

Office of the Secretary Patient-Centered Outcome Research Trust Fund Project HP-23-040:  
*Data Infrastructure Supporting Research on Refugee Medicaid Service Utilization and Outcomes*

ASPE Office of Human Services Policy and  
Administration for Children and Families Office of Refugee Resettlement

Final Report

August 6, 2025

*Section 1. Overview*

The following report is submitted as the final requirement associated with IDDA OS-ASPE-PCORTF-ACFORR-2023-012 between the Office of the Assistant Secretary for Planning and Evaluation (ASPE) Office of Health Policy (HP) and, jointly, the Office of Human Services Policy (HSP) and the Administration for Children and Families (ACF) Office of Refugee Resettlement (ORR).

The United States refugee resettlement program is the largest in the world, welcoming over three million people since the 1970s. Refugees may arrive with health issues and health care needs and can receive various benefits and services through the ORR refugee resettlement program. Practitioners, program managers, and policymakers need access to strong, high-quality evidence to understand and improve the efficiency of services provided to refugees. However, limitations in the availability of relevant data on refugee populations, including data on health outcomes, access to and utilization of health services, and social determinants of health, make it difficult for researchers to conduct the patient-centered outcomes research studies needed to generate this evidence to inform decisions and improve health outcomes.

The overarching goal of this project was to advance data capacity for patient-centered outcomes research by developing, evaluating, and supporting researchers' use of linked datasets on refugees served by the Medicaid program. The objectives of the HSP and ORR agreement under the IDDA included:

- Develop and evaluate linked Transformed Medicaid Statistical Information System (T-MSIS) and Refugee Arrivals Data System (RADS) dataset
- Produce research dataset and documentation
- Create data products for public use on refugee health

These objectives were not met, as the project was terminated by ORR. This report serves as the final summary of the aforementioned project, describing reasons for termination, tasks and deliverables completed, and lessons learned on building data infrastructure to link administrative datasets held by separate entities.

## Section 2. Reasons for Project Termination

This project relied upon, at a minimum, gaining access to two datasets and putting them in the same computing environment, in order to attempt linking: the Transformed Medicaid Statistical Information System (T-MSIS) from the Centers for Medicare & Medicaid Services (CMS) and the Refugee Arrivals Data System (RADS) from ORR. While the project team successfully obtained access to T-MSIS in the CMS Virtual Research Data Center (VRDC), ORR ultimately did not sign a Memorandum of Understanding with HSP to allow RADS to be transferred and uploaded into the CMS VRDC.

## Section 3. Tasks and Deliverables Completed

In December 2023, the project team executed a contract with the Urban Institute to conduct tasks required to achieve the project’s objectives. The General Services Administration (GSA) served as the contracting office and issued two task orders on the contract, for which work was completed by the Urban Institute: 47QFEA24F0004 (linking federal refugee and Medicaid administrative data) and 47QFEA25F0003 (state-based modeling).

As shown in Table 1, the Urban Institute, in collaboration with the project team, completed several Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) deliverables prior to the termination of the project, specifically: a signed and executed contract, a dataset review plan and a subsequent report, a data use agreement (DUA) between CMS and HSP to access T-MSIS data in the CMS VRDC, an initial data linking plan, an initial data archiving plan memo, and a stakeholder engagement plan. Table 1 also details the deliverables that were not completed due to early project termination.

**Table 1.** OS-PCORTF Deliverables and completion status

Objective	Deliverable(s)	Product Type	Product Description	Completed?
Objective 1: Project Management Support	Contractor support	Contract/grant	Signed & executed contract	Y
Objective 2: Assess feasibility of linking data and obtain agreements	Report on dataset review	Report	Report describing linking feasibility, strengths and weaknesses of data to be linked, and any analysis results	Y

	Data agreements	MOUs, DUAs	Agreements needed to secure access, linking, analysis, and redisclosure of datasets	Partial: DUA between CMS and HSP completed to access data, and project team planned to amend the DUA post-linking to cover redisclosure
Objective 3: Linkage and Evaluation for T-MSIS and RADS data	Report on linking methodology	Report	Report describing performance of data linking methods, and recommended method	N
	Internal-use linked dataset	Dataset	Dataset containing linked RADS and T-MSIS data	N
Objective 4: PCORTF Data Products	Research dataset and documentation	Datasets, summary tables; Codebooks, other documentation	Products based on deliverable in Objective 3, for public-use	N
	Data products for general public use	Tables of dataset aggregation, visualization, or other summaries	Products based on research-use dataset	N
	Data redisclosure plan	Project workplan	Plan describing approach to redisclosing data products and making available to public	Partial: initial data archiving platform memo produced but other parts not completed
	Presentation	Presentation	Presentation of research results demonstrating value of dataset, at a research conference or other similar venue	N
	Report or issue brief	Publication	Research report posted on ASPE and/or ACF websites demonstrating value of dataset, available to public	N

	Manuscript	Publication	Manuscript on research demonstrating value of dataset, published in a peer reviewed journal or as a white paper	N
Objective 5: Evaluation and Sustainability Planning	Public facing memo on lessons learned	Memo	Description of lessons learned throughout project	N
	Sustainability plan	Memo	Description of sustainability of data linking beyond project	N
Objective 6: Stakeholder Engagement	Dissemination plan	Project implementation plan	Plan specifying products, timeframe, audience for dissemination	N
	Roundtable report	Public report	Report on the results of the expert roundtable	N
Other	Participation in OS-PCORTF Consortium	Quarterly Progress Report	Participation in consortium activities	Y
	Quarterly Progress Reports	Quarterly Progress reports	Quarterly progress report submissions	Y
	Final Project Report	Final Project Report	Final Project Report	Y: this document

Section 3.1. Tasks completed with T-MSIS data

As part of the Urban Institute’s support of executing a data use agreement (DUA) between CMS and HSP to access T-MSIS data in the CMS VRDC, the project team applied for “non-standard” T-MSIS data to assist with the probabilistic linking between T-MSIS and RADS. Because there is no shared unique identifier in the two datasets, the project team requested a supplementary T-MSIS dataset called the Vital Status File, which contains the full name of the Medicaid beneficiary. Director and CMS Chief Data Officer in the Office of Enterprise Data and Analytics at CMS, Allison Oelschlaeger, approved the request, and the project team therefore obtained access to this non-standard data for 2016-2022, the project’s period of study.

The Urban Institute conducted analyses of the non-standard data in their preparation to link T-MSIS with RADS. Specifically, they examined data quality for full name, which was unknown to them and other researchers who have not worked with this data, since CMS does not publish data quality information on all variables. The Urban Institute validated data by checking that length of first and last name was greater than two (that is, that there are at least three characters). This will catch most issues and overestimate problems (for example, valid names that are only two characters long first and last name). They found that data quality for full name varied by state in 2016-2018, but

beginning in 2019, there was very low missingness (~0% rounded) for name completeness (first and last name and middle initial).

Next, the Urban Institute examined the validity of the citizenship variable, which they also planned to utilize in the probabilistic linking. Though CMS publishes data quality information by state and variable in their online [Data Quality Atlas](#), the citizenship variable is not included, despite being part of the standard data extracts. Therefore, the Urban Institute assessed this variable's quality, finding that in earlier years, they were more likely to find states that do not have complete citizenship information, but by 2020, this was largely resolved (missingness  $\leq 2\%$ ). Across some of the key variables to be used in the probabilistic linking (full name, citizenship, zip code and/or county), the Urban Institute used the Vital Status File validation results to remove states with poor data quality between 2016-2022. Removing states across all years with a percent of data missingness of  $>2\%$  reduced the number of states to 16.

### Section 3.2. Tasks completed in preparation to link datasets

The Urban Institute completed an initial data linking plan in August 2024, which included an overview of linking methods, including steps for data preparation, and tradeoffs and risks associated with each linking method described. Through review of methods and comparable projects, they had planned for the linking approach to be based on probabilistic methods based on the work of Fellegi and Sunter<sup>1</sup>. Below is a summary of steps and key decision points in the linking process and in the assessment of link quality, created by the Urban Institute in preparation for data linkage:

1. Clean and standardize data
2. Develop initial blocking structure manually
3. Develop targets and ceilings for link rates if possible
4. Generate probabilities of matching individual records between datasets to be linked
  - a. **Key decision point:** Which method is most appropriate?
5. Run linking iteration and evaluate
  - a. **Key decision point:** Review weights and ensure they are understandable and explainable. Does event timing for these linked records make sense? Are there any data quality issues?
  - b. **Key decision point:** Review and develop thresholds.
  - c. **Key decision point:** Review blocking strategy. Would machine learning add any value?
6. Adjust parameters, weights and blocks, run next linking iteration, and evaluate
  - a. **Key decision point:** Compare results to any developed targets, review unlinked records for any patterns or biases and adjust for the next round as needed.
7. Repeat steps 5 and 6 until satisfied with results.

---

<sup>1</sup> Fellegi, I. P., & Sunter, A. B. (1969). A Theory for Record Linkage. *Journal of the American Statistical Association*, 64(328), 1183–1210. <https://doi.org/10.1080/01621459.1969.10501049>

8. Evaluate quality of link results for use in research
  - a. **Key decision point:** Are results usable for analysis?

At each of these key decision points, the Urban Institute anticipated challenges, but they developed strategies to overcome them, as shown in Table 3.

**Table 3.** Key decision points, challenges, and methods to overcome

Key Decision Point	Potential Challenge	Methods to Overcome
Which method to use to calculate matching probabilities	Could be highly manual and/or difficult.	Software packages and related documentation.
Reviewing weights	Hard to explain or do not seem to make sense with respect to underlying data.	Previous linking projects, understanding of the data, and experience of the full team including HSP, ORR, and Technical Expert Panel.
Developing thresholds	Could be highly manual and time intensive.	Sampling results only when needed, using experience of prior linking projects and knowledge of string comparison algorithm performance.
Blocking strategy	May not have enough variables to make machine-learning a value-add.	Start small and devise a prototype before fully committing to machine-learning.
Review results	Time intensive. Learning curve with data.  Targets may be a challenge to develop.	Previous linking projects, understanding of the data, and experience of the full team including HSP, ORR, and Technical Expert Panel.
Evaluate results for use in research	Unknown how useable the results will be.	Experience of the full team including HSP, ORR, and Technical Expert Panel.

Notes: Table adapted from initial data linking plan created by the Urban Institute, provided to HSP and ORR.

### Section 3.3. Tasks completed to prepare for data archiving

The Urban Institute produced a memo in January 2025 on potential platforms for data archiving. While they evaluated several options, the Urban Institute ultimately suggested the utilization of two platforms, Chronic Conditions Warehouse (CCW) Virtual Research Data Center (VRDC) and Inter-university Consortium for Political and Social Research (ICPSR), depending on the type of data product and anticipated level of access to the product. The dimensions of evaluation for the two platforms are shown in Table 4.

The Chronic Conditions Warehouse (CCW) Virtual Research Data Center (VRDC) is CMS’s VRDC for T-MSIS data, as well as other datasets. While it satisfies all CMS privacy and security requirements and is accessible to researchers via a multi-step data use application process, it may be cost

prohibitive for some potential users (typically at least \$25,000/year/project and often more). The Urban institute suggested utilizing the CCW VRDC to host and distribute the linked RADS-T-MSIS dataset for highly restricted use.

Inter-university Consortium for Political and Social Research (ICPSR) hosts datasets utilized for social science research, and it is based virtually at the University of Michigan. ICPSR is widely known and trusted among researchers, easily accessible, and relatively low cost compared to the CCW. The Urban Institute suggested that ICPSR host a public-use version of the linked dataset (with personally identifiable information removed), accessible via a data use agreement application. ICPSR could also host aggregated public use data, including summary statistics and other descriptive cross-tabulations.

Ultimately, the public-use data could be available in both locations: the CCW VRDC and ICPSR. This would ensure that users could balance cost and data usability considerations themselves and would ensure a broad level of access to results.

**Table 4.** Summary of dimensions for evaluation, by platform

	Usability	Accessibility	Cost	Data Security	Data products
CCW VRDC	Highly favorable	Highly favorable, though long application review process	Cost-prohibitive to many potential users	Highly favorable	Federal use links, public-use links, summary statistics
ICPSR	Favorable, but may not include important/ data elements for analysis if restricted	Highly favorable	Highly favorable	Highly favorable	Public-use links, summary statistics

Notes: Table adapted from initial data linking plan created by the Urban Institute, provided to HSP and ORR.

### Section 3.4. Tasks completed to prepare to engage stakeholders

In August 2024, the Urban Institute provided a stakeholder engagement plan to the project team, outlining how they planned to engage stakeholders throughout the data linkage project. The purpose of stakeholder engagement was to ensure that the data products can appropriately inform policy and programmatic decisions. The plan included the main activities to be accomplished with stakeholders, a timeline in accordance with specific project tasks, and the topics they will be requested to consult on. Finally, the Urban Institute compiled a list of more than two dozen potential stakeholders, including researchers, refugee service providers, advocates and state office

representatives of refugee health services. Engagement did not occur as planned due to HHS policy changes.

### Section 3.5. Other tasks completed by the HSP and ORR project team

The HSP and ORR project team also completed several tasks independent of the Urban Institute. In October 2023, HSP published a [one-page brief](#) on the ASPE website describing the project and its objectives. In November 2023, HSP presented an overview of the project at the [Association for Public Policy Analysis and Management Conference](#) in Atlanta, GA. Both the brief and the presentation generated engagement with the project from external researchers.

The project team also organized a Technical Expert Panel (TEP) of relevant federal experts to convene and consult throughout the project duration. It consisted of about 15 federal government staff, including Medicaid and refugee health experts from ORR, CMS, the National Institutes of Health (NIH), the Centers for Disease Control and Prevention (CDC), the Agency for Healthcare Research and Quality (AHRQ), the Health Resources and Services Administration (HRSA), the Substance Abuse and Mental Health Services Administration (SAMHSA), and others. HSP convened the TEP twice virtually, in March 2023 for a project kickoff meeting and in June 2024 for a project update meeting. The project team also conducted meetings with individual TEP members at their request or as the project required their consultation.

### *Section 4. Lessons Learned*

Despite the project's early termination, the project team learned several lessons which can be applied to other projects seeking to link administrative data, and especially those utilizing T-MSIS data.

First, data use agreements (DUAs) and memoranda of understanding (MOUs) between government agencies or offices require a significant amount of time to obtain and, therefore, challenges may arise if/when there are changes in leadership or signatories. DUAs and MOUs often require many levels of approval in each government agency or office involved, which means that many signatures are needed. The project team had budgeted eight months to obtain all approvals needed, and did not begin the MOU negotiation process for the transfer of RADS data from ORR to HSP until fall 2024. Ultimately, the project team did not obtain an executed MOU. Future projects seeking to obtain DUAs or MOUs between agencies or offices should consider how the timeline interacts with anticipated organizational or leadership changes and backwards plan their timeline for beginning the negotiation process.

Second, and also related to DUAs and MOUs, CMS does not need to be involved in the DUA/MOU process as it relates to datasets owned by other entities. The project team initially thought that CMS would need to approve and sign the MOU between ORR and HSP to transfer the RADS dataset into the CMS VRDC, but ultimately, this was incorrect and caused a 1.5-month delay. Future researchers seeking to access CMS data in their VRDC only need to involve CMS in the DUA that permits access to the CMS VRDC, managed by the Research Data Assistance Center (ResDAC) and CCW. Once that DUA between CMS and the primary researcher's office (in this case, HSP) is

approved, that office is permitted to upload any data into the VRDC without further CMS involvement in another DUA or MOU.

Third, for projects that include a Technical Expert Panel (TEP), there should be periodic communication with the TEP via email in between convenings. In the 15 months that passed between the kickoff and the update TEP meetings (mentioned in Sec 3.5), multiple TEP members reached out to the project team to request status updates. While these served as informal touch points with individual TEP members, it may be more efficient to plan for periodic updates to the entire TEP.

Fourth and last, tasks completed provide lessons learned. Sec 3.1 (above) details the analyses that the Urban Institute completed regarding the T-MSIS “non-standard”/very limited access Vital Status File, which includes the full name of the beneficiary. The Urban Institute found that, beginning in 2019, there was very low missingness for name completeness (first and last name and middle initial). The Urban Institute also analyzed the data quality of the citizenship T-MSIS variable and found that, beginning in 2020, there was almost no missingness. Researchers who obtain access to the Vital Status File of T-MSIS with the full name or the citizenship variable in the Demographic and Eligibility File can consider these data quality analyses if they seek to use name or citizenship status for linking purposes. Finally, Sec 3.3 discusses the data platforms with the most promise for data archiving for projects that will have linked, personally identifiable data. In the Urban Institute’s scan of archive platform options, there were very limited platforms that were usable, accessible (e.g., via a DUA or application process), and secure. It is therefore important to consider data platform options upfront for projects seeking to share data products.