольку большинство зарегистрированных иммигрантов в США подают заявки на получение НСС, система Numident была выбрана в качестве основы ИБДММ для первоначальной интеграции данных и исследования
5. Данные Numident предоставляются ежеквартально и содержат записи обо всех лицах, которые когда-либо получали НСС. Помимо родившихся в стране, сюда входят заявки на НСС от родившихся за границей. Лица, родившиеся за границей, которые либо имеют разрешение на работу, либо стали натурализованными гражданами, имеют право на получение НСС. Используя комбинацию статуса гражданства, места рождения и даты создания записи, мы можем идентифицировать мигрантов, родившихся за границей, на национальном уровне по демографическим характеристикам. Связывание данных Numident с другими источниками может дать нам дополнительную информацию, такую как место жительства, что позволит нам создать субнациональные оценки численности лиц, родившихся за границей. Так мы получаем информацию еще и о «признаках жизни», которые дают нам дополнительную уверенность в том, что лицо, имеющее номер социального обеспечения, переехало в США на период времени, необходимый для получения вида на жительство.

B. Налоговая информация от Налоговой службы

6. НС представляет в Бюро переписи населения информацию на основе формы 1040 каждые четыре недели. Хотя эти данные не включают демографические характеристики, они включают информацию об адресах и ПКИ для основного заявителя, супругов и иждивенцев. Там, где это возможно, адресные данные связаны с Идентификатором главного адресного файла. Данные НС не включают информацию о статусе лиц иностранного происхождения, которая должна поступать из связанных файлов Numident. Кроме того, они не включают информацию о лицах, не подающих налоговые декларации (либо из-за отсутствия достаточного дохода, либо из-за того, что не заявляют о доходах). НСС, включенные в налоговые данные, обеспечивают возможность прямого сопоставления физических лиц с данными системы Numident.
7. Кроме того, можно идентифицировать налоговые данные, в которых используются индивидуальные идентификационные номера налогоплательщика (ИИНН), — уникальные идентификаторы, используемые физическими лицами, не имеющими НСС, для подачи налоговых деклараций. ИИНН выдаются только лицам, не являющимся гражданами США; таким образом, чтобы присвоить статус родившихся за границей, нет необходимости связывать этих людей с системой Numident. Хотя предполагается, что эта совокупность включает мигрантов, не имеющих разрешения на работу, она также включает неработающих иждивенцев легальных рабочих-мигрантов. Изменения в налоговом законодательстве 2017 года привели к резкому сокращению сообщаемых ИИНН, поскольку супруги и иждивенцы с тех пор имеют право на ИИНН, только если они имеют право на определенные вычеты

или подают свою собственную отдельную декларацию.¹ Эта ситуация значительно снижает полезность ИИНН для идентификации мигрантов после 2017 года. Дальнейшее усложнение использования ИИНН заключается в том, что от отдельных лиц периодически требуется повторно подавать заявку на получение ИИНН, и, таким образом, кто-то может иметь несколько ИИНН в течение жизни. Кроме того, держатели ИИНН могут позже подать заявку на получение НСС и, таким образом, для них в файле могут быть как ИИНН, так и НСС. Поскольку номера ИИНН не регистрируются в системе социального обеспечения, их нельзя напрямую сопоставить с данными системы Numident, хотя можно использовать вероятностные методы для сопоставления записей ИИНН с другими наборами данных. Текущая версия ИБДММ не включает держателей ИИНН, но в настоящее время проводятся дополнительные исследования, чтобы понять, как их можно включить в будущие анализы.

С. Создание ИБДММ

8. Как отмечалось ранее, система Numident является основой для выявления иммигрантов, родившихся за границей. Она используется на первом этапе обработки, когда лица, родившиеся за границей, идентифицируются с помощью переменных гражданства из системы Numident (включая иностранцев, имеющих разрешение на работу, и натурализованных граждан). Затем мы используем год создания записи в качестве приблизительного показателя для года въезда в США. Наконец, мы удаляем лиц, умерших в том же году, когда они мигрировали. Этот шаг приводит к оценке численности иммигрантов иностранного происхождения по годам с демографическими характеристиками, хотя и явно завышенной. Мы ожидаем, что здесь оценка будет завышена, так как эта совокупность включает заявителей на получение НСС, которые получили НСС, но либо приехали в США только на короткий период времени, либо никогда фактически не мигрировали в США. Информация о лицах, родившихся в США, сохраняется в файле, чтобы иметь группу сравнения с родившимися за границей. Данные о расе и латиноамериканском происхождении являются неполными или отсутствуют в системе Numident², поэтому необходимо использовать альтернативные методы для определения расы и латиноамериканского происхождения путем моделирования данных десятилетней переписи населения 2010 года и файлов ОАС. Такие методы определения расы в системе Numident не были включены в статью, что ограничивает анализ этих переменных.
9. На втором этапе обработки мы сопоставляем записи Numident с данными налоговых деклараций по форме 1040 НС, чтобы подтвердить въезд в США. Numident содержит все заявки на получение НСС, в том числе от тех, кто получил НСС, но фактически никогда не въезжал в США или проживал там в течение короткого периода времени. Чтобы исключить эту группу из наших оценок, мы сопоставляем данные НС, чтобы ограничить совокупность законными мигрантами, которые работали и платили налоги в США, а также как работающими, так и неработающими натурализованными гражданами. На этом этапе также присваиваются геокоды, что дает нам информацию об иммигрантах иностранного происхождения с демографическими характеристиками в разрезе страны, штата или округа. Здесь мы ожидаем занижение оценки численности населения иностранного происхождения, поскольку мы не учитываем мигрантов, не подающих налоговую декларацию, а также легальных неработающих мигрантов.
10. На данном этапе ИБДММ включает следующие группы иммигрантов: натурализованные граждане, иностранные граждане, имеющие разрешение на работу и подающие налоговые декларации, а также их неработающие иждивенцы и супруги. Не включены следующие группы населения: граждане США, родившиеся за границей от родителей-американцев, нелегальные мигранты, работающие мигранты, не подавшие налоговую декларацию, и неработающие иждивенцы, на которых не распространяется

¹ См. <https://www.irs.gov/individuals/individual-taxpayer-identification-number>.

² Почему исследователи теперь полагаются на опросы для получения данных о расе по программам OASDI и SSI: Сравнение четырех основных обследований (ssa.gov)

освобождение от уплаты налогов. Учитывая, что ОАС призвано быть репрезентативным для всего постоянного населения США, оно должно включать информацию о многих группах иностранного происхождения, которые в настоящее время отсутствуют в ИБДММ.

D. ОАС

11. В настоящем документе микроданные ОАС за 2019 год связываются с ИБДММ, чтобы помочь оценить охват ИБДММ, а также качество данных как ИБДММ, так и ОАС. ОАС представляет собой большое ежегодное непрерывное обследование домохозяйств США, в ходе которого запрашивается подробная информация, ранее собранная в развернутой форме десятилетней переписи. Полностью реализованное в 2005 году, оно в настоящее время охватывает около 3,5 миллионов адресов в год. Включение в выборку основано на проживании или планировании проживания не менее двух месяцев по адресу, включенному в выборку. В ходе ОАС респондентам задают вопросы социально-демографической и экономической направленности, которые включают переменные, относящиеся к иммиграции, такие как страна рождения, статус гражданства, год въезда в США и страна проживания год назад. Хотя информация о НСС не собирается в рамках ОАС, респондентам ОАС могут быть присвоены ПКИ с помощью системы проверки личности (СПЛ), которая определяет вероятность путем сопоставления информации об имени, поле и адресе.
12. Поскольку выборочная совокупность ОАС включает всех жителей США и не различает их по правовому статусу, мы считаем, что ОАС является потенциально хорошим источником информации о мигрантах, не учтенных в ИБДММ. Однако, учитывая сложность подсчета недавно въехавших и нелегальных мигрантов, вполне вероятно, что эти группы населения недостаточно представлены в ОАС (Jensen et al, 2015). Хотя мы не собираемся оценивать охват ОАС в этой статье, эту проблему следует иметь в виду при интерпретации некоторых наших результатов.

E. Система проверки личности (СПЛ)

13. СПЛ применяется Бюро переписи населения для определения и проверки номеров НСС и ПКИ для персональных записей в ходе обследований, переписей и административных записей. Бюро переписи пытается присвоить ПКИ каждой административной записи с помощью вероятностной модели, известной как Модель проверки личности (Wagner and Layne, 2014), и состоящей из четырех модулей. Во-первых, если административные данные содержат НСС, модуль верификации проверяет точное совпадение НСС с файлом Numident и проверяет достаточное соответствие элементов имени и даты рождения. Если они согласуются, НСС считается верифицированным и СПЛ присваивает соответствующий ПКИ записи о человеке. Если НСС отсутствует, как в случае с ОАС, СПЛ переходит к еще трем вероятностным модулям, пытаясь присвоить НСС административной записи, используя географические данные, имя и дату рождения. Приблизительно 94 % всех записей ОАС за 2010 год получили ПКИ, из чего следует, что только 6% всех записей не могут быть связаны с какими-либо административными данными.
14. Лица, которым не присвоены ПКИ, потенциально могут привести к смещению связанных данных. При изучении ОАС 2009 и 2010 годов был сделан вывод о том, что СПЛ с меньшей вероятностью верифицирует маленьких детей, представителей меньшинств, проживающих в помещениях для группового проживания, иммигрантов, недавно переехавших, лиц с низким доходом и безработных (Bond, 2014). В дополнение к неприсвоению ПКИ существует возможность ошибочного присвоения ПКИ отдельным лицам (также называемая ошибкой привязки записей), хотя частота этих неправильно присвоенных ПКИ для иностранцев неизвестна (Abowd et al., 2020). Это одна из возможных причин

несоответствия между переменными в ИБДММ и ОАС, когда сопоставление людей происходит неправильно.

Е. Привязка ОАС к ИБДММ

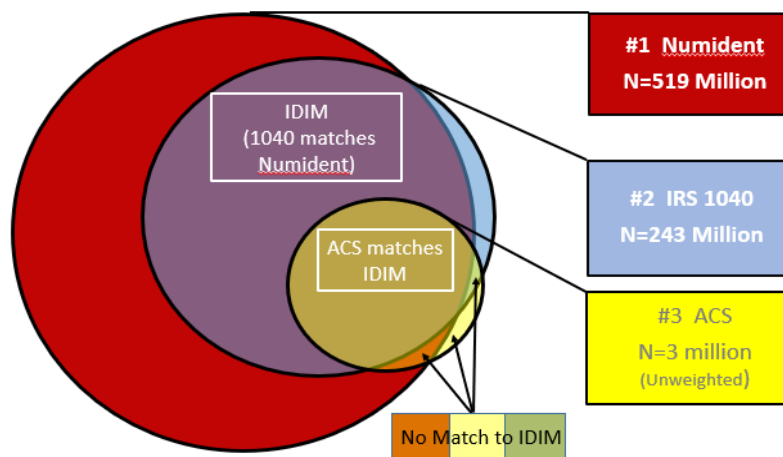
15. В этой статье мы связали данные ОАС за 2019 год с совокупностью ИБДММ за 2019 год. Поскольку совокупность ИБДММ определяется связанными записями из системы Numident и НС, которые характеризуются недостаточным охватом пожилых людей, мы ограничили нашу совокупность лицами моложе 65 лет, чтобы совокупность ОАС была более сопоставимой. Для нашей совокупности ОАС 14% идентифицированы как родившиеся за границей. Чтобы определить иностранца в ОАС, мы используем ответы на вопрос о гражданстве, полученные в рамках ОАС. Те, кто родился в США и за границей от родителей-американцев, определяются как «родившиеся в стране», в то время как натурализованные граждане США и лица, не являющиеся гражданами США, определяются как «родившиеся за границей». Доля родившихся за границей в ИБДММ составляет 12,6%. Родившиеся за границей определяются аналогично ОАС на основании переменной гражданства из-за опасений относительно качества данных о стране рождения и индикаторных переменных иностранного происхождения в системе Numident. В случае ИБДММ мы используем переменную, которая идентифицирует «граждан США» и «легальных иностранцев», в сочетании с переменной, которая указывает, был ли человек когда-либо натурализован. Ни определения ОАС, ни определения в ИБДММ для родившихся за границей не разбивают эту группу по статусу гражданства, что важно отметить из-за проблем с качеством данных для переменной натурализации как в ОАС (Van Hook and Bachmeirb, 2013), так и в Numident.
16. Для всей совокупности ОАС 91% выборки может быть присвоен индивидуальный ПКИ и, таким образом происходит сопоставление с системой Numident или присвоение ИИНН НС. Среди лиц иностранного происхождения, выявленных в рамках ОАС, 79% выборки могут быть присвоены ПКИ. При дальнейшей привязке к данным НС/НСС эти коэффициенты совпадения снижаются: через ПКИ можно привязать 82% всей совокупности ОАС и 71% выборки лиц иностранного происхождения. Такое снижение коэффициентов совпадения ожидаемо, поскольку ИБДММ ограничивается налоговыми декларантами с НСС. На основании предыдущих исследований, рассмотренных выше, ожидался более низкий коэффициент совпадения для совокупности лиц иностранного происхождения, что приводит к недостаточному учету лиц иностранного происхождения в ИБДММ.
17. И наоборот, показатель несовпадения для совокупности родившихся за границей в ОАС составляет 29%, что дает нам ключевую группу сравнения. Эта совокупность состоит из трех отдельных групп: учтенные в ОАС, кому нельзя присвоить ПКИ, те, у кого есть ИИНН и кто подает налоговые декларации, и те, кто может быть связан с системой Numident, но не подавал налоговые декларации или не фигурировал в налоговых декларациях, подаваемых в НС, как освобожденный от уплаты налога. В будущих исследованиях будет предпринята попытка сделать разбивку совокупности ОАС, не сопоставленной с совокупностью ИБДММ, но предполагается, что она включает большую часть нелегальных мигрантов, а также такие группы, как иностранные студенты и иждивенцы, которые не заявлены как освобожденные от налогов в налоговых декларациях, следовательно, те, кто считается отсутствующими в совокупности ИБДММ.

III. Сравнение совокупностей родившихся за границей в ИБДММ и ОАС

18. Существует несколько различных совокупностей, которые можно сравнить для оценки охвата ИБДММ. Рисунок 1 представляет собой концептуальную схему того, как создаются различные совокупности, связывающие между собой Numident, НС и ОАС. ИБДММ состоит из совпадений Numident и НС и в настоящее время исключает тех, у кого есть записи ИИНН (светло-голубого цвета, за пределами Numident и ИБДММ). На рисунке также обозначены важные группы сравнения,

использованные в этом анализе: (1) ИБДММ в целом (фиолетовый), (2) группа совпадения ОАС или ОАС-ИБДММ (желтый) и (3) группы только ОАС или группы несовпадения ОАС-ИБДММ. Группа только ОАС состоит из трех подгрупп: файлы ОАС, не поддерживающие ПКИ (светло-желтый), респонденты ОАС в Numident, но не в ИБДММ (оранжевый, например, не подающие налоговые декларации), и респонденты ОАС в НС, но не в Numident (зеленый, например, держатели ИИНН)

Рисунок 1. Оценочные совокупности Numident, НС, ИБДММ и ОАС



Источники: 2020 Census Numident, IRS 1040 TY19, 2019 American Community Survey

19. В целях сравнения в Таблице 1 приведены подробные демографические и социально-экономические характеристики каждой из этих совокупностей родившихся за границей. Все эти характеристики могут быть получены из ОАС, хотя ограниченное количество переменных также доступно в ИБДММ, включая пол, возраст, статус гражданства, год въезда и государство текущего проживания. Таким образом, для группы совпадения ОАС-ИБДММ некоторые переменные могут быть получены независимо как из ОАС, так и из ИБДММ, что позволит нам оценить качество данных в следующем разделе.

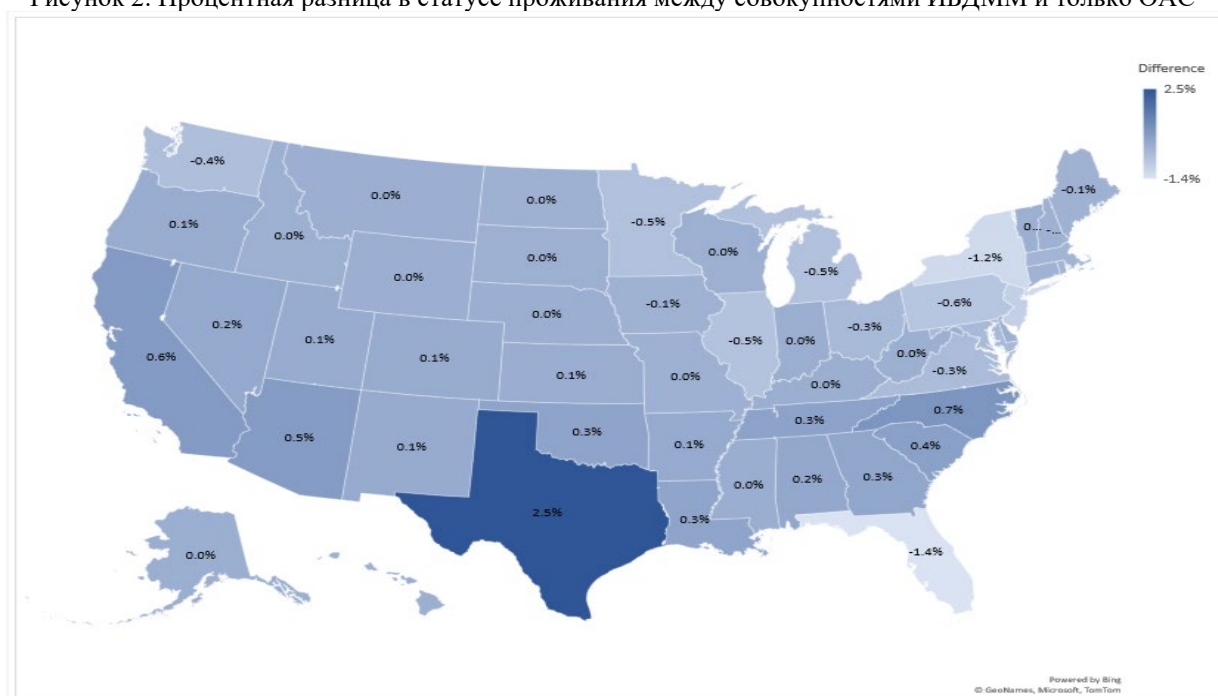
(См. Таблицу 1 в приложении)

20. Как и ожидалось, мы обнаружили четкие различия между сопоставленными совокупностями ИБДММ/ОАС-ИБДММ и совокупностями только ОАС. Совокупность только ОАС содержала больше мужчин, была более молодой (до 25 лет), менее азиатской, более латиноамериканской (мексиканской и центральноамериканской), менее образованной и в большей степени не работающей и бедной. Между группой ИБДММ и группой только ОАС были некоторые расхождения, касающиеся года въезда. Сравнение совокупностей только ОАС и сопоставленных с ОАС показало более резкие различия между группами, чем сравнения с ИБДММ в целом. Особый интерес представляли различия относительно статуса гражданства и года въезда.
21. Распределение по полу и возрасту можно получить как из ИБДММ, так и из ОАС для различных групп. Совокупность только ОАС была более мужской, чем совокупность ИБДММ и сопоставленная совокупность ОАС-ИБДММ. Возраст в ИБДММ был в целом выше (50 лет и старше), в то время как совокупность группы только ОАС была более молодой, включая больше людей студенческого возраста и детей.

22. Как обсуждалось ранее, раса и латиноамериканское происхождение в настоящее время не поддаются измерению в ИБДММ, поэтому это сравнение было ограничено группами только ОАС и группами, сопоставленными с ОАС. Были обнаружены четкие различия: гораздо больше азиатов в совокупности, сопоставленной с ОАС, и гораздо больше латиноамериканцев в совокупности только ОАС. Среди латиноамериканцев в совокупности только ОАС было гораздо больше лиц мексиканского или центральноамериканского происхождения по сравнению с совокупностью, сопоставленной с ОАС.
23. Социально-экономические переменные доступны только в ОАС, и были видны четкие различия между совокупностями, в которых используется только ОАС, и совокупностями, сопоставленными с ОАС. С точки зрения образования, совокупность только ОАС с гораздо большей вероятностью имела ниже среднего образование, в то время как совокупность, сопоставленная с ОАС, с большей вероятностью имела, по крайней мере, диплом колледжа. В связи с этим, группа сопоставления с ОАС с большей вероятностью работала и не жила в бедности, чем группа только ОАС.
24. Некоторые интересные результаты были обнаружены при рассмотрении переменных года въезда и гражданства, представленных как в ОАС, так и в ИБДММ. Год въезда определяется в ИБДММ как год, когда НСС был введен в систему Numident, в то время как ОАС спрашивает респондентов, в каком году они переехали в США, поэтому мы ожидаем увидеть различия между этими наборами данных. Это не относится к совокупностям только ОАС и ИБДММ, поскольку в них распределение по годам въезда очень похоже. Однако при сравнении совокупностей данных только ОАС и сопоставления с ОАС, в которых используется одна и та же переменная измерения для года въезда, различия действительно появляются. Совокупность только ОАС с большей вероятностью включала недавних мигрантов (с 2015 года), чем совокупность сопоставления с ОАС, представители которой с большей вероятностью прибыли до 1999 года. Учитывая аналогичное распределение года въезда между совокупностями ИБДММ и только ОАС, были обнаружены некоторые неожиданные различия между распределениями года въезда на основе ИБДММ при сопоставлении ОАС-ИБДММ, и распределениями года въезда на основе ОАС при сопоставлении ОАС-ИБДММ, которые будут более подробно рассматриваться в следующем разделе.
25. Статус гражданства также дает интересные результаты, которые ставят под сомнение качество этой переменной в ИБДММ. Лица иностранного происхождения в ИБДММ с гораздо большей вероятностью не являются гражданами США, чем в совокупности, сопоставленной с ОАС, что стало неожиданным результатом. Это говорит о том, что записи ИБДММ не обновляются регулярно после того, как человек натурализуется, или дает основания полагать, что данные ОАС для этой переменной низкого качества (или наблюдается сочетание обеих причин). Более показательным является сравнение совокупностей только ОАС и сопоставленных с ОАС, где обнаруживаются серьезные расхождения. Доля иностранцев и лиц без гражданства намного выше в совокупности только ОАС, в то время как, наоборот, доля натурализованных иностранцев, родившихся за границей, намного выше для совокупности сопоставления с ОАС. Это позволяет говорить о том, что совокупность только ОАС более репрезентативна для лиц, не имеющих права на НСС, включая нелегальных мигрантов и иждивенцев легальных мигрантов.
26. Тот или иной статус гражданства не имеет значения с точки зрения того, как мы используем ИБДММ для оценок лиц, родившихся за границей. Неважно, есть ли у кого-то в ИБДММ точная и актуальная информация о статусе натурализации, поскольку важно, родился он за границей или нет. Однако эти результаты могут иметь важное значение для других типов анализа с использованием административных данных и, вероятно, потребуют дополнительных связей с другими источниками данных (например, из Службы гражданства и иммиграции США (USCIS)) для точного определения статуса гражданства.

27. Наконец, на рисунке 2 показана разница в статусе проживания для тех, кто присутствует в файлах только ОАС и ИБДММ. Различия были относительно небольшими: в ИБДММ присутствовало немного больше представителей в таких штатах, как Флорида и Нью-Йорк, а в совокупности только ОАС был немного больше представлен Техас.

Рисунок 2. Процентная разница в статусе проживания между совокупностями ИБДММ и только ОАС



Источники: Бюро переписи населения США, Интегрированная база данных по международной миграции и Обследование американского сообщества 2019 года; Служба социального обеспечения; и Налоговая служба

28. Эти результаты свидетельствуют о том, что в ИБДММ очевидно недостаточно представлены определенные группы лиц иностранного происхождения. Несмотря на собственные потенциальные погрешности в охвате ОАС лиц, родившихся за границей, представляется, что оно более полно учитывает трудные для подсчета группы населения, родившегося за границей, отсутствующие в ИБДММ, и, следовательно, должно быть полезно для корректировки неполного охвата отдельных групп мигрантов в ИБДММ. Частично проблему недостаточной репрезентативности ИБДММ можно было бы решить путем включения ИИНН, но это сопряжено с рядом новых сложностей, которые еще предстоит решить. Вместо привязки дополнительных источников административных данных к

ИБДММ представляется целесообразным интегрировать ОАС в ИБДММ для частичного решения проблемы недостаточного охвата. Более подробно это обсуждается далее в статье.

IV. Анализ качества данных

29. Первоначальный анализ поднял некоторые вопросы о качестве данных как ИБДММ, так и ОАС, особенно в отношении переменных гражданства и года въезда. Чтобы получить представление о качестве данных, мы сравниваем общие переменные, полученные с помощью ОАС и ИБДММ, для людей, которые сопоставляются по обоим наборам данных. Как обсуждалось ранее, эти общие переменные включают пол, возраст, статус гражданства, год въезда и штат текущего проживания.
30. Результаты для переменных, полученных на основе данных ОАС и ИБДММ, как для пола, так и для возраста, являются многообещающими. Существует очень сильное совпадение индивидуальных половых переменных, полученных как из ИБДММ, так и из данных ОАС: более 97% респондентов-мужчин и 98% респондентов-женщин сообщили об одном и том же поле. Что касается возраста, мы ожидаем увидеть различия в зависимости от того, как возраст определяется в каждом источнике данных. В ходе ОАС запрашивается возраст на момент обследования (в дополнение к дате рождения), в то время как возраст в системе Numident привязан к середине года. По-видимому проблема именно в этом, поскольку только 48% респондентов имеют одинаковый возраст в обоих файлах. Однако, когда мы увеличиваем возраст до плюс-минус одного года, уровень совпадения увеличивается до 95%. Хотя мы не можем определить, какой источник данных имеет более точные значения, возрастная аккумуляция является известной проблемой для ответов о возрасте при опросах, и все же в целом соответствие возрастов между источниками данных кажется высоким.

Table 2. Citizenship Status Reported on the IDIM and ACS for Matched Respondents

	Citizenship Status		
	Reported on ACS		
	Native	Naturalized	Non-Citizen
Reported on IDIM			
Foreign-born Non-Citizen	9%	35%	56%
Naturalized	10%	84%	6%

Note: Values are unweighted

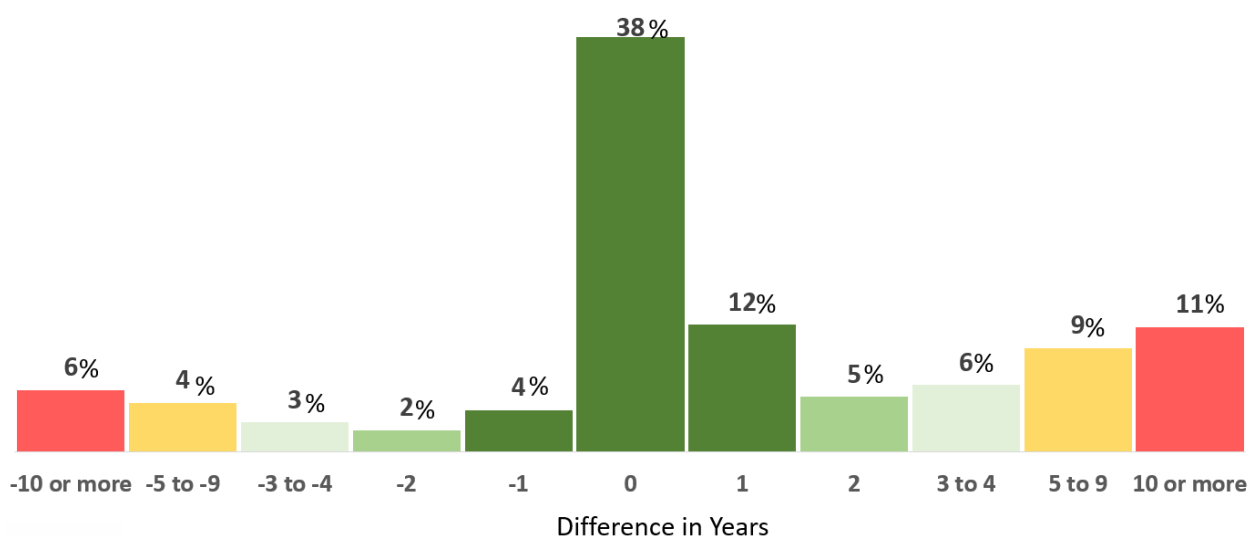
Sources: US Census Bureau, Integrated Database for International Migration and 2019 American Community Survey; Social Security Administration; and Internal Revenue Service

31. Статус гражданства определяется очень похоже в ОАС и в ИБДММ, при этом респонденты, родившиеся за границей, делятся на натурализованных и иностранцев. Более ранние результаты предполагали отсутствие связи между этими переменными, полученными с помощью ИБДММ и ОАС, что подтверждается этим анализом. Что касается родившихся за границей по данным ИБДММ, идентифицированных как иностранцы, только 56% из группы сопоставления с данными ОАС сообщают о том же статусе иностранца в соответствии с переменной гражданства в ОАС, в то время как 35% являются натурализованными гражданами и 9% являются коренными жителями согласно вопросу ОАС о гражданстве. Для натурализованных граждан, родившихся за границей, в данных ИБДММ, сопоставленных с ОАС, наблюдается лучшее совпадение: 84% натурализованных согласно вопросу ОАС и только 6% являются иностранцами, а 10% являются коренными жителями. Большие различия между статусами гражданства, полученными на основе ОАС и ИБДММ, могут свидетельствовать об отсутствии обновлений статуса натурализованного гражданина в записях ССО. Однако это также может свидетельствовать о том, что в рамках ОАС сообщаются неточные данные. Кроме того, вызывает беспокойство тот факт, что около 10% родившихся за границей по данным

системы Numident в ИБДММ классифицируются как коренные жители. Если рассмотреть иностранцев, представленных в данных ОАС, то более 10% также классифицируются как коренные жители в ОАС. Дальнейшее построение комбинационных таблиц по странам рождения может прояснить некоторые из этих выводов. Одним из частичных объяснений этого несоответствия может быть неправильная категоризация коренных жителей при ответе на вопрос ОАС о гражданстве «Родившийся за границей от родителей-американцев». По всей вероятности из-за неоднозначной формулировки вопроса о гражданстве в анкете сюда включены лица, чьи родители не были натурализованными гражданами на момент рождения респондента. В дополнение к ошибочным ответам в данных ОАС и ИБДММ причина кроется еще и в высокой импутации, связанной с переменной гражданства, и неверно подобранных ПКИ. Однако для тщательного изучения этого вопроса потребуются дополнительные исследования.

32. Как обсуждалось ранее, мы ожидаем несоответствия между значениями года въезда в ИБДММ и ОАС, учитывая разные способы измерения этой переменной в двух наборах данных, а также проблемы качества данных с этой переменной ОАС в связи с точностью отклика модели и возрастной аккумуляцией в ответах. Анализ, показанный на рисунке 3, подтверждает несоответствие между переменной года въезда в ИБДММ и ОАС для сопоставленных лиц. Значения года въезда из этих двух наборов данных совпадают только в 38% случаев. Если мы расширим этот диапазон до плюс-минус одного года, наблюдается улучшение только до 54% случаев, а плюс-минус два года улучшит этот результат до 61%. Даже с диапазоном плюс-минус 9 лет значения года въезда совпадают только в 83% случаев между наборами данных. Опять же, неизвестно, в какой степени это преимущественно связано с проблемами предоставления данных (как для ОАС, так и для Numident), высокой импутацией и (или) ошибкой привязки записей ПКИ. Очевидно, что необходимо дополнительное изучение переменных года въезда и статуса гражданства.

Рисунок 3. Распределение различий в ответах на вопрос о годе въезда среди сопоставленных лиц в данных ОАС и ИБДММ



Источник: Бюро переписи населения США, Интегрированная база данных по международной миграции и Обследование американского сообщества 2019 года; Служба социального обеспечения; и Налоговая служба

33. Наконец, для лиц, сопоставленных в ОАС-ИБДММ, мы видим сообщаемый статус проживания в ИБДММ и ОАС. Географическая переменная в ОАС зависит от места, где проживал респондент обследования на момент включения в обследование, а в ИБДММ - от того, где человек подавал свою налоговую декларацию. Вполне возможно, что человек переехал в другой штат в течение периода измерения, поэтому можно было бы ожидать некоторых различий по этой переменной между наборами данных. Оценка этой переменной показывает относительно высокое соответствие с 92% географических совпадений по данным ОАС и ИБДММ на уровне штата для связанных лиц. Различия вполне могут быть вызваны перемещениями между штатами в течение периода, хотя менее вероятно, что это связано с импутацией данных ОАС, поскольку эта переменная берется из выборочного списка адресов.
34. Если подытожить, то различия между некоторыми переменными, полученными из ИБДММ и ОАС, для сопоставленных лиц были неожиданно большими. Как упоминалось при обсуждении переменных статуса гражданства и года въезда, одним из возможных объяснений является то, что данные сообщаются неправильно по каждому набору данных. Существует также вероятность того, что высокие импутации для определенных переменных для родившихся за границей, таких как гражданство, место рождения и год въезда в ОАС, еще больше усиливают эти различия. Кроме того, могут быть ошибки при установлении связей между записями ИБДММ и ОАС, и, следовательно, они будут присвоены разными лицами, что вполне возможно, учитывая вероятностный метод, используемый для присвоения ПКИ при отсутствии информации о НСС. Это еще одно направление, которое требует дальнейшего изучения, чтобы развеять возможные опасения по поводу качества данных ИБДММ и ОАС.

V. Использование ОАС для корректировки ИБДММ

35. Цель этого зондирующего исследования заключалась не только в оценке охвата и качества данных ИБДММ, но и в предоставлении нам информации о том, можно ли использовать ОАС для корректировки ИБДММ с учетом подтвержденных ограничений охвата. Несмотря на потенциальные погрешности в ответах для целых домохозяйств и для отдельных вопросов об иностранном происхождении, по всей видимости ОАС действительно учитывает население иностранного происхождения, отсутствующее в ИБДММ, а именно нелегальных мигрантов и неофициально трудоустроенных мигрантов, которые не подают налоговых деклараций, иностранных студентов и некоторых иждивенцев налоговых декларантов НС.

36. Бюро переписи населения США производит оценки чистых международных миграционных потоков для страны, штата и округа в разбивке по возрасту, полу, расе и латиноамериканскому происхождению, в основном с использованием данных ОАС. Разработка ИБДММ проводилась не с целью заменить ОАС, скорее это была попытка использовать сильные стороны каждого набора данных посредством интеграции данных, что позволило бы улучшить наши оценки. ИБДММ может быть особенно полезна для улучшения оценок на уровне округов, где наши оценки основываются на сводных данных ОАС о численности иностранцев за 5 лет и по-прежнему имеют высокий уровень изменчивости выборки, особенно для небольших округов. В то же время все еще существует возможность использования ОАС для корректировки недостаточного охвата ИБДММ для оценок на уровне страны и округа, а также для национальных и субнациональных характеристик.
37. Бюро переписи населения ранее интегрировало административные данные на макроуровне, чтобы улучшить оценки миграции в Пуэрто-Рико и из Пуэрто-Рико после урагана «Мария» на основе национальных обследований, а также для учета воздействия пандемии COVID-19 на международные миграционные потоки в и из США. Эти методики использовали исторические тенденции между данными ОАС и административными данными для корректировки оценок ОАС на основе уровней, наблюдаемых на административных данных. Использование данных ОАС для корректировки ИБДММ будет примером использования данных обследования для корректировки административных записей и потенциально может происходить как на макро-, так и на микроуровне, учитывая характер процедур привязки данных.
38. Например, с точки зрения макроинтеграции, ОАС можно использовать для корректировки не представленных групп иностранных студентов, а также распределения по возрасту на субнациональном уровне путем добавления доли учащихся к общей численности по стране или путем применения данных ОАС о возрастном распределении на уровне округа для округов с большим количеством студентов. Также можно было бы использовать уровни и характеристики групп населения только по данным ОАС для учета отсутствующих в ИБДММ групп населения посредством либо пропорциональных, либо смоделированных поправок оценок. Эти методы могут помочь учесть часть отсутствующих в ИБДММ нелегальных мигрантов, а также другие группы населения, родившиеся за границей. Дальнейшая работа по совершенствованию разбивки населения по данным только ОАС на различные категории также улучшит нюансы любых корректировок, сделанных для этого населения.
39. С точки зрения микроинтеграции можно было бы использовать связанных в ИБДММ и ОАС домохозяйств, информация о членах семьи которых присутствует в ОАС, но не в ИБДММ, для корректировки данных о неучтенных иждивенцах, не включенных в налоговые декларации. Информация о размере семьи в данных ОАС и о том, в какой степени это население отсутствует в ИБДММ, может дать информацию для некоторых вероятностных методов оценки. Все эти методы интеграции макро- и микроданных еще предстоит разработать, но первоначальные результаты показывают, что ОАС может быть полезным инструментом для улучшения оценок миграции, полученных с помощью ИБДММ.

VI. Обсуждение

40. Как показано в настоящем документе, предстоит еще многое сделать для улучшения охвата ИБДММ и соответствующих оценок. На следующих этапах необходимо добавить в ИБДММ данные о расе и этнической принадлежности с помощью хорошо зарекомендовавшего себя метода, используемого другими подразделениями Бюро переписи населения, а именно: использование сопоставленной информации из переписей 2010 и 2020 годов о расе/этнической принадлежности для присвоения значений, а также моделирование отсутствующей информации о новых мигрантах на основе расового распределения по стране происхождения согласно ОАС. Это дало бы нам возможность получить все характеристики, необходимые для получения оценок миграции, исходя из ИБДММ. Применение ИБДММ для получения субнациональных оценок на уровне округов нуждается в дальнейшей оценке, даже если сохраняются проблемы охвата.

41. Может быть полезной и значительно поможет улучшить охват нелегальных мигрантов дальнейшая работа над возможностью добавления ИИНН в ИБДММ без дублирования. Кроме того, ИБДММ недостаточно учитывает маленьких детей. Это может быть частично решено за счет доработки процессов импутации для несоответствующих иждивенцев, но это потребует дальнейшего изучения. Настоящая статья предоставила дополнительную информацию об этой недостаточно учтенной группе населения, и, как обсуждалось ранее, дальнейшая разбивка населения на группы по данным только ОАС улучшило бы наше понимание ИБДММ. Использование дополнительных источников данных также может помочь в этом.
42. Было бы очень полезно связать ИБДММ с другими источниками данных, например файлами Службы гражданства и иммиграции США или Министерства здравоохранения и социальных служб. С указанными агентствами разрабатываются соглашения об обмене данными. В результате можно получить бесценную информацию не только о неучтенных группах населения, но и для проверки и улучшения качества данных в ИБДММ.
43. Аналогично, хотя в настоящее время у нас нет доступа к данным Иммиграционной и таможенной полиции США, данные Программы студентов и посетителей по обмену были бы идеальными для оценки потоков студентов и посетителей по обмену. Данные системы информации о въезде и выезде от Бюро таможенного и пограничного контроля могут помочь нам измерить потоки нелегальных мигрантов. Это примеры потенциальных источников данных, которые могут быть включены в ИБДММ в будущем.
44. Помимо улучшения охвата ИБДММ, необходимо провести дальнейшую работу, чтобы лучше понять качество данных, например относительно гражданства и года въезда. В любом случае Бюро переписи населения США будет продолжать попытки развивать и интегрировать административные источники с данными обследований, чтобы улучшить оценки чистой международной миграции.

Ссылки

Abowd, J., William R. Bell, J. David Brown, et al. (2020). Determination of the 2020 US Citizen Voting Age Population (CVAP) Using Administrative Records and Statistical Methodology. Center for Economic Studies Working Paper Series No. 20-23. Washington, DC: US Census Bureau.

Bond, B., J.D, Brown, A. Luque, and A. O'Hara. (2014). The Nature of the Bias when Studying Only Linkable Person Records: Evidence from the American Community Survey. Center for Administrative Records Research and Applications Working Paper Series No. 2014-08. Washington, DC: US Census Bureau.

Brown, J. David, Misty L. Heggeness, Suzanne M. Dorinski, Lawrence Warren and Moises Yi. (2019). Predicting the Effect of Adding a Citizenship Question to the 2020 Census. *Demography* 56:1173–1194.

Jensen, Eric b., Renuka Bhaskar, and Melissa Scopilliti. (2015). Demographic Analysis 2010: Estimates of Coverage of the Foreign-Born Population in the American Community Survey. US Census Bureau Working Paper No. 103. Washington, DC: US Census Bureau.

Luque, A., and R. Bhaskar. (2014). 2010 American Community Survey Match Study. Center for Administrative Records Research and Applications Series Working Paper No. 2014-03. Washington, DC: US Census Bureau.

Rastogi, S., and A. O'Hara. (2012). 2010 Census Match Study Report. 2010 Census Planning Memoranda Series No. 247. Washington, DC: US Census Bureau.

Van Hook, Jennifer and James D. Bachmeier. (2013). How Well Does the American Community Survey Count Naturalized Citizens? *Demographic Research* 29(1): 1–32.

Wagner, D., and M. Layne. (2014). The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software. Center for Administrative Records Research and Applications Working Paper Series No. 2014-01. Washington, DC: US Census Bureau.

Приложение

Демографические характеристики	Родившиеся за границей			
	ОАС в качестве основы		ИБДММ в качестве основы	
	Записи ОАС, сопоставленные с ИБДММ	Записи ОАС, не сопоставленные с ИБДММ	Административные записи ИБДММ	Подмножество ИБДММ для сопоставленных записей ОАС
Пол				
Мужской	48%	52%	48%	47%
Женский	52%	48%	52%	53%
Возраст				
0 -17	6%	10%	5%	5%
18 -24	7%	10%	8%	7%
25 -34	19%	21%	20%	18%
35 -44	25%	25%	24%	24%
45 -54	25%	20%	24%	26%
55 -64	20%	15%	18%	20%
Раса				
Только белые	54%	67%	X	X
Только черные	12%	10%	X	X
Только азиаты	31%	19%	X	X
Прочие	3%	4%	X	X
Нелатиноамериканцы	60%	39%	X	X
Латиноамериканцы				
Мексиканцы	21%	37%	X	X
Центральная Америка/Доминиканская республика	9%	16%	X	X

Прочие	10%	8%	X	X
Статус бедности				
Отсутствие бедности	91%	78%	X	X
Бедность	9%	22%	X	X
Статус занятости				
Занятость	78%	64%	X	X
Незанятость	3%	3%	X	X
Не входит в число работающих	19%	33%	X	X
Образование				
Ниже средней школы	20%	37%	X	X
Средняя школа	21%	26%	X	X
Высшее неоконченное/высшее образование	42%	29%	X	X
Последипломное образование	17%	8%	X	X
Статус гражданства				
Иностранцы	46%	71%	72%	67%
Натурализованные	54%	29%	28%	33%
Год въезда				
До 1990	23%	17%	17%	19%
с 1990 до 1999	24%	20%	22%	25%
2000 -2009	27%	29%	25%	26%
с 2010 до 2014	12%	12%	16%	15%
2015 и позже	13%	23%	20%	15%
N (тыс.)	26 450	11 040	29 690	283

Примечание: Значения ОАС являются взвешенными.

Источники: Бюро переписи населения США, Интегрированная база данных по международной миграции и Обследование американского сообщества 2019 года; Служба социального обеспечения; и Налоговая служба