

Towards a Vision for a New Data Infrastructure for Federal Statistics and  
Social and Economic Research in the 21st Century

---

**Workshop 3: Approaches to Sharing Blended Data  
in a 21st Century Data Infrastructure**

May 22, 2023

Sharing refugee survey data and privacy concerns with  
providing state level resettlement information or other  
information collected during enrollment

**Presenter:** Tim Triplett, Urban Institute

# Overview of Presentation

- Survey data: ASR (Annual Survey of Refugees)
- Administrative Data: RADS (Refugee Administrative Data)
- Disclosure Risks: Detailed demographic data
- Data requests: Mostly geography
- Other Challenges: RADS data becomes available after five years
- Recommendations for: Improving data quality

# What is the Annual Survey of Refugees

- The ASR is the only national survey on refugee self-sufficiency and integration
- A telephone survey conducted annually since the early 1980s
- Collects information on refugees arriving in the U.S. in the previous five years
- Refugee Act annual reporting requirement to Congress

# Annual Survey of Refugees: The Basics

- Sample of resettled refugees who came within last five years - does not include other humanitarian entrants
- Snapshot of each arriving year, grouped into 3 cohorts
- 1,500 Households → a little over three thousand eligible refugees (16+ years old)
- Telephone interviews Jan-April; CATI
- Linguistic & cultural competence, e.g., gender matching, offering choice of language
- 20 languages (78% coverage)

# RADS – Administrative Data

- Arrival intake form filled out by the Principal Applicant (often head of household)
- Last five years there were 203,041 new arrivals and 78,883 Principal Applicants
- RADS data is used to generate the sample of potential ASR respondents
- Not updated after 90 days
- Sent to the National Archives every after five years

# RADS Data

age	Age at time of arrival of individual
arrival_date	The Date this individual is estimated to arrive in the US or is confirmed to have arrived in the US.
birth_date	The Date of Birth of this individual.
case_guide	Identifies the Case this Individual belongs to.
case_no	Case Number (Case Prefix + Case Number + Case Suffix)
city_name	The City where this Case will be Resettled.
country_of_origin_name	ORR assigned Country of Origin Name (Aggregated for former USSR and Yugoslavia)
ethnicity_desc	Specifies the Ethnicity Description of the individual.
family_relationship_desc	Specifies the relationship description of this individual (case member), to the individual who is the Principal Applicant for this Case.
family_size_no	Specifies how many Members are currently included in this Case (including the Principal Applicant).
fiscal_year	Calculated Fiscal Year based on Arrival Date
gender_code	Gender of individual
ind_guid	Uniquely identifies this Individual.
language	Language
nationality_desc	Specifies the Nationality Desc of this Individual.
native_reading_code	Reading ability of Native language (WRAPS only and only from 2008 forward)

# RADS Data

education_level	The highest level of education desc. attained by this member
state_code	The State code where this individual will be resettled.
voluntary_agency_desc	Voluntary Agency Desc
zip_code	The zip code where individual is resettled
city_90day1	The 90-Day city where the member is resettled.
county_name_90day1	The 90-Day county name where the member is resettled.
addr_address_90day1	The 90-Day street address where the member is resettled.
addr_phone1_90day1	The R&P Primary Phone (if different than PA, Indicate if (H)ome or (W)ork or (C)ell)
addr_phone2_90day1	The R&P Second Phone (if different from PA, Indicate if (H)ome or (W)ork or (C)ell)
state_code_90day1	The R&P State Code where resettled (if different than PA)
zip_code_90day1	The R&P zip code where resettled (if different than PA)
addr_postalcode_90day1	The R&P zip code where resettled (if different than PA)
family_size_no_90day1	Specifies how many Members are currently included in this Case. Information is only available for PA.
addr_phone1_90day2	The 90-Day primary phone of PA (Indicates if (H)ome or (W)ork or (C)ell)
addr_phone2_90day2	The 90-Day second phone of PA (Indicate if (H)ome or (W)ork or (C)ell)
full_name	Full name of individual
maiden_name	Specifies the Maiden Name of this individual.

# Disclosure Risks

- Refugees are a vulnerable population which increases the risk beyond simple math formulas
- Fleeing a country for religious or ethnic persecution, war, threats, or working for the U.S.
- Reason to protect identity include:
  - Want to avoid being located by those threatening them
  - Protection for those left behind
- Sent to the National Archives every after five years



# Disclosure Risks

## Survey Data Masking

- Top code age
- Mask origination countries with less than 100 respondents
- Mask languages with less than 100 respondents
- Mask ethnicities with less than 100 respondents
- Collapse resettlement geography to the four major census regions
- Top code year of arrival
- Remove open ended responses and unusual family arrangements (i.e., foster kids)

# Disclosure Risks

- We do not include RADS data with the public use data file.
- We do use RADS data to impute for missing data.
- Geography (particularly State) is by far the biggest request that we get from data users.
- About a quarter of refugees have moved from the original resettlement location by the time of the survey.
- There is a fair amount discrepancy between the survey data and RADS on variables that should be the same
- RADS data is used for creating survey weights

# Data Requests

- State of resettlement – could we release larger States (current or original)
- Pool data across years to allow for analyses of more countries.
- Longitudinal weights
- Occupations

# Other Challenges

- An academic group has already used the archived RADS data and geo-coded city of relocation to report aggregated numbers
- Aggregated by city and year of arrival, plus country of origin from as recent as 5 years ago.
- 140,503 refugees successfully geo-coded with 26,182 cities aggregated as a single refugee.
- 102,500 of the geocoded refugees are aggregated in a group of 10 or less.
- This makes it harder to reconsider blending RADS and ASR data

# Recommendations

- Include on the public-use file more geographic data (larger states or smaller census regions)
- Help the RADS team better mask their data
- Creating multi-year public-use data files that included blended data
- Create restricted-use files that includes blended data
- Releasing the public-use data sooner

# Thank You

Timothy Triplett  
Senior Survey Methodologist  
Email: [ttriplett@urban.org](mailto:ttriplett@urban.org)

**U R B A N   I N S T I T U T E**  
Statistical Methods Group  
[www.urban.org](http://www.urban.org)