

Национальная экспериментальная статистика благосостояния (NEWS) Объединение данных обследований и административных данных для улучшения статистики доходов и бедности

Лиана Э. Фокс
Бюро переписи населения США

Группа экспертов ЕЭК ООН по измерению бедности и неравенства
28-29 ноября 2023 г.

*Любые мнения и выводы, выраженные в данном документе, принадлежат авторам и не отражают точку зрения Бюро переписи населения США. Бюро переписи населения проверило данный информационный продукт с целью обеспечения надлежащего доступа, использования и защиты от разглашения конфиденциальных исходных данных, использованных для создания данного продукта (номер системы управления данными (DMS): P-7524052, номер одобрения **Disclosure Review Board (DRB): CDRB-FY23-SEHSD003-025**).*

Атрибуция

- Адам Би, Джошуа Митчелл, Николас Миттаг, Джонатан Ротбаум, Карл Сандерс, Лоренс Шмидт и Мэтью Унрат

Оценки доходов и уровня бедности

- Неотвечаемость при обследовании домохозяйств растёт от 11% в 2013 году до 31% в 2023 году (мартовское текущее обследование населения)
- Среди тех, кто ответил на вопросы обследования, многие не отвечают на вопросы о доходах
 - ~ 45% дохода в официальной оценке бедности, вмененного за отсутствие ответов
- Среди тех, кто отвечает на вопросы о доходах, многие занижают данные
 - По нашим оценкам, число людей, живущих в бедности, на 1,1 процентного пункта (~3,5 млн. человек) меньше, чем по официальным оценкам

Что такое NEWS?

- Переосмысление методов подготовки статистики доходов и ресурсов
 - Какова наилучшая возможная оценка с учетом всех данных, имеющихся в настоящее время в распоряжении Census, для данной статистики доходов/ресурсов?
- Одновременное устранение нескольких источников предубеждений
 - Применять исследования для решения каждой

Как это делает NEWS?

- Собирает воедино все имеющиеся данные: опросы, переписи, административные записи, коммерческие (сторонние) данные
 - Для корректного решения проблемы смещения часто требуются **связанные** данные
- Делает все прозрачно, воспроизводимо, на основе фактических данных
- Привлекает исследовательское сообщество
 - Будет создана связанная база микроданных и кодов для доступа к ней в федеральных центрах данных статистических исследований
 - Код будет размещен в открытом доступе (с учетом ограничений на раскрытие информации)

Что мы сделали?

- Выпуск версии 1 - 14 февраля 14
 - Proof of concept
 - 1 год
 - Зеркало доходов и релизы бедности - денежные доходы (без налогов, кредитов, неденежных льгот)
 - Представить методы и подход к обратной связи
 - Документ и расчеты доступны по адресу
<https://www.census.gov/data/experimental-data-products/national-experimental-wellbeing-statistics.html>

Проблемы, связанные с измерениями Данные опроса

1. Непредвзятость ответов подразделений
 - Отказ от ответа на опрос
 - **Снижение** уровня бедности на 0,3-0,5 процентных пункта во время пандемии (Bee and Rothbaum, 2022)
2. Непредвзятость ответов на вопросы
 - Отказ от ответов на вопросы о доходах (~45% доходов в CPS ASEC являются вымышленными!)
 - **Снижение** уровня бедности на 0,5-1 процентный пункт (Bollinger et al., 2019; Hokayem et al., 2022)
3. Искажение и занижение данных
 - Неточные ответы
 - Бедность смещена в сторону **увеличения** на 2,5 процентных пункта для лиц 65+ (Bee and Mitchell, 2017)

Предвзятость может иметь различные знаки и величины, которые могут варьироваться в зависимости от группы

Проблемы измерения Административные данные

1. Выборка по административным данным
 - Не все должны подавать налоги или получать W-2 или другую информационную декларацию
 - Ларримор, Мортенсон и Сплинтер (2020) оценивают уровень бедности на основе административных данных, но при этом им приходится вменять существование и статус бедности 4-6 млн. человек
2. Административные данные "неответы"
 - Не представлена некоторая информация, которая должна была бы быть представлена
 - Работа "под столом" без W-2, например - 5% взрослых в CPS ASEC сообщают о заработках в рамках опроса без W-2
3. Административные искажения и занижения
 - Не всегда 100% точность
 - Незарегистрированные чаевые, заниженные доходы от самозанятости (см. анализ налоговых пробелов IRS)

Проблемы измерения Административные данные

4. Концептуальное рассогласование

- Администрация не всегда измеряет то, что мы хотим
- В W-2 исторически не указываются доходы, использованные для уплаты взносов на медицинское страхование, что занижает истинные доходы (перепись населения также не получает эту информацию, когда она доступна)

5. Неполный охват данных

- Данные по отдельным лицам или местам отсутствуют

6. Отбор в систему связи

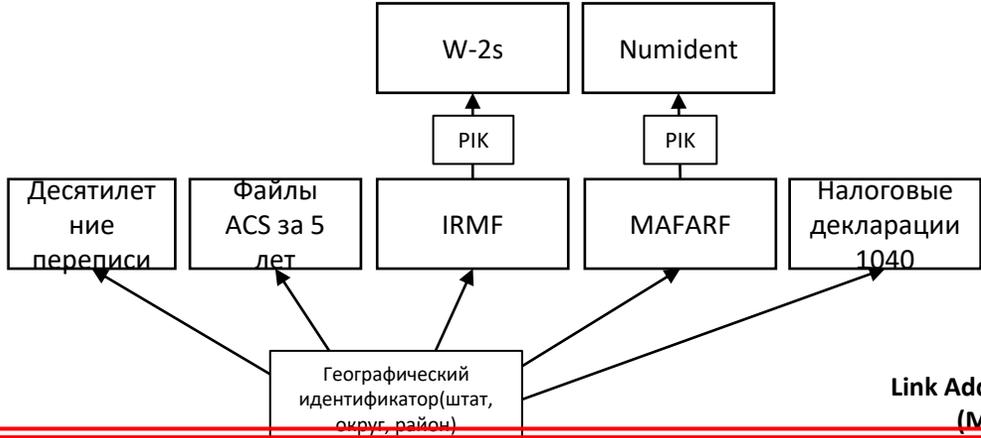
- Не все лица могут быть связаны между собой в разных источниках данных (см. Bond et al., 2014)

Решение проблем, связанных с измерениями

Шаг	Описание	Проблема измерения	Похожие работы
Взвешивание	Использование данных об адресах всех занятых единиц жилья для взвешивания респондентов, связанной выборки для обеспечения репрезентативности целевой совокупности домохозяйств	Неотвечающие единицы обследования Выбор в административные данные "Неотвечающие" административные данные Выборка в систему связей	Ротбаум и др. (2021) Ротбаум и Би (2022)
Импутация			
Доходы от проведения опроса	Вмененный заработок по результатам опроса, обусловленный данными опроса и административной информацией	Неотвечающие элементы обследования	Хокайем и др. (2022)
Валовой заработок администратора	Исчисление валового заработка в случае его отсутствия в административных данных	"Неответы" административных данных Концептуальное несоответствие Неполный охват данных	
Данные по программам с выплатой пособий	Вмененные данные по программам с оплатой по средствам для штатов, по которым отсутствуют административные данные	Неполный охват данных	Фокс и др. (2022)
Доходы неплательщиков	Вычет компенсации по страхованию от безработицы, процентов и дивидендов для неплательщиков	Выборка в административных данных Неполный охват данных	Ротбаум и др. (2023)
Оценка			
Совмещайте заработок на опросах и администрировании	Объединение данных обследования и административных доходов от заработной платы в соответствии с моделью ошибки измерения доходов NEWS	Искажение и занижение данных в обследованиях Административные искажения и занижения данных	Би и др. (2023)

Адресно-связанные данные (взвешивание)

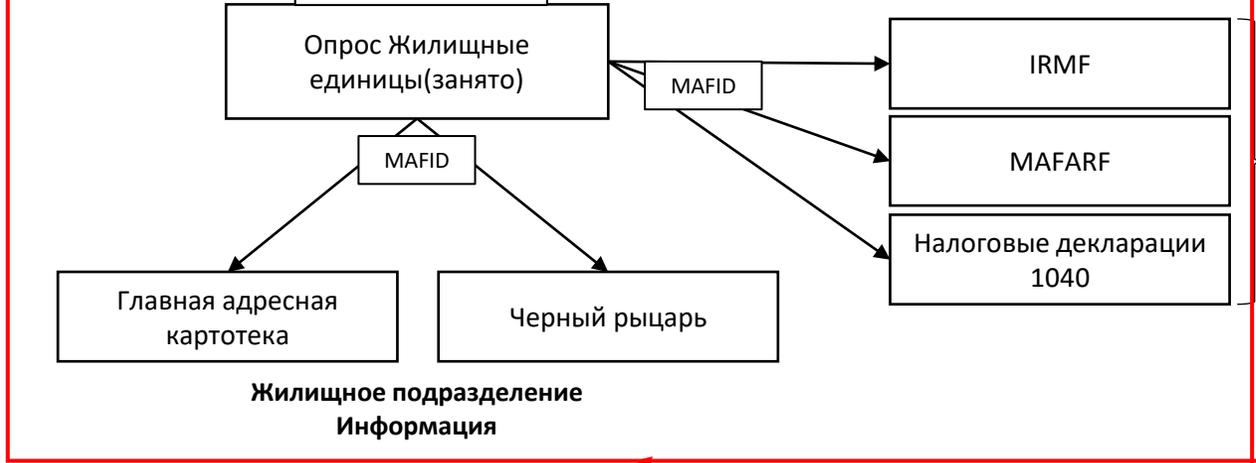
Краткие географические данные характеристик



Ссылки по географическому принципу

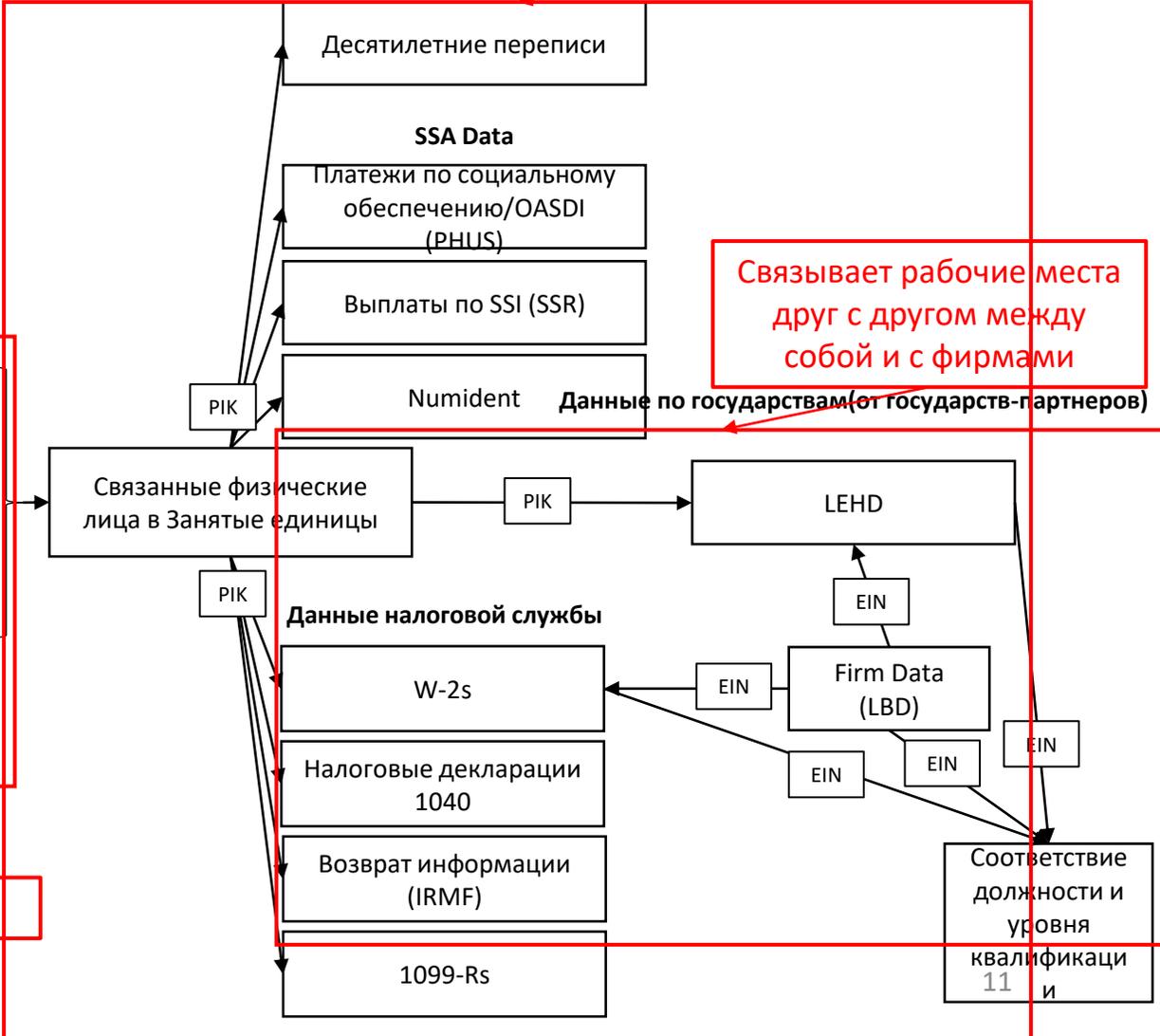
Ссылки на людей в Adresc по адресам

Link Addresses to People (MAFID → PIK)



Жилищное подразделение Информация

Ссылки по адресу



Связывает рабочие места друг с другом между собой и с фирмами

Данные по государствам (от государств-партнеров)



Оценка

Объединение результатов опроса и административного заработка

- Пять источников информации о заработной плате
 1. Обследование
 2. W-2s
 3. Подробный учет доходов
 4. LEHD - Продольная динамика "работодатель – домохозяйство"
 5. 1040 заработная плата и оклад

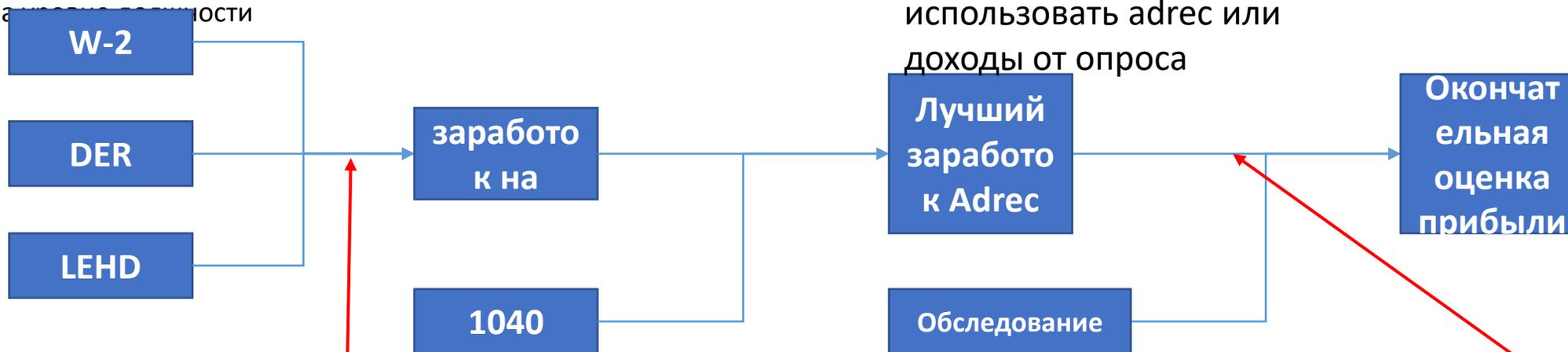
Полная картина - заработная плата и оклад

1. Использование информации об уровне должности для получения "максимально возможного" административного заработка на должности

2. Сравнение с 1040 для проверки отсутствующих доходов (на уровне налоговых единиц)

3. Сравните с результатами опроса и решите, для каких индивидуумов использовать адрес или доходы от опроса

4. Окончательная "наилучшая" оценка доходов для каждого человека/домохозяйства

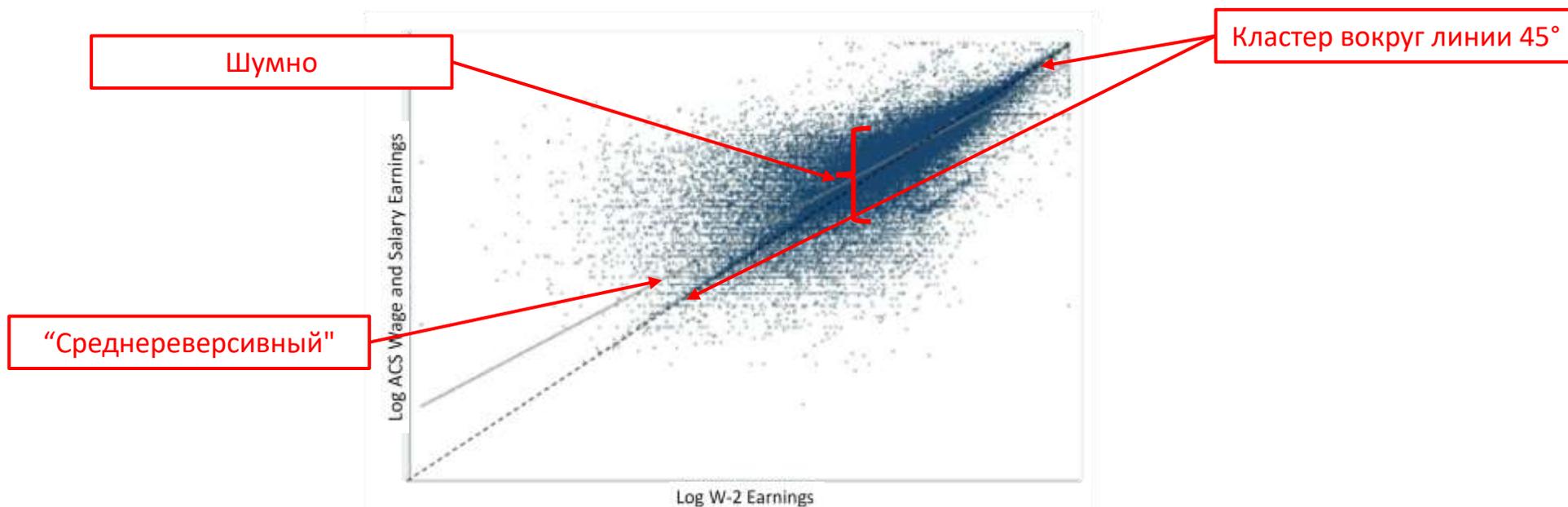


Если LEHD отсутствует (или имеются явные проблемы с качеством данных), то для каждого рабочего места (до 2) произведите интерполяцию валового заработка на основе административной и опросной информации

Улучшение вменений при обследовании

Как совместить опрос и административный заработок?

Различные источники доходов W-2 в сравнении с ответами на вопросы анкеты

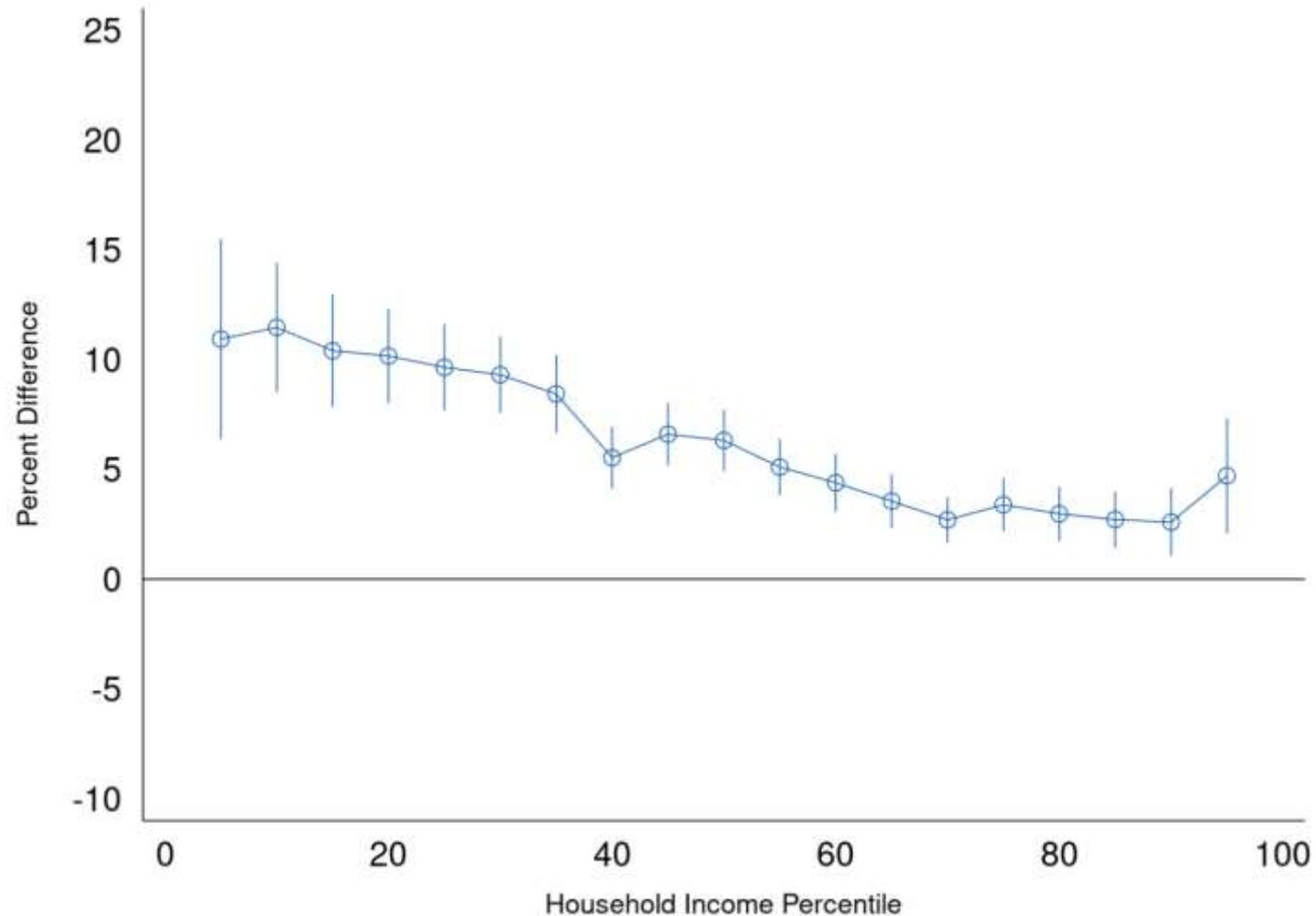


Источник: O'Hara et al. (2017) с использованием данных ACS 2011 года, связанных с записями W-2 2010 года.

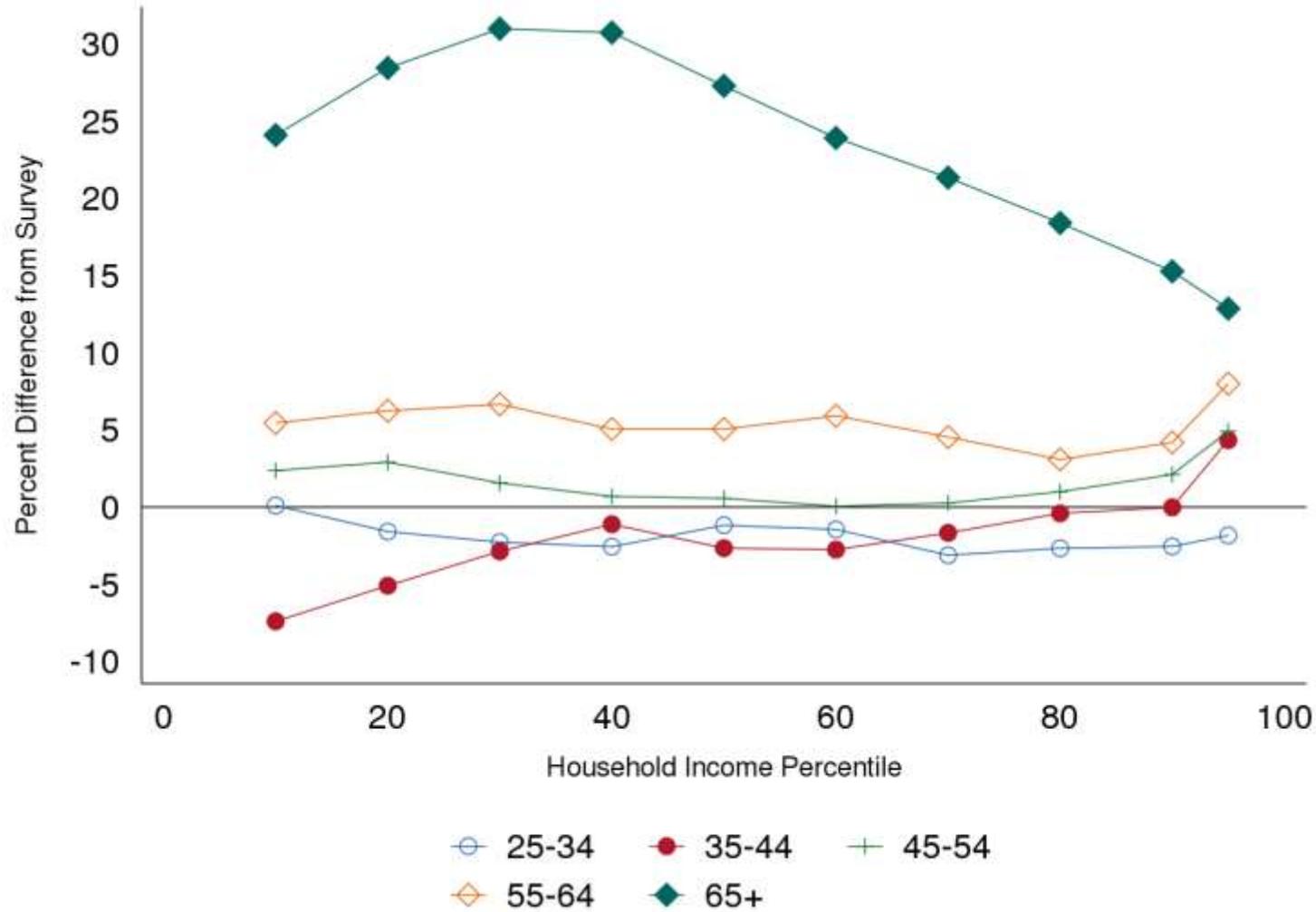
Использование заработков по результатам опроса

- 21% лиц
- Чаще всего для:
 - Работников, занятых в сфере недвижимости и строительства
 - Более молодых работников (25-44 года)
- Реже для:
 - Работников розничной торговли, образования, управления и здравоохранения
 - Пожилых работников (65+)
 - Темнокожих работников

Доходы населения в 2018 году: Оценка NEWS относительно опроса



Доходы населения в 2018 г: НОВИНКИ относительно опроса по возраст



Источник: Ежегодное социально-экономическое приложение к обследованию населения (2019 Current Population Survey Annual Social and Economic Supplement), связанное с административными данными, данными десятилетней переписи населения и данными сторонних организаций.

Результаты

- В целом медианный доход домохозяйств был на 6,3% выше, чем в оценочном исследовании, а уровень бедности - на 1,1 процентного пункта ниже.
- Результаты обусловлены лицами в возрасте 65 лет и старше:
 - Медианный доход домохозяйства был на 27,3% выше, чем в оценочном исследовании
 - Уровень бедности на 3,3 процентных пункта ниже, чем в оценке исследования.
- Существенного влияния на медианный доход домохозяйств для лиц моложе 65 лет и на уровень детской бедности не было.

Планы на будущее

- Несколько лет
 - Не все адресаты доступны во все годы
 - Не все переменные исследования доступны во все годы
- Больше географий
 - Использование ACS - менее подробная информация затрудняет объединение опросов и адресов
- Больше концепций доходов/ресурсов
 - Включает налоги, кредиты и трансферты в натуральной форме
 - Дополнительный показатель бедности
- Устранение дополнительных источников ошибок измерения
 - Доходы от самозанятости
 - Доходы на самом верху распределения (верхние 0,1%, 0,01%,...)
- Дальнейшее исследование допущений и проблем для других интересующих подгрупп
 - Неграждане, бездомные/не имеющие жилья (или лица с нестабильными жилищными условиями), групповые квартиры
- Обратная связь с опросами для улучшения вопросов и обработки

Обратная связь

Документ и расчеты доступны по адресу :

<https://www.census.gov/data/experimental-data-products/national-experimental-wellbeing-statistics.html>

Все комментарии, замечания, предложения и отзывы направляйте по адресу :

census.newsproject@census.gov