Lessons in Data Linkage: Four Examples from Home Visiting Research

HARC
Home Visiting Applied Research Collaborative

April 2026

INTRODUCTION

Linking administrative data across early childhood programs and sectors has the potential to improve service delivery for families, inform program and policy decisions, and track long-term outcomes. In fall 2024, HARC (we) conducted a scoping review of studies that linked home visiting program data with other administrative data sources. We examined data sources used, methods for linkage, study designs, research questions, and outcomes examined in the literature. This review demonstrated the value of linking data to answer research questions that cannot be explored with home visiting data alone. Findings illuminated a dearth of published information about the foundational efforts that went into the planning and visioning for these studies (Heberlein et al.,2025).

To better understand this gap, we selected four examples that demonstrated the breadth of approaches to using linked home visiting data including: universal home visiting program evaluation (Massachusetts), cost-benefit analysis (Arkansas), assessment of reach using GIS data (South Carolina), and child welfare involvement among home visiting participants (Texas). We interviewed authors from each study to gain additional insights into planning and logistics, including partnerships, data access, data governance,[1] and technology that are not traditionally provided in published articles. Interview questions were based on the Research Data Framework (RDaF), previously applied in HARC's **Resource List to Support Data Sharing for Collaborative Home Visiting Research**.

Potential Benefits of Linking Home Visiting and Administrative Data

- **Improve understanding of participant needs**: Linking data across systems reveals a more complete picture of families' experiences and challenges
- **Inform program and policy decisions**: Rigorous evaluation findings using linked data can guide resource allocation, program improvements, and policy advocacy
- **Track long-term outcomes**: Administrative data enable researchers to follow families beyond their home visiting enrollment, capturing outcomes that programs cannot observe through participant self-report alone or outcomes seen in other programs

The purpose of these examples is to inform state administrators and data leads of home visiting programs about the nuances of plans and partnerships necessary to do this work while also examining the concrete challenges and practical solutions. For state administrators of home visiting programs, linking home visiting data with administrative sources yields meaningful insights for program and policy improvement but demands sustained planning and effort. Study teams faced challenges in establishing legal agreements, ensuring data quality, and accurately matching records. State data leads should consider practical strategies to these challenges by involving legal and technical experts early, building flexibility into the timeline, and dedicating resources to manual verification of linkages. Success depended on strong individual and organizational relationships between involved parties. State administrators considering data linkages should establish strong partnerships and a long-term vision that includes clear governance, dedicated funding, and a phased implementation approach. This brief highlights

---

[1] Data governance refers to the people, policies, and procedures that support how data are used and protected (Actionable Intelligence for Social Policy, 2025).

examples from each study including the: 1) organizational and study context; 2) partnerships and data governance; 3) data sources and access, including which home visiting models were included in the linkage (e.g., combining data from different models); 4) study design, methods, and analyses; 5) findings and impact on program and policy; and 6) challenges and solutions.

Evaluation of Universal Welcome Families Home Visiting Program in Massachusetts

Massachusetts Department of Public Health (DPH) was interested in evaluating Welcome Family, a universal program available in five communities. Evaluators from Tufts Interdisciplinary Evaluation Research (TIER) at Tufts University linked Welcome Family enrollment and discharge data with several other state administrative data sources to examine whether participation in Welcome Family was associated with families' linkages to Evidence Based Home Visiting (EBHV) and Early Intervention (Part C).

Partnerships and Data Governance

Massachusetts DPH and the TIER study team have previously collaborated to evaluate home visiting programs through the Massachusetts MIECHV funds. Their longstanding relationship helped facilitate access to other state administrative data and funding sources (DPH and Medicaid) for home visiting program evaluation.

DPH support was critical for negotiating data access with other state agencies. Data governance processes varied across data sources, as each state agency had its own requirements. Data sharing agreements (DSAs)[2] were reviewed by Tufts legal counsel and data security teams, and the DPH IRB[3] provided evaluation oversight. The team worked with university IT staff to acquire enough server space and computing power to manage the large datasets of healthcare claims.

Data Sources and Access

Data sources for the evaluation included:

- Welcome Family enrollment and discharge data
- Pregnancy to Early Life Longitudinal Data System (PELL), including birth and hospital utilization records
- Massachusetts All-Payer Claims Database (APCD) healthcare claims
- Early Intervention (Part C) program data
- MIECHV program data (Early Head Start, Healthy Families America, Healthy Families Massachusetts, and Parents as Teachers)

The study team navigated multiple approval processes to access data. DPH provided home visiting program data, Early Intervention (Part C) program data, and PELL data via secure file transfer. Accessing APCD healthcare claims data involved stricter requirements and multi-step applications through DPH and APCD, with APCD providing access to the data and conducting periodic data audits.

Study Design, Linkage Methods, and Analysis

The study team collaborated with state partners to plan a quasi-experimental evaluation using linked administrative data to understand the impact of Welcome Family. Welcome Family provides families with one nurse home visit after the birth of a child and serves as an entry point into the system of care for families with newborns. The study examined whether Welcome Family participants were more likely than families with similar backgrounds in the same communities to enroll in EBHV or Early Intervention (Part C) programs to understand how well Welcome Family is linking families to other resources.

---

[2] Data sharing agreements (DSAs) reflect assent between two or more parties about which data will be shared and how the data can be used (National Center for Advancing Translational Sciences, 2025). They are sometimes referred to as data use agreements or DUAs.

[3] Tufts IRB ceded IRB oversight to DPH.

Data cleaning and matching decisions proved time intensive. For large claims files, the study team partnered with Tufts data experts and used a programming language called Python to parse and manage datasets. Individual records were linked using mothers' and children's names, dates of birth, and ZIP codes to produce the study dataset.

To develop a comparison sample, the study team used coarsened exact matching (CEM)[4] with birth certificate data to construct a comparison group of non-program participants with similar demographic and maternal health characteristics (i.e., covariates) living in the same service areas.

Findings and Impact on Programs and Policy

Findings showed that compared to non-program participants, Welcome Family participants were more than twice as likely to enroll in an EBHV program by age 1 and more likely to receive Early Intervention (Part C) services by the child's third birthday. Language moderated the reach of home visiting programs; English-speaking mothers experienced significant increases in EBHV enrollment, while non-English-speaking mothers did not. This disparity highlighted persistent access challenges tied to language. The variation in EBHV enrollment by families' primary language helped identify the need to hire bilingual Welcome Family staff, rather than relying on translators. Beyond the published journal article, results were shared through briefs and presentations developed to inform state-level decision making.

Linking home visiting program data with administrative data made it possible to construct a comparison sample without needing to conduct an experimental study that would exclude families from accessing the universal program. Administrative data also provided outcomes that may be influenced by home visiting but are not available in home visiting program data. The study demonstrated the feasibility and value of using linked administrative data for rigorous evaluation of universal home visiting programs. After addressing the original study questions, and in accordance with IRB and DSA requirements, study findings led to additional analyses. The follow-up analyses produced actionable findings that informed DPH and their discussions with MassHealth, the state Medicaid and Children's Health Insurance Program provider.

Related Journal Article

Kotake, C., Fauth, R. C., Stetler, K., Goldberg, J. L., Silva, C. F., & Manning, S. E. (2023). Improving connections to early childhood systems of care via a universal home visiting program in Massachusetts. *Children and Youth Services Review*, 150. https://doi.org/10.1016/j.childyouth.2023.106995

Evaluation of Child Maltreatment Involvement for Home Visiting Programs in Texas

As home visiting expanded throughout Texas, the state's home visiting team grew interested in understanding the impact of home visiting for families across the state. This evaluation looked at home visiting's impact on child maltreatment in the first 2 years of life.

Partnerships and Data Governance

This study built on an existing evaluation and consulting partnership between researchers at the University of Texas at Austin (UTA) and the Texas MIECHV program, first established in an implementation study of Texas home visiting programs in 2012. Through this initial partnership, the study team linked home visiting data from different home visiting models to support MIECHV performance measure reporting.

The study team at UTA worked with the Texas MIECHV program to develop research questions. The resulting evaluation, primarily funded by MIECHV evaluation funds, sought to understand the relationship between home visiting program participation and child welfare involvement and birth outcomes.

---

[4] Coarsened Exact Matching (CEM) describes a method that groups participants with the same values (raw or coarsened) on each of the covariates into strata that include at least one home visiting participant and one comparison group participant. CEM generates analytic weights for covariates to balance distribution across two groups to compare outcomes across groups.

Beyond partnering with state home visiting programs, the study team had also developed evaluation relationships with departments overseeing child welfare and vital statistics. This previous work helped demonstrate that the team could manage identifiable data securely and use data respectfully. In 2014, Texas moved home visiting programs to the same department as child welfare as part of a larger restructuring of the Health and Human Services Commission (HHSC), making it easier to get DSAs to link program data. The process for developing DSAs was also simplified because UTA is a state agency.

The study team worked with the UTA IT department to set up the technological components needed to accommodate the type of data being used, the required storage size, and access permissions. This included working with the Texas Advanced Computing Center at UTA to access the server storage space needed for the child welfare and vital statistics datasets. As part of IRB approval and DSAs, the research team also developed a data security plan delineating which staff could access the identifiable datasets and the process to de-identify the dataset after the linkage was completed.

Data Sources and Access

Data sources for the study included:

- Vital statistics birth records
- MIECHV program data (Nurse-Family Partnership, Parents as Teachers, and Home Instruction for Parents of Preschool Youngsters)
- Child welfare data for investigated and substantiated reports of child maltreatment

The study team developed a DSA between UTA and HHSC to access administrative data as part of its evaluation contract with the Texas MIECHV program.

Study Design, Linkage Methods, and Analysis

To understand the relationship between participation in home visiting and documented child maltreatment, the study team designed a quasi-experimental study using sequential logistic regression to compare program participants to non-participants. After acquiring home visiting and child welfare data, the study team spent significant time understanding and cleaning the data. The team’s previous experience with child welfare data helped members understand whether the data were accurate and complete; theiir deep knowledge of the home visiting programs helped members identify issues in data reporting practices.

Records were linked using exact and fuzzy matching;[5] the latter method allowed the team to link more records by accounting for data entry errors or inconsistencies that may not produce an exact match. The team set a lower limit for weighted match scores. Matches with scores below this limit were then manually reviewed. The study team achieved an 86% match rate for children enrolled in home visiting programs with their birth records with graduate student assistants supporting the matching process. Despite the significant time required to manually review matches, the study team found the review critical to include as many children as possible. Once the data had been cleaned and linked, the study team used CEM to develop a comparison group of families who had not participated in home visiting programs but had similar characteristics to program participants.

Findings and Impact on Programs and Policy

The study found that while families participating in home visiting were significantly more likely to experience a child maltreatment investigation before their child’s second birthday, they also had significantly lower odds of having their first child maltreatment investigation substantiated in comparison to families who did not participate in home visiting. Overall, families participating in home visiting did not differ significantly from families who had not participated in home visiting in their likelihood of experiencing a substantiated investigation over 2 years.

---

[5] Fuzzy matching is a method of linking data records using a combination of non-unique variables (e.g., last name, birth date) to calculate weights that describe the likelihood of two records being a match.

Linking home visiting program data with administrative data made it possible to answer questions about child welfare involvement that home visiting programs could not have answered with home visiting data alone. Home visiting programs typically rely on caregiver reports for data related to child welfare involvement; they also typically lose contact with families no longer in the program, limiting their ability to identify long-term outcomes.

The study team worked with home visiting programs to improve data quality, recommending changes such as creating a data structure to account for multiple caregivers. This recommendation improved data collection for fathers and thus, the ability to conduct evaluation on father engagement efforts. Analyses also identified a need for training on mandated reporting and how to discuss child welfare involvement with families.

Related Journal Article

Doe, H. A., Osborne, C., Huffman, J., Craig, S. M., & Shero, M. (2024). Home visiting and child welfare involvement: A matched comparison group study. *Child Maltreatment, 30*(2), 331–342. https://doi.org/10.1177/10775595241268227

Evaluating Reach of MIECHV-Supported Services in South Carolina

This study sought to assess how MIECHV-supported services were reaching families in communities facing high levels of adversity by identifying these communities at the ZIP code level. It built on a longstanding research and evaluation partnership between the University of South Carolina Rural Health Research Center (RHRC) and the Children's Trust of South Carolina (the Children's Trust), the state MIECHV program administrator.

Partnerships and Data Governance

The close partnership between RHRC and the Children's Trust helped the study team understand program needs and make its work relevant and useful for policy and program management. This partnership has benefitted from access to South Carolina's statewide integrated data system. Initiated by statute, the system is managed through interagency agreements covering a range of policy domains (e.g., education, child welfare, Medicaid enrollment and claims, vital records) and offers both a public retrieval system for aggregated data downloads and a request process for linked datasets. Data from the integrated system were used for the MIECHV needs assessment and used annually for reporting the MIECHV child maltreatment performance measure.

The Children's Trust funds RHRC to conduct MIECHV evaluations and report performance measures. The Children's Trust also manages the home visiting databases that store MIECHV performance measure data and other required reporting information. The state agency and university legal teams ensure that required elements and procedures for data linkage, transfer, and storage are included in DSAs between all entities (i.e., Revenue and Fiscal Affairs [RFA] Data Integration and Analysis Division, RHRC, Children's Trust).

The highlighted study used publicly available data from RFA so did not need to go through the standard request process. However, as part of RHRC's ongoing evaluation work with the Children's Trust, RHRC can submit research requests for linked data from RFA. The standard request process for linked datasets is not time intensive and is generally free or low cost. Requests require an official application and if Medicaid data will be linked, a Medicaid board review.

Data Sources and Access

Data sources for this study included:

- MIECHV program data (Healthy Families America, Nurse-Family Partnership, Parents as Teachers), specifically participant address data and demographic characteristics
- U.S. Census Bureau Topologically Integrated Geographic Encoding and Referencing (TIGER) files, address and boundary files for geocoding addresses to ZIP Code Tabulation Areas (ZCTAs)
- Indicators of family risk, aggregated at the ZIP code level (publicly available from the RFA data warehouse)
- County-level birth volume (publicly available from the RFA data warehouse)

Aggregated ZIP code data were downloaded from the publicly available South Carolina Community Assessment Network (SCAN) data retrieval system, which pulls from the RFA data warehouse. Home visiting household data were accessed from the statewide MIECHV database. Having MIECHV data consolidated in one database simplified this process rather than pulling data from multiple model data systems. TIGER files are available from the U.S. Census Bureau.

Study Design, Linkage Methods, and Analysis

The study team designed a retrospective, cross-sectional study using ZIP code-level analyses to examine whether MIECHV-supported services reach families living in communities facing high levels of adversity and to identify service gaps.

More than 3,000 (3,229) households were enrolled in MIECHV-supported home visiting services between 2012 and 2016. Residential addresses for program participants were reviewed to exclude missing or out-of-state addresses or PO boxes. Addresses were geocoded using ArcMap from ArcGIS. The study team reviewed any addresses with a match below 80%, corrected grammatical errors, and attempted to re-match entries. Households geocoded to a location outside of their county of residence were excluded. This resulted in a final analytic sample of 3,090 households. The RHRC used ArcMap to map MIECHV households to the ZIP codes and to present findings on indicators of risk (proportions of mothers who are unmarried, had Medicaid deliveries, are African-American, are under age 20, have less than a high school education, used tobacco during pregnancy, had inadequate prenatal care, received WIC during pregnancy, or delivered an infant who was preterm or low birthweight) by ZIP code in maps.

Because this project used indicators of risk at the ZIP code level rather than the individual level, the study team was able to access these publicly available data via the RFA warehouse.

Findings and Impact on Programs and Policy

Findings from the study indicated gaps in the delivery of home visiting services, despite a pre-implementation needs assessment, and led to specific recommendations for expanding services. Study results were shared with the Children's Trust and published in a peer-reviewed journal.

Linking home visiting program data with administrative data provided information not only on the population of families receiving home visiting services, but the population of families who might benefit from these services. South Carolina's integrated data system enabled the study team to show how community-level administrative data can guide home visiting resource allocation—one of the first demonstrations of its kind. The research has had direct policy impact, helping target resources, informing state-level advocacy for additional resources, and promoting MIECHV reauthorization at the federal level by showcasing home visiting's contribution to maternal and child health.

Sharing Actionable Results

RHRC prioritizes dissemination to help findings drive policy and so that other states can replicate their studies. Results are presented in annual MIECHV reports, one-page infographics, and executive summaries. The Children's Trust uses the results to advocate for change when meeting with state policymakers, demonstrate program impact, and support federal reauthorization efforts.

This study highlighted the benefits of research center and MIECHV program partnerships with access to a best-in-field integrated data system. The collaboration between the RHRC and Children's Trust continues, with the ongoing goal of disseminating findings that show home visiting's need and impact, and how to reach more families more effectively.

Related Journal Article

Radcliff, E., Breneman, C. B., Crouch, E., & Baldwin, I. (2019). Are we serving the most at-risk communities? Examining the reach of a South Carolina home visiting program. *Journal of Community Health, 44*(4), 764–771. https://doi.org/10.1007/s10900-018-00606-5

Healthcare Cost-Benefit Analysis of the Following Babies Back Home Model in Arkansas

Following Baby Back Home (FBBH) is a home visiting model that supports families whose infants have been discharged from the Neonatal Intensive Care Unit (NICU). This evaluation sought to demonstrate the model's effectiveness in reducing healthcare costs and improving health outcomes for this sub-population. The study team at the University of Arkansas for Medical Sciences has conducted evaluations for the model since its inception in 2009, when it was initially funded by a state Medicaid contract.

> *"Administrative data represent wonderful opportunities to be able to identify good comparison groups, in … a timely way. You don't have to wait 10 years for a longitudinal trial." – Dr. McKelvey*

Partnerships and Data Governance

This study, primarily funded by MIECHV evaluation funds, relied on an existing partnership between the team and FBBH. Before MIECHV was authorized by federal legislation in 2010, FBBH had established a process to include Medicaid data in evaluation efforts. This step enabled continued use of Medicaid data in subsequent evaluations for MIECHV. The Arkansas Biosciences Institute provided additional funding. The Arkansas Center for Health Improvement (ACHI) serves as an honest broker[6] for the Arkansas All-Payer Claims Database established via the Arkansas legislature with set-aside funding. The Arkansas Biosciences Institute has an agreement with ACHI that researchers can apply through the institute to access the database for analysis. The FBBH study application included its IRB approval and a research plan informed by available data demonstrating the model's influence on participant healthcare utilization, cost, and health outcomes.

Data Sources and Access

Data sources for the study included:

- FBBH model data
- Medicaid claims and enrollment
- Vital statistics birth records
- All-Payer Claims

The study team accessed the home visiting program data through an evaluation contract with the Department of Public Health and provided participant identifiers to ACHI via secure file transfer.

Study Design, Linkage Methods, and Analysis

The study team designed a quasi-experimental study that compared outcomes of interest for FBBH participants to those of non-participants. ACHI provided the study team with guidance on how to format their program data to meet their linkage requirements and to produce the best possible match rate.

FBBH records provided to ACHI by the study team were linked by ACHI to produce the study population sample and to create anonymized IDs of program participants. ACHI gave a list of anonymized IDs back to the study team. The study team accessed the anonymized health insurance records via a Virtual Private Network (VPN) managed by ACHI and were able to use the IDs to identify FBBH participants. The final study population included 416 children enrolled in FBBH. The study team then developed a comparison sample using propensity score matching to identify a matched control group of 382 children with similar demographic, socio-economic, and medical characteristics to the enrollees.

---

[6] Honest broker describes a neutral third party that removes sensitive information to deidentify participant data and facilitate research while protecting private information.

Findings and Impact on Programs and Policy

Findings demonstrated the efficacy of FBBH for families of high-risk, low birth weight preterm infants after hospital NICU discharge and supported its implementation for opioid-affected infants. The team calculated savings of $27.95 for every dollar spent on FBBH, with a net economic benefit of $83,020 for avoided deaths, program costs, and healthcare costs. The study also reported comparable medical care costs incurred by FBBH infants and the comparison group, despite the increased medical risk of FBBH infants. These findings expanded on previous study results, which found that infants enrolled in FBBH were more likely to have higher numbers of medical appointments and to be more compliant with immunization schedules.

Linking home visiting program data with administrative data made it possible to compare health outcomes for FBBH program participants to infants who were not enrolled in the program.

Related Journal Article

Lewis, K. N., Tilford, J. M., Goudie, A., Beavers, J., Casey, P. H., & McKelvey, L. M. (2023). Cost-benefit analysis of home visiting to reduce infant mortality among preterm infants. *Journal of Pediatric Nursing, 71*, e112–e119. https://doi.org/10.1016/j.pedn.2023.05.003

CHALLENGES AND SOLUTIONS

The study authors interviewed for this brief identified challenges and solutions throughout the data linkage processes. This section presents their insights, organized across the following categories: resources and timeline, legal requirements and agreements, data quality and preparation, and linking records. It also specifies the relevant RDaF stages and substages to consult in the **HARC 2024 list of resources** to help address these challenges.

Resources and Timeline

Challenges	Potential Solutions
• Significant time needed to nurture partnerships and trust between state home visiting programs and agencies with available data sources • Delays in progress due to shifts in agencies' priorities and timelines and legal and security reviews • Lack of funding and time to complete the work required, including time needed to form DSAs and acquire data • Limited shelf life of one-time linkage projects (i.e., those that are not part of an ongoing integrated data system) for informing programs as data become outdated • Significant server space and computational power required for large datasets such as healthcare claims	• Identifying key contacts, familiarizing them with the data linkage process, and introducing them to one another early in the process to navigate legal and technology needs • Building in flexibility for timelines, assuming that things will take longer than planned • Reviewing preliminary data early and holding discussions between agencies and the study team to understand nuances • Acquiring data in waves (e.g., by program years, one dataset at a time) to begin data inventory and cleaning while navigating legal and technology needs for other datasets • Documenting all aspects of the process and engaging with agencies early and often about ongoing collaboration needs, and using this documentation and communication to keep DSAs up to date and relevant to program needs
RDaF Stages and Substages: Envision (buy-in & data culture, collaboration, funding), **Plan** (data governance, technology & personnel)	

Legal Requirements and Agreements

Challenges	Potential Solutions
• Data sources that have been de-identified or that lack variables needed to link files (e.g., addresses for geocoding) • Different standards from different data clearinghouses for the same variables • Lengthy application processes for data access and burdensome processes to request additional variables or revision (e.g., changing how program dosage is provided) • Different interpretations of data sharing rules by lawyers from different institutions (e.g., interpretations of FERPA requirements for sharing education data with identifiers for matching) • Misalignment of academic needs for publication with agency rules for publicly releasing data analyses, especially for results with null or unfavorable findings or cases with known data limitations (e.g., incomplete, inconsistent reporting)	• Facilitating regular discussions between the study team, data management experts, and agency data managers • Maintaining separate DSAs for each dataset, so updates for one DSA do not delay the full project • Developing detailed data security plans to simplify IRB and other governance applications • Leveraging evaluation requirements to support data access (e.g., state Medicaid agencies may be open to contributing claims data to support evaluations of home visiting programs that receive Medicaid funding) • Exploring opportunities for studies using data from sources housed within the same agency to streamline the DSA process • Discussing academic publication needs with programs and agencies early and throughout the process
RDaF Stages and Substages: Plan (legal considerations, data governance), **Acquire** (establishing data exchange standards)	

Data Quality and Preparation

Challenges	Potential Solutions
• Incomplete or inconsistent home visiting datasets • Unique analytic challenges from data sources such as healthcare claims, including bundled payments for prenatal care (which limits study potential for adequacy of prenatal care measures and attendance at postpartum visits), and inconsistent reporting of specific diagnoses or health conditions • Data purging policies (e.g., deletion of low-risk child welfare cases after 18 months) that limit the look-back period for retrospective studies	• Collaborating with programs to identify critical data fields and understand field-specific data entry practices • Partnering with programs to help staff who collect, enter, and submit data understand the purpose, benefits, and requirements for high-quality data and identify how to improve processes • Engaging in feedback loops with families, home visitors, and programs whenever possible to share findings back to them and increase their buy-in • Investing time upfront in data cleaning and validation to pinpoint issues in data collection forms that can be resolved to save time • Facilitating discussions between research and administrative data team experts before applying for data to understand variable definitions and

Challenges	Potential Solutions
	limitations upfront and to decide on the best approach to variable construction
RDaF Stages and Substages: Acquire (establishing data exchange standards, data inventory)	

Linking Records

Challenges	Potential Solutions
• Significant time needed to review records that were expected to link but did not • Inconsistent match rates for uncommon names or spelling variations • Difficulty matching mothers and infants due to name changes related to parent marital status or multiple child aliases • Potential underreporting of miscarriages • Multiple records to match and reconcile due to families enrolling in home visiting multiple times with different children or re-enrolling in programs • Higher rates of missing birth dates for infants enrolled prenatally compared to families who enrolled after the birth of the child	• Developing a review protocol with rules for exact versus approximate matches[7] and building in time to manually review matches to increase linkage rates and reduce linkage errors • Hiring and training university students for manual review processes in a mutually beneficial arrangement that provides real-world experience managing large datasets • Developing a matching protocol that considers name commonality and group differences in spelling and family structures • Obtaining detailed documents from data contributors explaining how they create identifiers • Building cross-disciplinary study teams (e.g., statistics, GIS, epidemiology) to maximize learnings • Checking for documentation that may detail secondary names or aliases from data providers • Using estimated due date windows when birth dates are unavailable
RDaF Stages and Substages: Process or Analyze (data linkage, analysis)	

Systems-Level Barriers to Linked Data Research

Interviewees noted several system-level barriers in need of coordinated action across the field because they are unlikely to be resolved by individual research teams and state home visiting programs.

Fragmented data systems: Home visiting data are scattered across multiple model, agency, funder, and state-specific systems with limited interoperability, making cross-model analysis difficult and increasing data entry burden for home visitors.

Lack of alignment across early childhood programs: Inconsistent data collection practices, variable definitions, and outcome measures across federal and state early childhood programs complicate data linkage and limit comparability of findings.

---

[7] An exact match occurs when all matching variables in a record match exactly across data sources (e.g., date of birth, first and last name, ZIP code). Approximate matches occurs when some variables match but others do not (e.g., birth year is a year off, ZIP code is missing in one source). Depending on the matching protocols in the study design, approximate matches may require manual review.

Misaligned incentives for data sharing: Current structures do not consistently provide benefits to programs that share data, such as reduced reporting burden, enhanced data literacy support, or opportunities for staff to contribute to peer-reviewed publications.

Inconsistent study quality standards: Definitions of study quality criteria around baseline equivalence do not always align across research and evaluation clearinghouses and could limit recognition of rigorous quasi-experimental designs using linked administrative data, particularly for population-specific models where randomized controlled trials may be infeasible.

Absence of field-wide standards: Lack of consensus on using healthcare claims data to define diagnoses and conditions reduces continuity and comparability across studies and publications.

CONCLUSION

The profiled studies demonstrate that linking home visiting program data with other administrative data sources is both feasible and valuable for rigorous evaluation, despite significant resource and coordination requirements. All projects depended on strong, trusting partnerships between researchers and state agencies, often built over years of collaboration. Despite their diverse approaches—universal program evaluation (Massachusetts), cost-benefit analysis (Arkansas), GIS study of home visiting reach (South Carolina), and study of child welfare involvement (Texas)—the studies showcase common elements of leveraging existing relationships, using MIECHV funding, and investing substantial time in data governance and quality improvement. Each effort also produced actionable findings that informed program improvements, policy decisions, and resource allocation at the state level. Studies resulted in at least one published journal article each, as well as policy briefs, infographics, presentations, performance measure reporting, and internal reports for quality improvement.

The barriers identified across projects highlight the complexity of data linkage work, particularly related to legal agreements, data quality, and matching processes. The solutions employed by research teams—such as engaging IT and legal experts early, building in timeline flexibility, and investing in manual review processes—provide a practical roadmap for others to follow. Notably, the centralized data warehouse in South Carolina and All-Payer Claims Database in Arkansas enabled more streamlined, cost-effective access to linked data. Integrated data systems also reduced the burden of negotiating multiple DSAs, facilitated ongoing data updates for continued evaluation, and allowed researchers to extend analyses through simple addendums. While these systems are typically developed for program management and coordination rather than evaluation purposes (the Arkansas and South Carolina systems were established by state statute), they benefit research by reducing administrative burden and enabling more timely, iterative analyses. The contrast between states with and without such infrastructure underscores how investments in integrated early childhood data systems could significantly reduce barriers and expand opportunities for future evaluations.

As the home visiting field moves forward, the studies highlighted in this brief underscore the importance of viewing data linkage not as a one-time project but as an ongoing, collaborative process that benefits multiple groups. Partnerships among researchers, evaluators, and community groups strengthen data quality, build state capacity for data-driven decision making, provide technical assistance to programs, and demonstrate home visiting's impact on child and family outcomes in ways that inform local policy and practice. Keys to success include a commitment to long-term partnerships, sustained funding, and attention to technical and governance challenges, linked data approaches can become a more widely used tool for advancing home visiting research and improving services for families.

ACKNOWLEDGEMENTS

We extend our thanks to the researchers, evaluators, and program partners who participated in interviews for the case studies featured in this brief. Their generosity in sharing time, expertise, and candid insights into the realities of planning, governing, and implementing data linkage efforts was essential to this work and greatly enriched the case studies presented here. We also thank our federal colleagues—Maggie Kane, Soohyun Kim, Aleta Meyer — for their thoughtful review and constructive comments on earlier drafts of this document, which strengthened the clarity and relevance of the brief. In addition, we are grateful to Jill Filene and Joelle Ruben for their careful review of multiple drafts and for their substantive feedback, which helped shape and refine the content and framing of this brief.

SUGGESTED CITATION

Bruning, J., Heberlein, E., & Sturmfels, N. (2026). *Lessons in Data Linkage: Four Examples from Home Visiting Research*. Home Visiting Applied Research Collaborative.

REFERENCES

Actionable Intelligence for Social Policy. (2025). Quality framework for integrated data systems: Governance. University of Pennsylvania. https://aisp.upenn.edu/quality-framework-for-integrated-data-systems/

Doe, H. A., Osborne, C., Huffman, J., Craig, S. M., & Shero, M. (2024). Home visiting and child welfare involvement: A matched comparison group study. *Child Maltreatment,* 30(2), 331–342. https://doi.org/10.1177/10775595241268227

Heberlein, E., Sturmfels, N., & Bruning, J. (2025). A scoping review of home visiting research using linked data. *Home Visiting Applied Research Collaborative*. https://hvresearch.org/resources/a-scoping-review-of-home-visiting-research-using-linked-data/

Hanisch, R., Kaiser, D., Yuan, A., Medina-Smith, A., Carroll, B. and Campo, E. (2024). NIST research data framework (RDaF): Version 2.0, Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, [online], https://doi.org/10.6028/NIST.SP.1500-18r2, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=957222

Kotake, C., Fauth, R. C., Stetler, K., Goldberg, J. L., Silva, C. F., & Manning, S. E. (2023). Improving connections to early childhood systems of care via a universal home visiting program in Massachusetts. *Children and Youth Services Review*, 150. https://doi.org/10.1016/j.childyouth.2023.106995

Lewis, K. N., Tilford, J. M., Goudie, A., Beavers, J., Casey, P. H., & McKelvey, L. M. (2023). Cost-benefit analysis of home visiting to reduce infant mortality among preterm infants. *Journal of Pediatric Nursing*, 71, e112–e119. https://doi.org/10.1016/j.pedn.2023.05.003

National Center for Advancing Translational Sciences. (2025). Toolkit for patient-focused therapy development: Data sharing agreement. U.S. Department of Health and Human Services. National Institutes of Health. https://toolkit.ncats.nih.gov/glossary/data-sharing-agreement/

Radcliff, E., Breneman, C. B., Crouch, E., & Baldwin, I. (2019). Are we serving the most at-risk communities? Examining the reach of a South Carolina home visiting program. *Journal of Community Health,* 44(4), 764–771. https://doi.org/10.1007/s10900-018-00606-5