Cheaper, Faster, Better:
Are State Administrative Data the Answer?

The Mother and Infant Home Visiting Program Evaluation-Strong Start
Second Annual Report

OPRE Report 2015-09

January 2015

Cheaper, Faster, Better: Are State Administrative Data the Answer? The Mother and Infant Home Visiting Program Evaluation-Strong Start Second Annual Report

OPRE Report 2015-09

January 2015

Authors: Helen Lee, Anne Warren, and Lakhpreet Gill, MDRC

Submitted to:

Nancy Geyelin Margie, Project Officer
Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Project Directors: Virginia Knox and Charles Michalopoulos
MDRC
16 East 34th Street
New York, NY 10016

Contract Number: HHSP23320095644WC

This report is in the public domain. Permission to reproduce is not necessary.

Suggested citation: Lee, Helen, Anne Warren, Lakhpreet Gill (2015). *Cheaper, Faster, Better: Are State Administrative Data the Answer? The Mother and Infant Home Visiting Program Evaluation-Strong Start Second Annual Report.* OPRE Report 2015-09. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Disclaimer: The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

This report and other reports sponsored by the Office of Planning, Research and Evaluation are available at http://www.acf.hhs.gov/programs/opre.

The work in this publication was performed under Contract No. HHSP23320095644WC awarded by the U.S. Department of Health and Human Services (HHS) to contractor MDRC and subcontractors James Bell Associates, Johns Hopkins University, Mathematica Policy Research, and New York University.

The content of this publication does not necessarily reflect the views or policies of HHS, nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. government.

Dissemination of MDRC publications is supported by the following funders that help finance MDRC's public policy outreach and expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: The Annie E. Casey Foundation, The Harry and Jeanette Weinberg Foundation, Inc., The Kresge Foundation, Laura and John Arnold Foundation, Sandler Foundation, and The Starr Foundation.

In addition, earnings from the MDRC Endowment help sustain our dissemination efforts. Contributors to the MDRC Endowment include Alcoa Foundation, The Ambrose Monell Foundation, Anheuser-Busch Foundation, Bristol-Myers Squibb Foundation, Charles Stewart Mott Foundation, Ford Foundation, The George Gund Foundation, The Grable Foundation, The Lizabeth and Frank Newman Charitable Foundation, The New York Times Company Foundation, Jan Nicholson, Paul H. O'Neill Charitable Foundation, John S. Reed, Sandler Foundation, and The Stupski Family Fund, as well as other individual contributors.

For information about MDRC and copies of our publications, see our website: www.mdrc.org.

THIS PAGE INTENTIONALLY LEFT BLANK

Overview

Home visiting programs, which provide individualized education, support, and referral resources, often serve disadvantaged women whose families are at risk for adverse health outcomes. The Mother and Infant Home Visiting Program Evaluation-Strong Start (MIHOPE-Strong Start) is examining the effectiveness of home visiting services on improving birth and maternal health outcomes for women who are enrolled in Medicaid or the Children's Health Insurance Program (CHIP), as well as their effectiveness at reducing costly health care encounters. MIHOPE-Strong Start is sponsored by the Center for Medicare and Medicaid Innovation (CMMI) of the Centers for Medicare and Medicaid Services (CMS), the Office of Planning, Research and Evaluation (OPRE) in the Administration for Children and Families (ACF), and the Maternal and Child Health Bureau (MCHB) of the Health Resources and Services Administration (HRSA). The study uses a random assignment design, and is conducted by MDRC in partnership with James Bell Associates, Johns Hopkins University, and Mathematica Policy Research.

This report details MIHOPE-Strong Start's process of acquiring administrative vital records and Medicaid data from 20 states and more than 40 state agencies. The study relies on administrative data to measure infant and maternal health, health care use, and cost outcomes. Policymakers have increasingly encouraged greater access to and use of administrative data to produce timely, rigorous, and lower-cost evaluations of health and social programs, since these records may be less costly and more accurate than information collected directly from families. The MIHOPE-Strong Start experience sheds light on the process of acquiring permission to access such data. Specifically:

- **As of this report's writing, legal agreements to acquire administrative records have been successfully reached with 19 different state agencies in 20 states.** Among these first 19 agreements, it has taken an average of 11 months to complete all the necessary steps to execute them.

- **The MIHOPE-Strong Start team has faced a number of challenges in gaining access to administrative records, the most significant of which is time.** Administrative data acquisition — beginning with initial outreach, progressing to reviews of application materials and informed consent forms, and culminating with a legal agreement to receive the data — can be a lengthy process, but the time frame is highly variable. In MIHOPE-Strong Start, the process has lasted anywhere from a few months to two years. And not all states are willing to share the Medicaid and birth record data with personal identifiers that MIHOPE-Strong Start requires.

- **Lags in data availability can be lengthy as well.** For vital records, the wait time can be longer than two years from the beginning of the calendar year in which the birth occurred.

- **The process may be easier with state agency support or with the use of national databases.** Administrative data acquisition is likely to be smoother if the research study is supported or commissioned by the state agency that collects the data needed. Alternatively, national administrative databases can potentially help overcome the hurdles associated with accessing administrative data across multiple state agencies. Delays in the availability of these data and limitations in linking them to other data sources made them unusable for MIHOPE-Strong Start.

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

Overview v
List of Exhibits ix
Acknowledgments xi
Executive Summary ES-1

Introduction 1

Overview of MIHOPE-Strong Start 4

Using Administrative Data in Research 7

Description of the Process Used to Collect Data 13

Data Availability 24

Discussion 29

References 33

THIS PAGE INTENTIONALLY LEFT BLANK

List of Exhibits

Table

1 Administrative Data Acquisition Summary 17

Figure

ES.1 Data Acquisition Process: Typical Length of Time to Complete Each Step ES-3

1 Section of U.S. Standard Certificate of Live Birth, 2003 Revision 9

2 Data Acquisition Process: Typical Length of Time to Complete Each Step 15

3 Availability of Birth Certificate Data to Researchers, After the Beginning of a Calendar Year 25

Box

1 Collecting Vital Records: Making an Initial Outreach and Submitting an Application 18

2 Collecting Medicaid Data: Modifying an Informed Consent Form and Completing a Legal Agreement 22

THIS PAGE INTENTIONALLY LEFT BLANK

Acknowledgments

We would like to acknowledge a number of people who offered guidance on the structure and content of this report. We received thoughtful comments on early drafts from Virginia Knox, John Martinez, Charles Michalopoulos, and Alice Tufel at MDRC and from Carolyn Hill at Georgetown University. This report also benefited from the valuable contributions of Rebecca Hughes at MDRC on MIHOPE-Strong Start state selection and site recruitment, and we received helpful insights from Stephen Freedman, Desiree Alderson, and Bret Barden at MDRC regarding MDRC's experience with acquiring administrative data. Ellen Fried, MDRC's legal consultant for MIHOPE-Strong Start, provided insights and assistance with developing the technical language in this report.

The administrative data acquisition efforts for MIHOPE-Strong Start would not have been possible without the instrumental contributions from Noemi Altman, Megan Millenky, and Katie Rue at MDRC as well as Liz Clary and Brittany English at Mathematica Policy Research, who were directly responsible for outreach to states' data agencies. Additionally, Jennifer Somers and Katie Egan at MDRC worked on applications to state and site Institutional Review Boards. Keith Kranker and Mary Kay Fox at Mathematica Policy Research and Todor Mijanovich at New York University shared their expertise working with Medicaid and vital records data and contributed to our understanding of acquiring these data sources.

The report also reflects suggestions from the staff at the Centers for Medicare and Medicaid Services (CMS), the Administration for Children and Families (ACF), and the Health Resources and Services Administration, including Nancy Geyelin Margie and Lauren Supplee from ACF and Caitlin Cross-Barnet and Susan Jackson from CMS.

Finally, Suzanne Finkel at MDRC provided excellent assistance with all aspects of producing the report. Katie Rue assisted with producing exhibits in this report. Jennie Kaufman edited the report, and it was prepared for publication by Carolyn Thomas and Stephanie Cowell.

The Authors

THIS PAGE INTENTIONALLY LEFT BLANK

Executive Summary

Adverse birth outcomes, including low birth weight and preterm births, are strong determinants of compromised health and development in infancy and early childhood. These outcomes are also financially costly to both families and the Medicaid program, which provides health care for many low-income families. Home visiting, which provides individualized education, support, and referrals to community resources, has been found to improve prenatal and infant health when provided to expectant women and families with infants. Because home visiting programs often serve women who may be otherwise disconnected from health and social safety net services, these programs reach the most socially isolated families, who have high levels of unmet health and other social service needs.

The Mother and Infant Home Visiting Program Evaluation-Strong Start (MIHOPE-Strong Start) is the largest study to date to examine the effectiveness of home visiting services on improving birth outcomes and infant and maternal health care use. By including a national sample of pregnant women on Medicaid, the study also hopes to provide information on whether home visiting programs can reduce short-term Medicaid costs. The study is being sponsored by the Center for Medicare and Medicaid Innovation (CMMI) of the Centers for Medicare and Medicaid Services (CMS) and the Office of Planning, Research and Evaluation (OPRE) in the Administration for Children and Families (ACF) in partnership with the Maternal and Child Health Bureau (MCHB) of the Health Resources and Services Administration (HRSA). The study is being conducted by MDRC in partnership with James Bell Associates, Johns Hopkins University, and Mathematica Policy Research.

MIHOPE-Strong Start is studying the effects of home visiting on Medicaid recipients for two national models with prior evidence of improving birth outcomes: Healthy Families America (HFA) and Nurse-Family Partnership (NFP).[1] Both HFA and NFP have been widely implemented across the country, although previous studies documenting impacts have been constrained to a few communities or are somewhat dated. In order to provide the most reliable estimates of the effects of home visiting, eligible women are being randomly assigned either to a program group, which can receive home visiting services from the program sites in the study,

[1]To determine which national models are evidence-based, the U.S. Department of Health and Human Services (HHS) funded the Home Visiting Evidence of Effectiveness (HomVEE) review, conducted by Mathematica Policy Research (Avellar and Paulsell 2011), which assessed the quality of the research evidence and documented impacts of home visiting programs on a range of domains, including child health, child development and school readiness, maternal health, positive parenting behaviors, child maltreatment, crime and domestic violence, family economic self-sufficiency, and referrals and coordination. Sarah Avellar and Diane Paulsell, *Lessons Learned from the Home Visiting Evidence of Effectiveness Review* (Princeton, NJ: Mathematica Policy Research, Inc., 2011).

or to a control group, which receives referrals to other services in the community. In addition to estimating impacts, MIHOPE-Strong Start is studying how local programs operate and are implemented to improve birth and health care outcomes.

Information on birth outcomes, health care use, and associated costs will come from state vital (birth certificate) records and Medicaid systems. Administrative data are primarily designed to allow for the effective administration of programs and services, not necessarily for research purposes. However, policymakers, program administrators, and researchers have increasingly emphasized the potential value of using — and are thus encouraging the broader sharing of — administrative records already collected and housed by state agencies to evaluate a range of complex health and social programs more quickly, effectively, and cost-efficiently. A detailed delineation of the administrative data acquisition process in MIHOPE-Strong Start can inform this broader discussion.

The current report describes the process of acquiring administrative data, specifically Medicaid data and vital records, across 20 different states and over 40 agencies.[2] Among the key findings are the following:

- **Gaining access to administrative records across numerous states and across different agencies within a state is a many-faceted process** (Figure ES.1). For MIHOPE-Strong Start, this process typically includes a series of conversations with a relevant contact at each agency, getting approval for the study's informed consent form, submitting an application to access data, and entering into a legal agreement. The procedural step of getting approval from a review committee for MIHOPE-Strong Start to obtain data is somewhat surprising in that most data agencies do not defer to MDRC's Institutional Review Board, which means that the study has to go through multiple reviews by different state departments and data agencies, at times resulting in changes to the study's informed consent form.

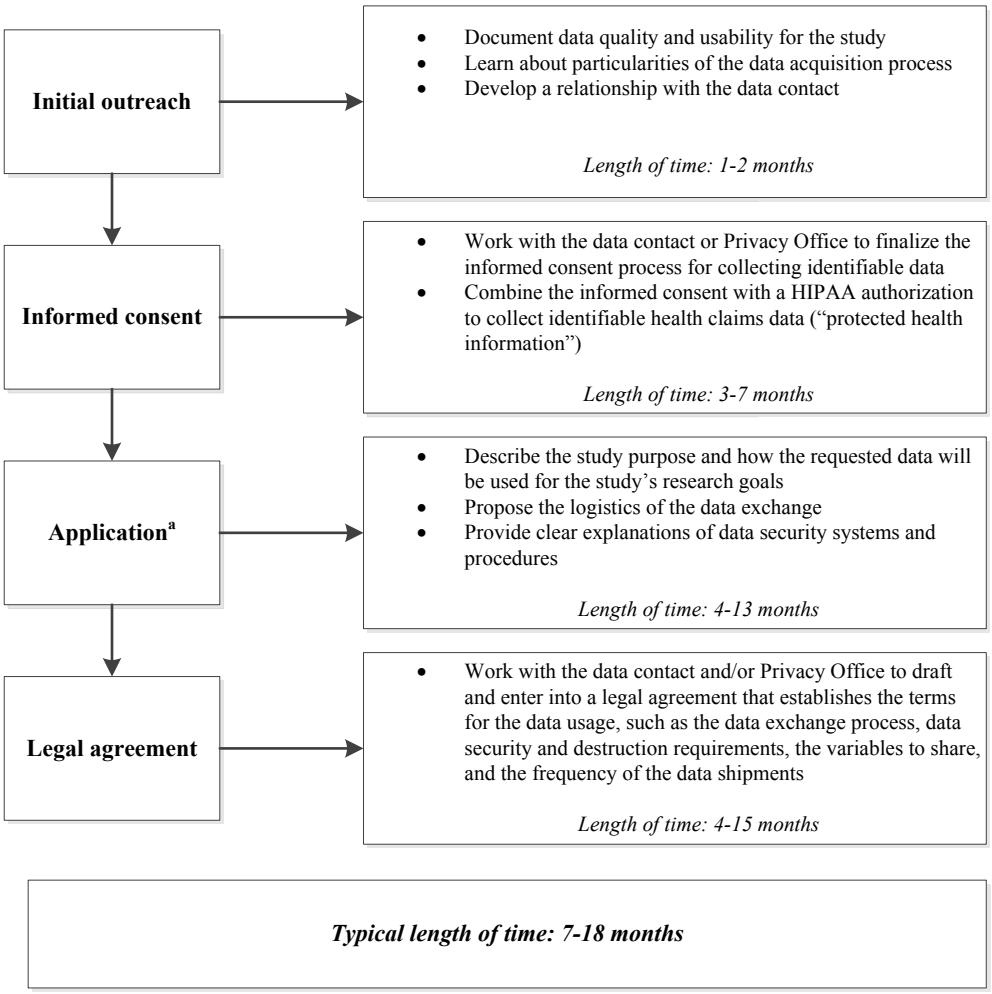
- **Entering into legal agreements often requires legal consultation and negotiations around data usage and destruction as well as around publications review and approval procedures.** Most agencies provide their own legal agreements for sharing data, but these agreements often are not conducive to MIHOPE-Strong Start's design or deliverables and require extensive negotiating to modify the terms.

[2]MIHOPE-Strong Start began in September 2012. Final results from the study, including implementation and impact results for the full MIHOPE-Strong Start sample, will be produced in 2017.

The Mother and Infant Home Visiting Program Evaluation-Strong Start

Figure ES.1

Data Acquisition Process: Typical Length of Time to Complete Each Step



SOURCE: MDRC calculations based on administrative data acquisition efforts with 42 data agencies for birth certificate and Medicaid data as of November 2014.

NOTES: Estimates of the typical length of time (in months) to complete each data acquisition stage are based on the middle 50 percent of targeted MIHOPE-Strong Start states. Data acquisition activities may be completed sequentially or concurrently depending on the data agency's protocol. These estimates are still preliminary since administrative data acquisition has not been completed in all states. HIPAA = Health Insurance Portability and Accountability Act of 1996.

[a]Some agencies may not have an application process and require only an approved consent form and a signed legal agreement. Therefore, the data acquisition process shown in this figure can be reduced to three out of the four stages when a data agency does not require an application to release identifiable data.

- **Because of these multiple steps, the process for reaching an agreement with state agencies has been quite long.** Underscoring how variable the process is, it can take between 4 and 24 months to reach a legal agreement, with half the agencies taking between 7 and 18 months to reach a legal agreement. As of this report's writing, 19 legal agreements have been executed — about half the legal agreements required to access Medicaid and vital records data. Among these first 19 agreements, it has taken 11 months, on average, to complete all the steps necessary to execute them.

- **Not all state agencies are willing to provide Medicaid and birth certificate data.** MIHOPE-Strong Start seeks identified administrative data — records that include personal information on individuals (such as name, date of birth, and residential location) — so that all data received can be linked together for analysis purposes. While most vital records agencies have been able to provide identified records, one state agency will not do so due to state statutes that permit the release of only deidentified data for researchers who are external to the department. Additionally, three Medicaid agencies thus far have indicated that they cannot provide any data to the study. Two of the three agencies noted that they have limited resources to provide data to external researchers. State Medicaid data agencies generally are more accustomed to providing data to a researcher or organization that is conducting services on behalf of the department or state, and thus are more willing to provide data for a research project where the agency can identify a direct programmatic benefit, is sponsoring the research, or is provided with findings from the study about its own state. With Medicaid being a state-administered program, each Medicaid data agency is different.

- **Some administrative data does not lend itself to rapid evaluation, due to the long lag in availability to researchers.** On average, the study team has found that Medicaid data are available more quickly than birth certificate data. However, as noted above, not all Medicaid and vital records agencies were willing to release the data for use in MIHOPE-Strong Start.

- **Many of the MIHOPE-Strong Start study participants are likely to be enrolled in Medicaid managed care organizations, which may make it more difficult to obtain accurate information on health care use.** Data quality may be lower from managed care organizations because they are not reimbursed per service and often use capitated payment plans, so there is less motivation to accurately record the specific services provided than there might be under other arrangements. It may also take longer for such data to

be available since they are not being promptly reported to the state for reimbursement purposes. This may make it difficult to extract information about exactly which services were rendered and their associated costs, particularly in states that have high penetrations of managed care organizations.

- **For most of the data agencies targeted in MIHOPE-Strong Start, birth certificate data will be available for the study to use within two years after the beginning of the calendar year in which the birth occurred.** Birth certificate data typically are processed by data agencies on a calendar-year basis, and are thus not usable in a real-time manner. For example, a birth record from January 2014 may not be available for a data agency to extract and provide to external researchers until late 2015.

Implications

Both lags in data availability and potentially lengthy waiting periods at each data acquisition step are important considerations for research studies such as MIHOPE-Strong Start, especially if the timely analysis of administrative records is of key interest or necessity. Researchers need to learn about each agency's expected lag in data availability when considering how long it will take to complete the proposed research, and program administrators and agencies interested in studies with a quick turnaround should remain cognizant of this potential barrier.

Although the data acquisition efforts in MIHOPE-Strong Start have proven labor intensive and the length of the procedural steps has varied across agencies in ways that could not always be predicted, the process is likely to be smoother if the study involves working with administrative data at only one or a few state agencies, or if the research study is funded or supported by the state that collects the data needed. This underscores a key challenge of MIHOPE-Strong Start, which is that the study is funded by a federal agency and relies on data from up to 20 states. Although many state data agencies are willing to provide data out of goodwill or for the perceived value of the research study, this has not been universal.

Researchers should further anticipate that most state agencies have their own Institutional Review Board (or other review committee), which will need to approve the research study. This process may involve multiple iterations and levels of approval. Research studies involving contact with human subjects and receiving informed consent must adjust the timing of program and participant recruitment according to the time it may take to gain approval from data agencies for these processes. Without confirmed programs for participation, some data agencies contacted by the MIHOPE-Strong Start research team would not review application materials. At the same time, recruitment of individuals in these programs could not occur until the intake procedures for informed consent had been approved by the data agencies and access to data was ensured.

Finally, the navigation of varying state systems that is required for acquiring data from numerous state agencies is likely to require significant staffing capabilities and the expertise to negotiate legal agreements for sharing data, including the guidance of an attorney with extensive knowledge about the research study, state statutes, and privacy protection laws. State agencies are increasingly requiring detailed data security plans, the written assurance and legal agreement that the organization will abide by the agency's own security requirements, copies of signed informed consent forms, and lengthy reviews of the research study and the organization's right to access the agency's identifiable data. These procedural requirements, while increasing a research project's time frame, staff efforts, and study costs, also signify the value placed on safeguarding the personal information of individuals.

Some circumstances may afford easier access to data. Medicaid data agencies are more accustomed to providing data to a researcher or organization that is conducting services on behalf of the department or state. In some cases, national administrative databases may offer an alternative to the difficulties associated with using administrative data from various state agencies. These national databases could not be used for MIHOPE-Strong Start because they typically do not identify individuals (and thus could not be matched to other data used in MIHOPE-Strong Start) or are not available for use by researchers for several years.

In short, while state administrative data can provide reliable information for a large number of families, the process for acquiring those data can be time consuming and labor intensive, reducing their utility for conducting fast studies with fewer resources.

Introduction

Although the majority of infants born today are full term and healthy at delivery, 12 out of every 100 infants are born prematurely and 8 of every 100 have low birth weight.[1] Reducing the number of infants who are born too early or too small has been a longstanding focus of health policy in the United States for several important reasons, including the higher risk of morbidity and mortality for such infants, increased health care spending for infants born in poor health, and the long-term developmental and economic tolls of poor birth outcomes for children and their families.[2] In addition, the risk for and consequences of poor birth outcomes tend to be concentrated among the economically and socially disadvantaged.[3]

Home visiting for low-income, pregnant women — whereby individualized services (including direct education, assessments, and referrals to community resources) are provided to families in their homes — has been identified as one promising avenue for reaching women who are vulnerable to adverse birth and health outcomes. The Mother and Infant Home Visiting Program Evaluation-Strong Start (MIHOPE-Strong Start) is designed to help researchers, practitioners, policymakers, and others understand the effects of home visiting on improving birth outcomes, as well as maternal and infant health outcomes up to age 1, among women enrolled in Medicaid or the Children's Health Insurance Program (CHIP). MIHOPE-Strong Start is a randomized controlled trial being conducted in up to 20 states, thus representing the largest cross-state study of home visiting programs to date. Local home visiting programs included in MIHOPE-Strong Start use one of two home visiting service models that have demonstrated evidence of effectiveness in improving birth outcomes: Healthy Families America (HFA) and Nurse-Family Partnership (NFP).[4] The study is the result of interagency partnerships between the Center for Medicare and Medicaid Innovation (CMMI) of the Centers for Medicare and Medicaid Services (CMS), the Office of Planning, Research and Evaluation (OPRE) in the Administration for Children and Families (ACF), and the Maternal and Child Health Bureau (MCHB) of the Health Resources and Services Administration (HRSA). MIHOPE-Strong Start is being conducted by MDRC in partnership with James Bell Associates, Johns Hopkins University, and Mathematica Policy Research. MIHOPE-Strong Start began in late 2012 and will culminate in a final report to be published in 2017, which will present program implemen-

[1]Martin et al. (2013a).

[2]Bhutta et al. (2002); Cuevas et al. (2005); Hack, Klein, and Taylor (1995); Institute of Medicine (IOM) (2007); Lackritz (2004); March of Dimes (2003); March of Dimes, Partnership for Maternal, Newborn and Child Health, Save the Children, and World Health Organization (2012); Murphy, Xu, and Kochanek (2013); Paneth (1995); Shiono and Behrman (1995).

[3]Lu and Halfon (2003); Shiono and Behrman (1995). In addition to those with low income status, very young women, older women (age 35 or older), African-Americans, women with low education levels, and women who have multiparous births are also at greater risk of poor birth outcomes (IOM 2007).

[4]Avellar and Paulsell (2011).

tation and impact results for the full sample of MIHOPE-Strong Start enrollees. As of the writing of this report, site recruitment is ongoing and will be continued through the first quarter of 2015. The study team has thus far approached numerous sites across 27 states, working closely with the national model developers and state representatives to identify which states had the most programs and to identify those programs that would be strong candidates for participation in the evaluation. The team has also carried out numerous informational discussions with state and local program representatives and stakeholders, as well as with state administrative data officials to inquire about accessibility of Medicaid and vital records.

To measure home visiting programs' impacts on health and health care spending — including birth outcomes, health care interactions, and associated costs — MIHOPE-Strong Start is using state vital records (birth certificate) data and Medicaid files.[5] These data contain accurate information on the clinical and cost outcomes of interest across a large number of individuals, including details on the infant's weight and gestational age at birth (birth certificate data) and health care use and costs (Medicaid files). In addition to administrative records, MIHOPE-Strong Start is collecting 2010 Census data, survey information from local home visiting program staff members and families, HFA and NFP model developer information, and management information systems data. These combined sources will connect information on local community context and program features to dosage and participation levels, to family characteristics, and ultimately, to health outcomes and costs.

This report, the second in a series of four reports about the project, describes the study's efforts to acquire identifiable birth certificate records and Medicaid data from states targeted for MIHOPE-Strong Start in order to accurately assess the key health outcomes of interest. The report is especially timely given recent efforts by the federal Office of Management and Budget to encourage agencies to make administrative data more readily accessible to researchers as a way to lower the costs of conducting research without sacrificing quality, by reducing the need to collect information through surveys or other potentially expensive means.[6] It has also been posited that sharing and encouraging greater use of administrative records across agencies could save time, compared with conducting surveys or observations, and result in "rapid" or quick-turnaround evaluations.[7] Some national-level administrative data are publicly available, but records containing personally identifiable information (such as name, date of birth, residential location, and health information) are not always available or are available to external researchers only after data security procedures and the integrity of the

[5]The state vital records of interest in MIHOPE-Strong Start include birth certificate data and fetal death data. This report focuses on the process of acquiring birth certificate data to measure birth outcomes from vital records agencies.

[6]Card, Chetty, Feldstein, and Saez (2011); Office of Management and Budget (2014).

[7]Office of Management and Budget (2013, 2014).

research objectives are ensured. Because MIHOPE-Strong Start is collecting a broad array of data, it is seeking identifiable administrative data so that all data collected can be linked together into a single analytic data set.

The main finding of this report is that the process of acquiring administrative data across numerous states and multiple agencies within a state in a timely manner carries particular challenges. Specifically, these challenges include identifying an appropriate data contact to help facilitate the process and field questions, receiving approval for the consent procedures to obtain identifiable data, preparing and submitting applications to multiple review committees, negotiating legal agreements for the particular needs of the study design, and working within the bounds of data availability lags.

For example, for most of the data agencies targeted in MIHOPE-Strong Start, birth certificate data will be available for the study to use within 24 months of the beginning of the calendar year in which the birth occurred, indicating that these data cannot be used in "real time." Furthermore, despite the study's securing approval from MDRC's Institutional Review Board (IRB) — which has federalwide assurance — and a Certificate of Confidentiality from the U.S. Department of Health and Human Services (HHS), most state data agencies requested their own review process, which could entail multiple levels of review and result in additional time lags. In large part, the time intensity of the study team's efforts reflects the challenge of research that entails working with more than one state or one agency, as well as the need for and requirements of data stewards to protect the identity and privacy of individuals. Data acquisition efforts are still under way; therefore, this report should be seen as a summary of the acquisition status as of this writing. Yet, in clearly and comprehensively detailing the numerous steps that have been needed to acquire administrative data for MIHOPE-Strong Start thus far, the report can serve as a broader resource as to the benefits and trade-offs of accessing such resources for program or policy evaluation purposes.

Before describing the process that the MIHOPE-Strong Start team is using to acquire administrative data and documenting the varying lags in data availability for the targeted study states, the report presents a brief description of MIHOPE-Strong Start, along with background information on how birth certificate and Medicaid files have typically been used. The report concludes with a discussion of implications for researchers and policymakers drawn from the MIHOPE-Strong Start experience.

Overview of MIHOPE-Strong Start

Research Motivation and Questions

MIHOPE-Strong Start is part of CMMI's Strong Start for Mothers and Newborns initiative, which is testing various approaches to improve outcomes for newborns and pregnant women.[8] As described in the first annual report,[9] MIHOPE-Strong Start is examining the effects of two national home visiting models — HFA and NFP — on birth, infant, and maternal health outcomes and health care costs and will investigate the features of local programs that are associated with differential effects for families.[10]

Both HFA and NFP have previously demonstrated impacts on birth outcomes, as noted earlier. Specifically, a multisite evaluation in New York State found that HFA reduced the number of babies born with low birth weight, while a study in Elmira, New York, found that NFP reduced the number of preterm births among some of the families it served. In addition, NFP trials have shown evidence of improving a number of maternal health outcomes, particularly during the prenatal period, and have reduced the number of emergency department visits among infants. HFA trials have also shown some evidence of impacts on maternal and infant outcomes beyond birth outcomes, such as reduced maternal alcohol use, increases in the number of well-child visits, and improved likelihood of infant health care coverage over time. These studies thus provide some evidence of the potential of home visiting programs to influence the health of mothers during pregnancy as well as the health of their infants. Unfortunately, many of the studies noted above are based on small samples or are focused on one geographic area, and some are now dated.[11] Moreover, these studies did not examine in depth the mechanisms by which programs achieve their impacts, such as through the dosage of services delivered, the screening of health risks, or assistance with referrals to other community services. MIHOPE-Strong Start aims to help answer questions about mechanisms as well as impacts.

The study's broad research questions are thus as follows:

- What is the impact of home visiting programs on birth outcomes, infant and maternal health, and health care use? How do impacts vary for key subgroups?

[8]Centers for Medicare and Medicaid Services Innovation Center (n.d.).
[9]Filene et al. (2013) is available at http://www.mdrc.org/.
[10]Further details of each program model can be found in Filene et al. (2013).
[11]Kitzman et al. (1997); Lee et al. (2009); Olds, Henderson, Chamberlin, and Tatelbaum (1986); Olds et al. (2002).

- How do home visiting programs achieve their results on birth outcomes and maternal and infant health?
- What is the relationship between dosage or enrollment length and program impacts?

CMS may further use the results of the evaluation to write regulations that would allow qualifying home visiting programs to be reimbursed for services through Medicaid.

Study Design

In order to provide the most reliable estimates of the effects of home visiting programs, MIHOPE-Strong Start uses a random assignment design, which involves a lottery-like process that randomly places voluntary study participants into either a home visiting group (program group) that can receive home visiting services from the program sites in the study or to a control group that can receive other services available in the community.[12] Random assignment in MIHOPE-Strong Start occurs after a home visiting program determines that a woman is eligible and interested in the program, but before she is enrolled in the program. The random assignment design ensures that the program and control groups are expected to be similar when they enter the study, so that systematic differences in outcomes — or *impacts* — that are observed between the two groups can be attributed to the home visiting services rather than to some other characteristic or program. Women are considered eligible for MIHOPE-Strong Start if they speak English or Spanish, are at least eight weeks from their due date for delivery, and are at least 15 years old.

To be deemed eligible for MIHOPE-Strong Start, local home visiting programs must have been in operation for at least two years before beginning study procedures, employ at least three full-time home visitors, have an excess demand for services (to allow for the ethical creation of a control group), and serve a prenatal client population where nearly all are expected to be covered by Medicaid or CHIP by the time of the infant's birth. As of this writing, the MIHOPE-Strong Start team is actively seeking local home visiting programs to participate in the study, and this process will be ongoing through the first quarter of 2015.

The organizations conducting MIHOPE-Strong Start are also conducting a companion study, the Mother and Infant Home Visiting Program Evaluation (MIHOPE). MIHOPE is the ongoing evaluation of the Maternal, Infant, and Early Childhood Home Visiting Program

[12]These other services exclude the evidence-based program that is participating in the study but do not exclude receipt of services from other home visiting programs in the community.

(MIECHV or the Home Visiting Program).[13] The MIECHV program, created through an amendment of Title V of the Social Security Act, is a federal funding source for some of the local home visiting programs in MIHOPE-Strong Start. Twelve states — California, Georgia, Illinois, Iowa, Kansas, Michigan, Nevada, New Jersey, Pennsylvania, South Carolina, Washington, and Wisconsin — are participating in MIHOPE. Information for MIHOPE families engaged with either HFA or NFP programs who meet the MIHOPE-Strong Start criteria will be used in the MIHOPE-Strong Start analysis. To identify additional states and sites to include in MIHOPE-Strong Start beyond the 12 MIHOPE states, the study team gathered information and recommendations from the national program model offices and state-level administrators of the MIECHV program (who are familiar with the home visiting landscape in their states). Because the study relies on administrative records as part of the evaluation design, the MIHOPE-Strong Start team also investigated the accessibility of these data sources in identifying other potential states.

In particular, MIHOPE-Strong Start is collecting survey data from program participants when they enroll in the study and is acquiring birth certificate records and maternal and infant Medicaid files to assess program impacts. The study is also collecting information on program implementation from home visiting program staff members. Because it is costly to conduct surveys to measure all the outcomes of interest, and not all outcomes (particularly health care costs) will be known by participants or accurately reported in surveys, outcome measures will rely on administrative records. Outcome measures originating from birth certificate data are likely to include infant weight at birth and gestational age and potential indicators of maternal health and behaviors. Outcomes derived from Medicaid files will likely center on infant health care use — such as well-child visits, hospitalizations, and emergency department visits — and associated costs of health events. It is important to underscore that while administrative records will be used to measure outcomes, the collection of survey data from families is an integral part of the study design as well. Several important measures of participant characteristics at the time of enrollment, such as relationship and cohabiting status, psychosocial well-being, and unmet medical needs, can be assessed only through directly surveying respondents. This information is critical in order to understand the role of participant characteristics in program impacts.

[13]In addition to maternal and infant health outcomes, MIHOPE is examining the effectiveness of home visiting on improving outcomes in a range of other domains, including child development, economic self-sufficiency, crime or domestic violence, parenting, child maltreatment, and referrals and coordination of services (Michalopoulos et al. 2013).

Using Administrative Data in Research

Background on the Collection and Use of Administrative Data

Administrative records traditionally have been used for the management of programmatic and public services, but use has evolved beyond program management as a result of momentum at the national level to improve the accessibility and quality of administrative data.[14] The advent of electronic data in the 1980s made certain types of administrative data, such as unemployment and wage records, more amenable to research purposes by removing the interim step of data entry by researchers. Early electronic data sources came with challenges, such as storage in decentralized locations, which made the data inconsistent across locations and difficult to extract. New policies and programs during the 1990s in such areas as social welfare and public health, which required heightened monitoring and reporting by state and federal governments, introduced incentives for modernizing electronic administrative databases and increasing the uniformity of those data for reporting purposes. Demonstrating metrics became tied to funding requirements and, as a result, data were more consistently recorded. The creation of national databases emerged, and state and local data that funneled into these larger databases improved. While national databases can have limitations on use and complicated acquisition processes, their growth contributed to advancements in access to and quality of administrative data.

For example, the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRWORA) enacted the National Directory of New Hires (NDNH). This directory centralized data on new hires, quarterly wages, and unemployment from all states, with the goal of enforcing child support payments by noncustodial parents who live in different states from their children by withholding the amounts from their paychecks.[15] PRWORA also permitted researchers to use the NDNH by giving them access to individual-level, deidentified (not personally identifiable) data.[16] Thus, since its creation, policymakers and researchers have used the NDNH for understanding a range of policy topics related to labor market functioning and social welfare, and, over time, efforts grew to improve the data for more varied forms of programmatic and research activities.[17] However, permission to use the NDNH is granted on a limited basis, and access to the data comes with a number of restrictions that include working in a "clean" data room (or a workspace disconnected from any networks and peer services) and

[14]Office of Management and Budget (2014).
[15]Solomon-Fears (2014).
[16]42 U.S.C. § 653 (j)(5).
[17]Wheaton, Durham, and Loprest (2012); Office of Child Support and Enforcement (2012a).

having the Office of Child Support Enhancement at HHS process the data for analysis to minimize the risk of identification.[18]

Similarly, the birth certificates and Medicaid data that will be used in MIHOPE-Strong Start have seen improvements in accessibility and usability as part of efforts to advance programmatic functioning. Birth certificates are one of the oldest administrative data sources used in research; births were registered by the colonial settlers in the seventeenth century.[19] In 1902, the U.S. Bureau of the Census started collecting annual vital records from all states, and the establishment of a national system began.[20] Minimum reporting items for all states' vital records were mandated, and the Model Vital Statistics Act of 1907 established guidelines for states to update their systems, law, and regulations.[21] These guidelines have been revised over the years to meet the growing uses of vital records data.[22] Established in 1912, the Child Health Bureau later expanded into the Maternal and Child Health Bureau and influenced the establishment of the birth registry, becoming an important driver of the use of vital records data for public health purposes.[23]

Today, states annually report vital records data to the National Center for Health Statistics (NCHS) in the Centers for Disease Control and Prevention (CDC) using a standard certificate. The NCHS data play a central role in government planning, public health surveillance, and social welfare policy, as the richness of information contained in the U.S. Standard Certificate has applicability to a variety of policy and research areas. For example, with the recognition that some chronic diseases diagnosed in adulthood can begin at the fetal stage, some maternal and child health policymakers now use a life course approach in which programs target maternal prenatal health to improve children's prenatal development with the end goal of preventing serious disease later in life (for example, heart disease, stroke, and diabetes). Vital records are one of the sources used to measure effects.[24]

The current version of the birth certificate (2003 revision) includes detailed information about the mother's and infant's health and health care use during and after delivery (Figure 1), as well as personal information such as mother and infant names, dates of birth, and address. The data are collected from official reports that hospitals prepare to record births.

---

[18]Office of Child Support and Enforcement (2012b).

[19]Hetzel (1997); Hotz, Goerge, Balzekas, and Margolin (1996).

[20]The U.S. Census Bureau first used the national registration system to collect death certificates, and then, by 1915, expanded the national system to collect birth certificates. Today, vital records encompass births, deaths, fetal deaths, marriages, divorces, and adoptions (Hetzel 1997).

[21]Hetzel (1997).

[22]Centers for Disease Control and Prevention (1994).

[23]Harwood, Yu, and Kavanagh (n.d.).

[24]Fine, Kotelchuck, Adess, and Pies (2009); Kogan, Barfield, and Kroelinger (2014).

The Mother and Infant Home Visiting Program Evaluation-Strong Start

Figure 1

Section of U.S. Standard Certificate of Live Birth, 2003 Revision

MOTHER

29a. DATE OF FIRST PRENATAL CARE VISIT
_____/_____/_____ ☐ No Prenatal Care
M M D D YYYY

29b. DATE OF LAST PRENATAL CARE VISIT
_____/_____/_____
M M D D YYYY

30. TOTAL NUMBER OF PRENATAL VISITS FOR THIS PREGNANCY
_____ (If none, enter ".0".)

31. MOTHER'S HEIGHT
_____ (feet/inches)

32. MOTHER'S PREPREGNANCY WEIGHT
_____ (pounds)

33. MOTHER'S WEIGHT AT DELIVERY
_____ (pounds)

34. DID MOTHER GET WIC FOOD FOR HERSELF DURING THIS PREGNANCY? ☐ Yes ☐ No

35. NUMBER OF PREVIOUS LIVE BIRTHS (Do not include this child)

35a. Now Living
Number _____
☐ None

35b. Now Dead
Number _____
☐ None

36. NUMBER OF OTHER PREGNANCY OUTCOMES (spontaneous or induced losses or ectopic pregnancies)

36a. Other Outcomes
Number _____
☐ None

37. CIGARETTE SMOKING BEFORE AND DURING PREGNANCY
For each time period, enter either the number of cigarettes or the number of packs of cigarettes smoked. IF NONE, ENTER ".0".

Average number of cigarettes or packs of cigarettes smoked per day.

	# of cigarettes		# of packs
Three Months Before Pregnancy	_____	OR	_____
First Three Months of Pregnancy	_____	OR	_____
Second Three Months of Pregnancy	_____	OR	_____
Third Trimester of Pregnancy	_____	OR	_____

38. PRINCIPAL SOURCE OF PAYMENT FOR THIS DELIVERY
☐ Private Insurance
☐ Medicaid
☐ Self-pay
☐ Other (Specify) _____

35c. DATE OF LAST LIVE BIRTH
_____/_____
MM YYYY

36b. DATE OF LAST OTHER PREGNANCY OUTCOME
_____/_____
MM YYYY

39. DATE LAST NORMAL MENSES BEGAN
_____/_____/_____
M M D D YYYY

40. MOTHER'S MEDICAL RECORD NUMBER

MEDICAL AND HEALTH INFORMATION

41. RISK FACTORS IN THIS PREGNANCY (Check all that apply)

Diabetes
☐ Prepregnancy (Diagnosis prior to this pregnancy)
☐ Gestational (Diagnosis in this pregnancy)

Hypertension
☐ Prepregnancy (Chronic)
☐ Gestational (PIH, preeclampsia)
☐ Eclampsia

☐ Previous preterm birth

☐ Other previous poor pregnancy outcome (Includes perinatal death, small-for-gestational age/intrauterine growth restricted birth)

☐ Pregnancy resulted from infertility treatment-If yes, check all that apply:
☐ Fertility-enhancing drugs, Artificial insemination or Intrauterine insemination
☐ Assisted reproductive technology (e.g., in vitro fertilization (IVF), gamete intrafallopian transfer (GIFT))

☐ Mother had a previous cesarean delivery
If yes, how many _____

☐ None of the above

42. INFECTIONS PRESENT AND/OR TREATED DURING THIS PREGNANCY (Check all that apply)
☐ Gonorrhea
☐ Syphilis
☐ Chlamydia
☐ Hepatitis B
☐ Hepatitis C
☐ None of the above

43. OBSTETRIC PROCEDURES (Check all that apply)
☐ Cervical cerclage
☐ Tocolysis

External cephalic version:
☐ Successful
☐ Failed

☐ None of the above

44. ONSET OF LABOR (Check all that apply)
☐ Premature Rupture of the Membranes (prolonged, ∃12 hrs.)
☐ Precipitous Labor (<3 hrs.)
☐ Prolonged Labor (∃ 20 hrs.)
☐ None of the above

45. CHARACTERISTICS OF LABOR AND DELIVERY (Check all that apply)
☐ Induction of labor
☐ Augmentation of labor
☐ Non-vertex presentation
☐ Steroids (glucocorticoids) for fetal lung maturation received by the mother prior to delivery
☐ Antibiotics received by the mother during labor
☐ Clinical chorioamnionitis diagnosed during labor or maternal temperature ≥38°C (100.4°F)
☐ Moderate/heavy meconium staining of the amniotic fluid
☐ Fetal intolerance of labor such that one or more of the following actions was taken: in-utero resuscitative measures, further fetal assessment, or operative delivery
☐ Epidural or spinal anesthesia during labor
☐ None of the above

46. METHOD OF DELIVERY

A. Was delivery with forceps attempted but unsuccessful?
☐ Yes ☐ No

B. Was delivery with vacuum extraction attempted but unsuccessful?
☐ Yes ☐ No

C. Fetal presentation at birth
☐ Cephalic
☐ Breech
☐ Other

D. Final route and method of delivery (Check one)
☐ Vaginal/Spontaneous
☐ Vaginal/Forceps
☐ Vaginal/Vacuum
☐ Cesarean
If cesarean, was a trial of labor attempted?
☐ Yes
☐ No

47. MATERNAL MORBIDITY (Check all that apply) (Complications associated with labor and delivery)
☐ Maternal transfusion
☐ Third or fourth degree perineal laceration
☐ Ruptured uterus
☐ Unplanned hysterectomy
☐ Admission to intensive care unit
☐ Unplanned operating room procedure following delivery
☐ None of the above

NEWBORN INFORMATION

NEWBORN

48. NEWBORN MEDICAL RECORD NUMBER

49. BIRTHWEIGHT (grams preferred, specify unit)
_____
9 grams 9 lb/oz

50. OBSTETRIC ESTIMATE OF GESTATION:
_____ (completed weeks)

51. APGAR SCORE:
Score at 5 minutes:_____
If 5 minute score is less than 6,
Score at 10 minutes: _____

52. PLURALITY - Single, Twin, Triplet, etc.
(Specify)_____

53. IF NOT SINGLE BIRTH - Born First, Second, Third, etc. (Specify) _____

54. ABNORMAL CONDITIONS OF THE NEWBORN (Check all that apply)
☐ Assisted ventilation required immediately following delivery
☐ Assisted ventilation required for more than six hours
☐ NICU admission
☐ Newborn given surfactant replacement therapy
☐ Antibiotics received by the newborn for suspected neonatal sepsis
☐ Seizure or serious neurologic dysfunction
☐ Significant birth injury (skeletal fracture(s), peripheral nerve injury, and/or soft tissue/solid organ hemorrhage which requires intervention)
9 None of the above

55. CONGENITAL ANOMALIES OF THE NEWBORN (Check all that apply)
☐ Anencephaly
☐ Meningomyelocele/Spina bifida
☐ Cyanotic congenital heart disease
☐ Congenital diaphragmatic hernia
☐ Omphalocele
☐ Gastroschisis
☐ Limb reduction defect (excluding congenital amputation and dwarfing syndromes)
☐ Cleft Lip with or without Cleft Palate
☐ Cleft Palate alone
☐ Down Syndrome
☐ Karyotype confirmed
☐ Karyotype pending
☐ Suspected chromosomal disorder
☐ Karyotype confirmed
☐ Karyotype pending
☐ Hypospadias
☐ None of the anomalies listed above

56. WAS INFANT TRANSFERRED WITHIN 24 HOURS OF DELIVERY? 9 Yes 9 No
IF YES, NAME OF FACILITY INFANT TRANSFERRED
TO:_____

57. IS INFANT LIVING AT TIME OF REPORT?
☐ Yes ☐ No ☐ Infant transferred, status unknown

58. IS THE INFANT BEING BREASTFED AT DISCHARGE?
☐ Yes ☐ No

Mother's Name _____ Mother's Medical Record No. _____

Hospitals submit these reports to states through an electronic registration system. States are federally required to clean and process the data and to send completed vital records files to the NCHS for disseminating official vital statistics.[25]

While data housed by the NCHS contain much of the information displayed in Figure 1, the NCHS data are deidentified.[26] As a consequence, these data cannot be combined with other administrative or survey data sets to answer such policy research questions as whether improvements in infant health at birth (birth certificate data) can reduce postbirth hospitalizations and health care spending in the first year of infancy (medical records and cost data). The MIHOPE-Strong Start team therefore had to acquire identified vital records data from each state agency.

Medicaid data have a shorter history. The Medicaid program was established by the Social Security Act of 1965 and is jointly funded by state and federal governments. Historically, states have had a lot of flexibility in the administration of Medicaid, which has resulted in varying data practices. The Balanced Budget Act of 1997 initiated important changes in its administration.[27] Most relevantly, the act mandated a uniform process for states to submit Medicaid claims and eligibility data to CMS and conditioned federal matching dollars upon accurate submission of data.[28] The system used for collecting these data, called the Medicaid Statistical Information System (MSIS), has improved data uniformity across states, although eligibility requirements and Medicaid-covered services remain distinct by state.[29] Medicaid data use has experienced a marked shift in recent years, from being primarily administrative to having a research objective, through the release of the Medicaid Analytic eXtract (MAX),[30] which converts data in the MSIS from quarterly files into annual files designed to be more usable by the research community. While MAX data represent a major advancement for using Medicaid data for research and are available in an identifiable form, they take at least two years

[25]42 U.S.C. § 242k.

[26]Aggregated data from the NCHS are publicly available online. Researchers interested in deidentified record-level data must submit an application to NCHS and agree to the terms of a data use agreement.

[27]Klemm (2000).

[28]42 U.S.C. 1396b § 1903(r)(1)(F).

[29]Kaiser Commission on Medicaid and the Uninsured (2004). The Balanced Budget Act also simplified the process by which Medicaid beneficiaries could enroll in managed care health plans, and it paved the path for state Medicaid programs to increase their use of managed care. The act did not require encounter data (data received from Medicaid managed care organizations) to abide by the same MSIS submission standards as claims data (data received from Medicaid fee-for-service providers), given the complexity of managed care billing practices. However, the Patient Protection and Affordable Care Act recently tied the receipt of federal matching dollars to the accurate submission of encounter data to the MSIS, and it is expected that encounter data will become more comparable to fee-for-service claims data (Byrd and Verdier 2011).

[30]Previously called State Medicaid Research Files (SMRFs).

to be compiled and processed, which is too long for MIHOPE-Strong Start. Therefore, the research team has had to collect Medicaid data directly from state agencies.[31]

Data Quality Considerations

Although the use of administrative records has particular strengths compared with the more resource-intensive effort to collect survey data, and administrative records are prized for their accuracy, they are not without some data quality concerns for research purposes. While many of the administrative data-based measures of interest in MIHOPE-Strong Start have been found to be highly accurate, other indicators may be less so.

Birth Certificate Data

Data elements that are commonly used to monitor fertility, birth, and mortality trends have been found to have high accuracy. Specifically, studies that validated birth certificate data using medical chart abstractions have found high rates of agreement for birth weight, Apgar score,[32] obstetric estimate of gestational age, and methods of delivery. For example, when birth certificate records were compared with data abstracted from medical charts, the percentage of infants classified as low birth weight and very low birth weight on both sources had over 95 percent agreement.[33]

In contrast, information from birth certificates about pregnancy history and prenatal care is less valid because this information is typically based on mothers' self-reports (which can be subject to memory or other types of bias) and entered onto the forms by hospital staff members (adding an additional opportunity for misreporting). Studies indicate less agreement for these items when compared with medical health records. The latest (2003) revision to the U.S. Standard Certificate improved data quality and reporting rates, but implementation has been delayed in numerous states, and there is a continued lack of standardization across hospitals.

---

[31]Newer BetaMAX files release data more quickly but still have a lag of over one year. A third version of MAX files, called AlphaMAX, is in development, with the advantage that these files will be prepared for release as they are received, instead of after seven quarters of data are received, which is generally how the MAX and BetaMAX files are prepared.

[32]The Apgar score comes from a test performed on a baby at one minute and five minutes after birth, measuring breathing effort, heart rate, muscle tone, reflexes, and skin color. The one-minute score determines the baby's response to the birthing process. The five-minute score indicates how the infant is faring in the environment.

[33]Martin et al. (2013b).

Medicaid Data

Medicaid claims data record health care services provided to beneficiaries, associated diagnoses that justify the services, amounts charged by providers, and amounts Medicaid reimbursed to the provider. Each state, however, has its own federal medical assistance percentage, which is the specified percentage of program expenditures that the federal government will pay to states. This can make using data across different states more difficult for research, because reimbursement prices will vary for the same service or prescription medication. A movement toward payment arrangements in which packages of related services are collapsed into a single code (a practice sometimes referred to as bundled payments or global billing) can further affect the accuracy of the data because it is difficult to isolate specific events.

Other data quality concerns can be attributed to patterns in the stability of coverage; Medicaid files capture services only during the time someone was receiving Medicaid benefits, so data completeness can be of concern for individuals who lose coverage or disenroll for a period of time.

The form and quality of Medicaid data depend on whether people are in a fee-for-service (FFS) plan or a Medicaid managed care organization (MCO).[34] CMS has developed strict guidelines for submitting FFS claims for payment that give providers an incentive to enter information accurately about services rendered. Managed care organizations, however, have different payment arrangements from those of fee-for-service providers. Managed care organizations may get reimbursed on a monthly capitation basis (that is, they receive a set fee per patient regardless of the treatment or service provided) or may have network-specific guidelines. If capitated payments are used, the state does not need to see or track each service, so there is less motivation to record these accurately than there might be under fee-for-service arrangements. This makes it difficult to extract information about exactly which services were rendered and their associated costs, especially in states that have a high penetration of managed care organizations. This also means that the data extracted may be less accurate.

Most MIHOPE-Strong Start states use managed care for a majority of Medicaid beneficiaries, and the quality of certain types of managed care encounter data is likely to vary across states. (Note that Medicaid reports received from managed care organizations are called encounters rather than claims.) In some MIHOPE-Strong Start states, managed care penetration is likely to be high; several states (as of 2010) reported that over 70 percent of their Medicaid

[34]FFS is a payment model in which providers are paid for the specific services they provide. By contrast, MCOs implement a range of activities to try to reduce health care costs while maintaining health care quality.

beneficiaries were in managed care organizations.[35] Other MIHOPE-Strong Start states have previously reported very low levels of managed care penetration, but the use of managed care has grown over time, and pregnant women are more likely to be in managed care compared with other Medicaid recipients. The extent to which study participants are enrolled in managed care organizations, and how this varies across states, will remain unclear until all records are received.

Similarly, the extent to which the required encounter data may or may not be usable in the records maintained by state Medicaid agencies remains unknown. Prior research on the usability of managed care data for research purposes has revealed that most of the states approached for MIHOPE-Strong Start are known to have usable managed care data for the primary infant health outcomes described earlier.[36] However, with respect to both outpatient and inpatient care data, Medicaid encounter data may not be usable in several states.[37]

Thus, although several of the salient potential outcomes in MIHOPE-Strong Start have been shown to be captured accurately in administrative data sources, the quality of information for other indicators may be suboptimal in some states or for some data elements, and will need to be closely examined when the data are received. And while recent efforts to make birth certificate and Medicaid data available nationally have furthered some research agendas and improved the quality of records submitted to federal agencies,[38] these national databases will not work for studies like MIHOPE-Strong Start that need individual-level, identifiable data as soon as possible after a birth event occurs or a health care service is rendered and paid. The following section describes MIHOPE-Strong Start's experience securing birth certificate and Medicaid data directly from the states targeted for the study.

Description of the Process Used to Collect Data

As of the writing of this report, MIHOPE-Strong Start had initiated data acquisition efforts with 21 Medicaid agencies and 21 vital records agencies in 20 states.[39]

---

[35]Centers for Medicare and Medicaid Services (2010).

[36]In assessing whether data is usable for research purposes, Byrd and Dodd (2012) consider completeness of encounter records (assessed by volume, such as the number of claims per enrollee and the percentage of enrollees with claims) and the quality or amount of information provided about the encounter (for example, whether diagnosis and procedure codes were filled in and with specificity).

[37]Byrd and Dodd (2012).

[38]National Association for Public Health Statistics and Information Systems (2013).

[39]The total number of targeted Medicaid agencies includes one state that has two relevant agencies: one for collecting fee-for-service data and one for collecting managed care data. The latter is the state's "all-payer claims database," a repository of claims from all health insurance payers in the state, including programs such as Medicaid and private insurers such as Blue Cross Blue Shield, in which the data are standardized across all

(continued)

Figure 2 shows the typical data acquisition process, which starts with outreach to an appropriate data contact and ends with a legal agreement that ensures the study team's right to receive data from the agency. (The process of having the state send the administrative data to MDRC is not discussed.) Key intermediary steps are having the data agency review and approve the MIHOPE-Strong Start informed consent form and, in most cases, submitting an application for data use.

The ranges of months given in Figure 2 are estimates of the time needed to complete each data acquisition step for the middle 50 percent of agencies targeted for MIHOPE-Strong Start based on data the study team has collected to track progress. They include projections, because not every step has been completed for all the agencies. The longest step is entering into a legal agreement, which can take between 4 and 15 months. The different data acquisition stages may overlap (for example, the application step may be combined with the review of the consent form or with the legal agreement step) or may occur sequentially, which tends to lengthen the process. The overall amount of time to complete all these steps varies greatly; it has taken between 7 and 18 months to complete the process for the middle 50 percent of all state agencies. The full range of time for all agencies contacted by MIHOPE-Strong Start suggests that the process can extend from just 4 months to as long as 24 months (not shown).

Some agencies are outliers — taking a much shorter or longer time than expected — making it difficult to predict exactly how long the administrative data acquisition process will take. At the time of this report's writing, the team had entered into legal agreements with about half (19) of the data agencies. Among those 19 agencies, it took 11 months, on average, from the time the team started to identify a contact at the data agency to the date the agreement was executed. It should be noted that outreach to the 42 agencies did not occur all at once.

The remainder of this section provides greater detail about each of the steps taken in this process.

Initial Outreach

The first step in the administrative data acquisition effort is to identify an appropriate contact with whom to speak at a data agency. Once this person is identified, the team reaches out for an initial consultation. Through this conversation, the team learns about the quality of the agency's data and whether they will be usable for MIHOPE-Strong Start's research goals.

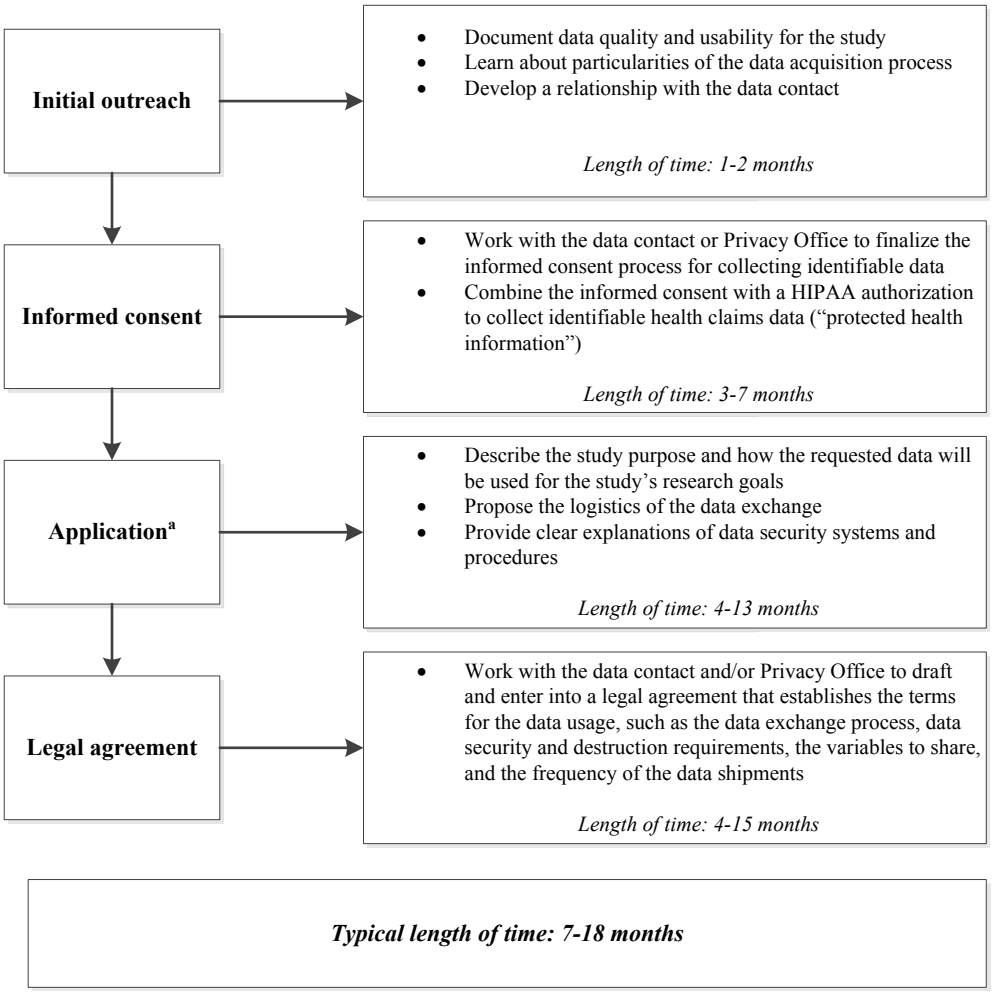
---

payers. The total number of vital records agencies also includes one state that has two separate data agencies housing vital records data for different parts of the state. Therefore, data collection activities targeted 21 Medicaid and 21 vital records agencies in 20 states.

The Mother and Infant Home Visiting Program Evaluation-Strong Start

Figure 2

Data Acquisition Process: Typical Length of Time to Complete Each Step



SOURCE: MDRC calculations based on administrative data acquisition efforts with 42 data agencies for birth certificate and Medicaid data as of November 2014.

NOTES: Estimates of the typical length of time (in months) to complete each data acquisition stage are based on the middle 50 percent of targeted MIHOPE-Strong Start states. Data acquisition activities may be completed sequentially or concurrently depending on the data agency's protocol. These estimates are still preliminary since administrative data acquisition has not been completed in all states. HIPAA = Health Insurance Portability and Accountability Act of 1996.

[a]Some agencies may not have an application process and require only an approved consent form and a signed legal agreement. The data acquisition process shown in this figure can be reduced to three out of the four stages when a data agency does not require an application to release identifiable data.

These initial conversations with agency contacts help the team learn whether an application will have to be completed to access data or to initiate human subjects contact, whether the agency wants to change the study's process for obtaining informed consent, and whether the agency has its own legal documents that MDRC will need to sign. The team also asks detailed questions about data quality, such as when data will be complete and how long it will take for the agency to send the data files to the researchers. Another key piece of information needed in planning study intake is which identifiers the agency needs in order to match to the MIHOPE-Strong Start sample (for example, name, date of birth, Social Security number, or Medicaid identification number) and how the agency will conduct this match. Additionally, for budgeting purposes, the team asks about the costs associated with the requested data extractions. Fees have varied widely — from no cost to $15,000 for two data extractions.

Table 1 presents information learned from the data acquisition process undertaken for MIHOPE-Strong Start. As shown, at the time of writing, initial outreach has been completed in nearly all the targeted data agencies (20 out of 21 Medicaid agencies; 20 out of 21 vital records agencies). In both of the cases where initial outreach has not been completed, it is because research activities in the state have begun only recently. Additionally, as detailed later in the report, in the cases of three Medicaid agencies and one vital records agency, the team learned early in the process that the agency would be unable to provide the data needed for the study.

The case study in Box 1, a compilation of the team's experiences with different agencies, illustrates the importance of direct outreach at the start of the acquisition process by presenting an example of how early conversations can affect later stages.

Informed Consent

For MIHOPE-Strong Start, all study participants provide informed consent. The standard informed consent form used for MIHOPE-Strong Start (hereafter referred to as the "standard consent form") details the activities asked of study participants, explains how the study will use a participant's information, covers anticipated risks and benefits, and provides a means to remove oneself from the study by revoking consent.

All Medicaid data agencies and most vital records agencies have a template consent form, but the study team's preference was to use the MIHOPE-Strong Start standard consent form. Typically, a data agency's legal counsel reviews the study's standard consent form to verify whether it meets the agency's standards for allowing the researchers to obtain identifiable data and use the data as proposed. While both the data agencies and the study team share the explicit goal of conducting an ethical research study and protecting sample members' identities, some data agencies have required changes to the standard consent form. Some agencies are bound by state statutes that require the informed consent process to follow specific guidelines, for example, to disclose that certain types of protected health information (PHI) will be released

The Mother and Infant Home Visiting Program Evaluation-Strong Start

Table 1

Administrative Data Acquisition Summary

Number of Agencies	Medicaid	Vital Records	Total
Targeted[a]	21	21	42
Initial outreach completed	20	20	40
Agency required modifications to standard informed consent form[b]	6	5	11
Agency required Institutional Review Board (or similar review committee) application[c]	14	20	34
Agency provided a template legal agreement[d]	19	19	38
Entered into a legal agreement	10	9	19

SOURCE: MDRC calculations based on administrative data acquisition efforts as of November 2014.

NOTES: [a]Targeted agencies are those where outreach occurred to learn more about the availability of data. The total number of targeted Medicaid agencies includes one state that has two relevant agencies: one for collecting fee-for-service data and one for collecting managed care data. The total number of vital records agencies also includes one state that has two separate data agencies for collecting data from different parts of the state.

[b]An additional five Medicaid and three vital records agencies have not approved the standard informed consent form and, therefore, still may request changes.

[c]In the case of one Medicaid agency and one vital records agency, it is still unknown whether they will require an application.

[d]Legal agreements can be part of an application for data or a stand-alone document. There is one Medicaid agency and one vital records agency for which it is unknown whether they have a template agreement.

to researchers.[40] Other agencies have simply wanted the team to be clearer about exactly which types of data will be collected or how identifiers will be used. The study team has to balance

[40]This process is further complicated when collecting Medicaid data, which is governed by the Health Insurance Portability and Accountability Act of 1996 (HIPAA). All protected health information (PHI), including identifiable Medicaid data, has specific regulations for its acquisition and use. See the Code of Federal Regulations, 45 C.F.R. § 160.103, for the legal definitions of PHI and covered entities. A covered entity cannot disclose the PHI of an individual whose services received or other characteristics are described in the data without receiving a formal authorization from the individual for this disclosure. Therefore, a consent form that seeks to collect PHI must include a set of core elements (see 45 C.F.R. §164.508(c)(1)) and required statements (see 45 C.F.R. § 164.508(c)(2)) that are necessary parts of a HIPAA authorization to release PHI.

Box 1

Collecting Vital Records: Making an Initial Outreach and Submitting an Application

A researcher seeks identifiable birth certificate data for a study. She needs identifiable data for subgroup analyses. To begin the process, she searches on the relevant data agency's website to learn where the data are housed and finds some information there about requesting data for research. She develops a few questions, based on the materials she found and the study's data needs, and compiles those into an introductory memo. She emails a contact listed on the department's website, asking to hold a brief call to discuss the study and to review the memo she has attached.

A few weeks later, an agency staff member responsible for data requests contacts the researcher to talk about the study. The researcher learns that the department's application process for her study will be different from what is listed on the website and that there is a fluid schedule for submitting applications for review. She also gains some pointers for improving the odds of having her application approved. The review committee is particularly interested in learning how the data will be deidentified once acquired by the researcher and wants to see a lot of detail about this process. Also, the data agency reports being understaffed and only takes on projects that further the department's mission. The data contact encourages the researcher to demonstrate how her study will have benefit to the department, a requirement that was not conveyed in the materials or on the website. The researcher makes sure to provide evidence for such benefit in a number of places in the application.

Once she completes the application, the researcher submits it to the agency. Her application includes a detailed data security plan, a set of signatures from all staff members who will have access to the data, a copy of the informed consent form that the study will ask participants to sign, and a series of flow charts that illustrate the data deidentification process and data sharing. She plans to begin her study in three months and has been told that the department reviews applications on an ad hoc basis as their schedule allows. It is important that the review happen before enrollment into the study begins, in case the review committee has any particular requirements about the wording of the informed consent form. The review committee provides preliminary feedback that the justifications for the data elements she requested are not adequate and that the data shipments will need to be sent later than requested. She revises and resubmits. At this point, she has one month until her study is scheduled to begin. She calls the data contact to explain the growing urgency and he helps connect her with the review committee to discuss the remaining issues. The review committee is still concerned about how the data will be sufficiently deidentified. They are particularly concerned about the possibility of the researcher using follow-back procedures, meaning that she would use the identifiable data to contact families for additional research. The researcher clarifies during this conversation that she does not intend to use follow-back procedures based on the data that she will receive from the data agency. The review committee has this clarified in the application and then sends it to the State Registrar to process the official approval.

these requests from data agencies with MIHOPE-Strong Start's intent to make the informed consent process as easy and understandable to study participants as possible.

The study team needs the data agencies to approve the informed consent process (regardless of whether the team had to make changes) before the team starts recruiting sample members into the study in that state, to avoid the possibility of enrolling participants for whom data will not be provided. The study team cannot initiate enrollment of study participants until the standard consent form has been approved.

At the time of writing, 5 Medicaid and 3 vital records agencies have not yet indicated whether the informed consent form is sufficient, while approximately one-third (11 out of 34) of the remaining agencies required changes to the study's standard consent form (Table 1). Many of the Medicaid agencies (4 out of 6) requested modifications to the standard consent form in order to specify that certain types of sensitive health information, such as the treatment and status of HIV and AIDS, mental health conditions, and substance abuse, would be released.

Application

Most data agencies (34 out of 41; unknown in the case of 1) have required the MIHOPE-Strong Start team to submit an application to either an IRB, a departmental review committee, or the agency itself (Table 1). Nearly all vital records agencies have required some level of formal review (20 out of 21), compared with two-thirds of the Medicaid agencies (14 out of 20; unknown for 1). Typically, applications ask the data requester to describe the purpose of the study, how the requested data will be used to complete the study's research goals, the types of deliverables that will result from the study, the data security procedures that will be followed, and details about the data request (such as the list of data elements, the number of data shipments, the time periods of data coverage for each shipment, and the preferred method of data transmission).

Important to all data agencies is evidence that a requester is able to protect their data. Applications have to provide detailed accounts of the involved organizations' security systems, including how data will be encrypted in transit and at rest to ensure the confidentiality of all sample members and how access to the identifiable administrative data will be limited to staff members on a "need-to-know" basis.

In some cases, the applications involve a review of human subjects contact, which means that the research design and contact procedures for the study are under consideration and research activities cannot be started until the team receives approval. Applications of this "full-project" scope may be for an IRB or for a data agency application and include a review of all study protocols involving the subjects, such as the study design, intake procedures, survey instruments, and whether the interactions with subjects abide by ethical standards.

A few states even have multiple review committees for access to a single data source. This occurs in some cases because a state's data are stored within different organizations. For example, one state requires one application for fee-for-service Medicaid claims and another for managed care encounter data. Another state has two separate agencies that house vital records data for different parts of the state. Multiple applications and reviews also occur when a state requires applications to each data agency but also has an overarching, statewide review committee with jurisdiction over the individual data agencies. Without approval from both the data agencies and the statewide IRB, data access will not be granted.

Finally, the review process has varied greatly in terms of the number of follow-up questions that an agency asks, whether a resubmittal is required, and the length of time it takes the reviewers to approve an application. A typical review process, which includes resubmitting revised materials if requested and receiving approval from the committee, can take as little as one week or as long as three to six months. While it requires a significant amount of research staff time to handle these applications, having gone through a systematic application procedure seems to speed the process once the committee completes its review and a legal agreement is being negotiated. In cases where an application is submitted and accepted, reviewers seem less likely to ask for information or changes in an ad hoc manner in the course of negotiating the language for the legal agreement.

Legal Agreement

Most agencies (38 out of 40; unknown for 2) have template legal agreements, which serve as the starting point for negotiating a final agreement stating that the agency will provide MDRC with data to be used for the MIHOPE-Strong Start study (Table 1). These template agreements range from data use, business associate, and nondisclosure agreements to a series of confidentiality agreements to be signed by each study staff member with access to data.[41] They vary widely; one agency provided an agreement that was nearly 60 pages long, while another provided a one-page data use document. The legal agreement may also be combined with the application, rolling the two steps into one, which makes the content of the application part of the legal obligations for the requester and the data agency.

The template legal agreements from data agencies often have required much negotiation because they do not meet the requirements of a complex study such as MIHOPE-Strong

[41]Data use agreements are typically used to broker the data exchange between a data agency and a requester. Nondisclosure and confidentiality agreements tend to be paired with a data use agreement and more explicitly obligate the research organization and its staff to protect the data by having, for example, all staff members involved sign the agreement. Business associate agreements are provided by Medicaid agencies, as they are contracts between a HIPAA-covered entity and a business associate (who provides a service to the covered entity); they are intended to protect PHI in accordance with HIPAA guidelines.

Start. Depending on the nature and number of requested modifications, the negotiations between MDRC and the data agency may span multiple iterations and can take from a couple of months to over a year to execute. MDRC has had to ensure that it has the staff to navigate these negotiations, including an attorney with in-depth knowledge about data security, various state statutes, and the Health Insurance Portability and Accountability Act of 1996 (HIPAA).

The case study in Box 2 presents an example of how the study team might interact with a data agency's privacy office to modify a consent form and legal agreement. It is based on the study team's experiences with multiple agencies rather than one particular agency.

Key areas of the agreements that have consistently met with resistance from data agencies but that are necessary for MIHOPE-Strong Start are described below.

Destruction of Identifiable Data and Retention of Deidentified Data

One goal of MIHOPE-Strong Start is to further the knowledge base of social policy researchers by archiving a deidentified analysis file for secondary use by interested researchers (called a "restricted access file"). This restricted access file will combine all data sources collected for MIHOPE-Strong Start but will strip the file of personal identifiers and mask "small cells" so that sample members' identities cannot be determined by users of the file. This will be done by collapsing small cells in the table of data (for example, by combining specific racial and ethnic groups that are uncommon among the study's sample members into a broad "other" category) and by conducting cross-tabulations of information to ensure that identities cannot be determined by combining information. For example, if there were only two 15-year-old study participants who were married, then the marital status of those individuals would be removed so that a small cell size of two would not appear in the data set. Additionally, users of the restricted access file will have to meet and follow specific requirements, including data security requirements for storing and accessing the data, using the data for research purposes only, and being associated with a verifiable research institution, in order to access the data.

OPRE, a sponsor of the evaluation, requires most of its large studies to archive their data so that other researchers can explore new research questions and replicate original findings once the primary study is complete. Creating a restricted access file lends transparency to MIHOPE-Strong Start by allowing other researchers to do independent analyses and to determine whether they agree with the conclusions of the original study. It also ensures that the time, effort, and resources spent to collect the data for MIHOPE-Strong Start will have use beyond this specific project and that this information can help inform other research as well as program and policy agendas in the future.

Box 2

Collecting Medicaid Data: Modifying an Informed Consent Form and Completing a Legal Agreement

A researcher is working with the privacy office of a state health department to revise an informed consent form that includes a HIPAA authorization. He wants to obtain identifiable Medicaid data, which is protected health information. He learned that the proposed consent form is not adequate for collecting identifiable Medicaid data because it does not clearly explain that eligibility records will be used in addition to claims records. It also does not mention that the data will be linked with other types of data using Social Security numbers, which the department requires researchers to note. The privacy office also would like certain legal language that is part of their template HIPAA authorization to be included in the informed consent form. HIPAA requires that authorizations are written in plain language, but the template from the privacy office contains a lot of legalese, so the researcher is trying to strike a balance between the language the department has proposed and what he believes study participants will be able to understand. An attorney from the researcher's organization helps him to negotiate the changes to the informed consent form so that it meets both parties' needs. They finally agree to supplement the researcher's informed consent form with the department's HIPAA authorization, while stripping some of the duplicative information from the HIPAA authorization.

After this, the researcher starts to work with the privacy office to finalize a legal agreement. The researcher receives a template data use agreement to be signed by the organization's responsible party. It incorporates aspects of the study that the researcher talked about with the privacy office, such as the purpose of the study and the details of the data shipments. The researcher's attorney reviews the data use agreement with the researcher and notes that it specifies that the data must be encrypted in transit and at rest if they are not deidentified. The attorney also notes a provision in the agreement that would curtail the researcher's ability to use the data as needed.

The researcher had intended to encrypt only data that have personally identifiable information, such as names and Social Security numbers, but not protected health information, such as the calendar month of admission to a hospital. The researcher works with his information technology department to make sure his study is in compliance with the department's data security requirements. At the same time, the researcher's attorney negotiates with the privacy office at the data agency to have the provision modified in the agreement. Since those at the privacy office are also attorneys, preferred protocol is to have peer-to-peer communications. The negotiations take over three months, but the parties eventually arrive at a mutually agreeable solution and sign the modified agreement.

Some agencies initially stated a preference for the destruction of deidentified (rather than just identifiable) data at the end of the study and therefore were concerned about having it included in the MIHOPE-Strong Start restricted access file, since their data (in deidentified form) would continue to be used beyond the termination date of the legal agreement. However, providing agencies with a detailed plan for how the data in the restricted access file would be created, used, and stored has generally satisfied their concerns. Both vital records and Medicaid data agencies have responded positively when the study team has explained that the file will be deidentified in accordance with HIPAA.

Review of Publications Before Release

Another common point of negotiation concerns allowing MIHOPE-Strong Start to produce publications using the agency's data without the agency's review and approval of these publications. This is necessary because it is OPRE's policy to protect its research studies from the appearance or possibility of external influence and bias in the findings. Informing the agencies that the analyses provided in publications will be aggregated across all participating home visiting programs, that MIHOPE-Strong Start is a national evaluation involving many states' data, that agencies and their policies will not be included by name in any reports, and that small cell sizes will be suppressed in the restricted access file has served to assure most agencies that their prior review of publications will not be necessary.

Linking Across Data Sources

As noted earlier, MIHOPE-Strong Start will link various data sources together in order to enhance their usefulness and evaluation capacity. For example, an individual's survey data at the time of enrollment, which includes socio-demographic and family information not found in administrative records, will be linked to Medicaid health claims and birth certificate data through name, date of birth, Social Security number, and Medicaid identification number. This allows the study team to have a richer data set for analysis purposes, affording them the ability to conduct subgroup analyses of impacts on Medicaid data using socio-demographic and psychosocial information (such as marital status or depressive symptoms) to define the subgroup. Some data agencies, however, have been concerned that having their data linked with these other sources will open them up to an unintended risk of unauthorized use of the data. Therefore, the MIHOPE-Strong Start team has had to demonstrate that this part of the study will not subject the data agency to unnecessary risk and that, in fact, protecting the identities of participants is of primary concern to the MIHOPE-Strong Start researchers. This is often achieved by detailing the data processing procedures, such as how data are stored on an encrypted server, which can be accessed only by a limited number of technical staff members who deidentify the data and assign a unique random identification number to each participant before linking data sources into an analytic file.

Data Availability

In addition to the time required to contact data agencies and ultimately to execute agreements to access administrative data, MIHOPE-Strong Start has had to build in time to receive these data. Identifiable data from these sources do not undergo the same amount of processing by the agency as deidentified or aggregated data, but they are still not usable on a "real time" basis. As described in more detail below, states' birth certificate data, which typically are released on an annual basis, may not be available until 24 months after the start of the calendar year in which a birth occurs. For example, the record for a birth in January 2014 may not be available for data extraction until December 2015. Medicaid data typically have a shorter lag, but it still takes some time from the date the service is rendered to when it is paid for or processed by the agency.

The final report for MIHOPE-Strong Start is to be published in September 2017, and receiving timely data is essential for meeting the report schedule. The study team has worked closely with data agencies to understand lags and negotiate for earlier releases of data when needed so that all data can be received by January 2017.

Timing of Birth Certificate Data

Birth certificates reflect all births to residents of a state, and the time it takes to compile and process this information depends on whether the data are classified as "preliminary" or "cleaned." The raw data that data agencies receive from hospital submissions are called preliminary (or provisional) data. Preliminary data are available shortly after the birth, but are not always available to researchers and may have more problems with data quality (for example, duplicate reports or inconsistently labeled information). Cleaned data are produced through the federal mandate for states to submit official vital statistics annually to the NCHS at the CDC; cleaned data files are produced at the end of the calendar year so that all births are captured and undergo a reconciliation process with the NCHS and hospitals to ensure the data are accurate, which usually takes between 6 to 12 months.[42]

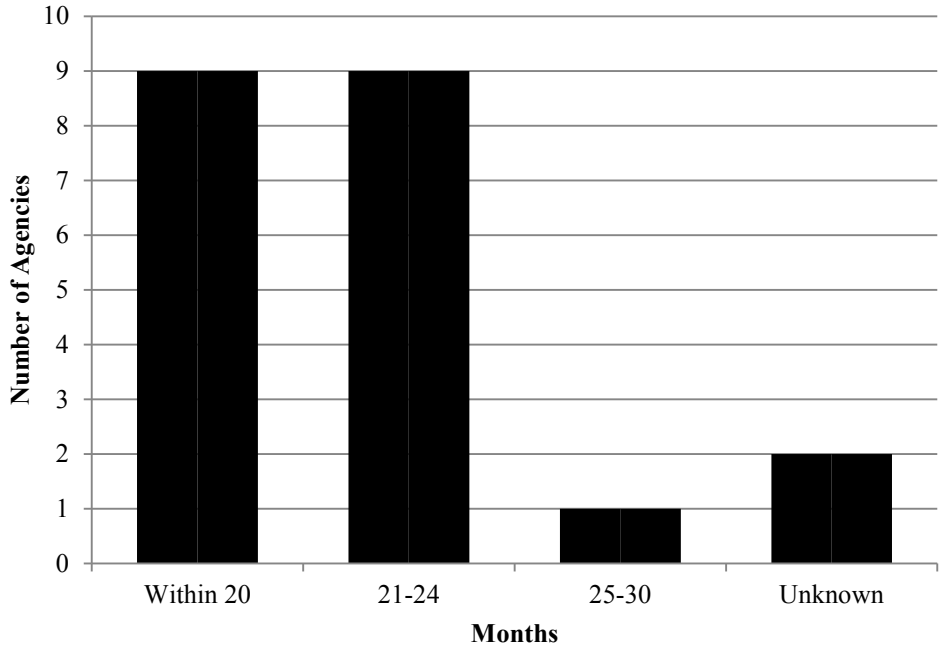
Figure 3 shows that, for most of the data agencies targeted in MIHOPE-Strong Start, birth certificate data (either cleaned or preliminary) are available for the study to use within 24 months of the beginning of the calendar year in which the birth occurred (18 out of 19 agencies; for 2 agencies the timing is unknown). This means that, for births that occur on January 1, it can take up to two years for the records to appear in most states' data systems and to become available for extraction. Additional time is required for the state to send data to researchers and for researchers to process and analyze the data. In many of these cases (13 out of 18 agencies),

[42] 42 U.S.C. § 242k.

The Mother and Infant Home Visiting Program Evaluation-Strong Start

Figure 3

Availability of Birth Certificate Data to Researchers, After the Beginning of a Calendar Year



SOURCE: MDRC calculations based on administrative data acquisition efforts as of November 2014.

NOTES: The dates in this figure show when birth certificate data (either cleaned or preliminary) are available for births that occurred in the beginning of a calendar year. For example, a record for a birth in January 2013 that is available "within 20 months" would be available by August 2014. The figure includes one agency with a data lag between 21 and 24 months that will provide only aggregate data for medical and health information, with a secondary file consisting exclusively of identifiers. The figure reflects the availability of cleaned data if the cleaned data are available within 24 months of the beginning of the calendar year. If it takes longer than 24 months for cleaned data to become available, then the figure relies on the dates that preliminary data are available. Dates available do not include the amount of time it takes for data shipments to occur.

cleaned data will be available for MIHOPE-Strong Start to use within 24 months, but for the other 5 states, MIHOPE-Strong Start will have to use preliminary data if the study team wishes to access records within the same time period.

Timing of Medicaid Data

Medicaid data are available more quickly than birth certificate data and can typically be requested from a state on a real-time basis.[43] In practice, however, researchers need to wait for claims to be submitted by providers and paid or processed by the Medicaid agency, which is typically between three and six months after the date of service, although this can vary by state, type of service, payment method, and data warehousing by the Medicaid agency.

In both fee-for-service claims and managed care encounters, inpatient data tend to have a longer lag than data on other services, such as prescription drug data, which are usually available very quickly. Claims data are typically available earlier than encounter data. Providers often do not file encounter data as quickly as claims because encounter data are not used for reimbursement purposes, and they involve the input of multiple entities within the managed care network.[44] In fact, depending on a state's Medicaid managed care model, the amount of time for encounter data to become available can exceed six months.

Matching to the MIHOPE-Strong Start Sample and Extracting Data

The discussion above focuses on how long it takes for data to become available from a state's system for extraction, but the MIHOPE-Strong Start team must also factor in time for the data agencies to match the research sample to their data systems and to ship the data extracts.[45] In this process, the MIHOPE-Strong Start team sends the sample members' available identifiers to the agency, which uses these identifiers to match to their data system. The data agency then sends its data for the specific MIHOPE-Strong Start sample to the research team.

---

[43]One of the motivations in some states' recent efforts to create an "all-payer claims database" has been to improve access to health claims data (including Medicaid claims and encounters) for researchers. A state data agency may refer a requester to such a database if it exists. The lag time for an all-payer claims database tends to be about 12 months after the close of the calendar year because the data must be received from all health insurance payers and processed into a uniform format.

[44]Encounter data record services rendered to a client but may lack some information on specific charges and reimbursement amounts to providers. This is because managed care organizations have different payment arrangements from fee-for-service providers, which are reimbursed per service (as noted earlier). Medicaid uses encounter data for rate setting for managed care organizations and for monitoring managed care organization performance.

[45]At least one data agency has said it cannot perform the match between its data system and the MIHOPE-Strong Start research sample, so MDRC will conduct this match itself. Data security complications can arise with receiving all data from a data agency's system, so MIHOPE-Strong Start elected to use a "two-stage" matching process in which a data agency sends the study team all identifiers in a data system, including the data system's unique record identification numbers. The MIHOPE-Strong Start study team then matches the identifiers to the research sample and returns all matched identifiers to the data agency. The agency can then extract the identifiable data for the research sample with minimal effort by using these unique record identifiers. While it requires significantly more resources for the study team to perform these matches, the team can be confident in how well the match was performed.

A challenge the MIHOPE-Strong Start team has encountered is that the data agencies vary in terms of which identifiers they need to conduct the match to their data systems. All mothers in the MIHOPE-Strong Start sample provide their name and date of birth at study intake, and most also supply their Social Security number. Relatively few mothers provide their Medicaid identification number because many do not know this information offhand and do not have their identification cards with them at study intake. And because mothers enter the MIHOPE-Strong Start study when they are pregnant, identifiers for their infants cannot be collected at intake.

These aspects of study intake — having neither Medicaid identification number nor Social Security number for some mothers and having no infants' identifiers — have been a challenge for some data agencies. For example, a few Medicaid agencies have said they can extract data from their systems only by using the Medicaid identification number. To collect infants' Medicaid data, some agencies need the infants' own identifiers; they cannot identify them through the mothers. Even in cases where the data agency can link infants to their mothers, it would be more labor intensive to do this type of match.

Finally, data extraction for a defined sample can be resource intensive for the data agencies, and establishing each data agency's match process for MIHOPE-Strong Start has taken time. The MIHOPE-Strong Start team has to negotiate a preferred match process and the identifiers to be included as part of that process based on the data agency's matching capabilities. Many agencies use "exact" matching strategies, where specific identifiers (for example, name, date of birth, and Social Security number) are matched to the identifiers in their data systems but minor discrepancies are not accounted for (such as typos in date of birth or reversed first name and last name), so some records are not considered to be matches even though they may well be. Some agencies employ a more advanced "fuzzy" matching strategy, which accounts for the likelihood that a given set of identifiers may not correspond exactly to those in a data system, but may be close enough to be considered a good match.[46] For example, a date of birth given as 04/28/1982 and 05/28/1982 in two different records might be considered a match even though the date has a one-month discrepancy if other identifiers, such as name and Social Security number, matched well. Depending on the data agency's matching strategy, the personal identifiers that are needed to yield a high match rate may be expansive and flexible or may be narrowly defined. The MIHOPE-Strong Start team prefers match processes that incorporate fuzzy matching logic and use many different identifiers.

[46]Vital records agencies are more likely than Medicaid data agencies to use a fuzzy matching strategy. The CDC has developed software called LinkPlus to facilitate data linkages and matching, and this has established a precedent of using such software for extracting matched data.

Between matching and shipping the data, it can take an agency several weeks or months to send an initial data extract. But once a system is in place to match, extract, and send the data, it should take much less time for an agency to honor subsequent data requests. Having clear data specifications in the legal agreement or in a memo regarding the transmission method, extract date, and matching protocol should help to ensure that this process runs smoothly.

Limitations on Access to Data

For some states, the decision to release data to researchers is a matter of individual choice that is made by an individual when he or she agrees to participate in a study. In this situation, a data agency will supply the information as long as there is a valid consent form and other procedural and data security considerations are met.[47] But not all state agencies targeted for MIHOPE-Strong Start are willing to provide identified Medicaid and birth certificate data based on these conditions alone. Data agencies have a number of other considerations to balance that are a confluence of state, departmental, and programmatic dictates.

One vital records agency is not able to provide identifiable data for the study due to existing state statutes that prohibit release of identifiable data to external researchers. MIHOPE-Strong Start relies on medical and health information that is found in the birth certificate as well as the dates on which certain events occurred, such as the date of the first prenatal care visit. In this case the research team worked collaboratively with the agency to come up with a solution so that a core set of analyses can still be performed. The data agency will provide the study with record-level deidentified data that is matched to the MIHOPE-Strong Start sample, include selected variables from the other data sources in the study, and convert dates to relative measures in weeks, so that instead of having a variable for the date of the first prenatal care visit, the data file will contain a variable for the number of weeks from conception to the first prenatal care visit.

However, three Medicaid agencies have indicated that they cannot provide the study with any data. Two of the three agencies noted that they have limited resources to provide data to external researchers. State Medicaid data agencies generally are more accustomed to providing data to a researcher or organization that is conducting services on behalf of the department or state and thus are more willing to provide data for a research project where the agency can identify a direct programmatic benefit, is sponsoring the research, or is provided with findings

[47]Both the "Common Rule," which is the federal policy for the protection of human research subjects, and the HIPAA Privacy Rule provide guidelines for releasing data for research purposes. Essential for the release of data is a valid consent form. The main components of a valid consent form are codified in the Common Rule (45 C.F.R. Part 46, Subpart A) and in the HIPAA Privacy Regulations (45 C.F.R. §164.508(c)(1) and 45 C.F.R. § 164.508(c)(2)).

from the study about its own state. With Medicaid being a state-administered program, the direction and needs of the Medicaid data agencies can differ by state.

Discussion

This second annual report from the Mother and Infant Home Visiting Program Evaluation-Strong Start (MIHOPE-Strong Start) provides an overview of a central component of the study: acquiring, across multiple states, birth certificate and Medicaid records, which will be used to assess the impacts of home visiting programs on birth outcomes and maternal and infant health care use. Medicaid claims information will also allow the evaluation to identify potential sources of cost savings. These records, as widely recognized by researchers and policymakers alike, contain comprehensive and (mostly) high-quality information on a large number of individuals. When these data are combined and analyzed with other information collected directly from surveys and programs, researchers can build rigorous evidence on whether social and health interventions, of which home visiting programs are an example, achieve improvements in outcomes of policy relevance.

MIHOPE-Strong Start is expansive in scope, gathering information directly from the program staff and families, the U.S. Census, and management information systems, in addition to administrative records. Recent discussions among federal agencies, however, have particularly focused on the untapped potential advantages of using and pooling various administrative data sources to conduct program evaluations and address pressing policy questions in a more timely and cost-efficient manner. Indeed, the information of interest has often already been gathered and does not need to be collected again through costly and duplicative surveys.[48] Administrative data can also be extracted throughout the course of a project without necessitating follow-up with study participants. In addition to cost concerns, an academic advisory panel on the use of administrative data for analyzing policy has noted that it is often impossible to examine state-level policies and initiatives using existing national surveys because they do not have large enough state-specific samples.[49] And existing survey data, the panel noted, are not always up to date. In contrast, administrative records contain information on events as they occur or shortly thereafter, and the information is often provided by staff members who use the

[48]Office of Management and Budget (2014).

[49]The Advisory Panel on Research Uses of Administrative Data (which included researchers, state and federal officials, and experts in data protection) was formed in 1996 at the Northwestern University/University of Chicago Joint Center for Poverty Research. The panel's work was funded by the Department of Health and Human Services to assess the development and dissemination of state administrative data sources that could be used for academic and policy research in the areas of public assistance, public health, and welfare (Hotz, Goerge, Balzekas, and Margolin, 1996).

data for day-to-day programmatic purposes and therefore have an interest in its being detailed and accurate.

Along with these advantages, administrative data house detailed programmatic or clinical information and contain large numbers of observations by which to conduct fine-grained analysis with statistical confidence. In fact, vital records contain the universe of all births to women living in a state, and Medicaid provides health insurance for about one in five Americans. The Medicaid expansion provisions in the Patient Protection and Affordable Care Act, which will increase the income eligibility threshold from 100 percent of the federal poverty level to about 138 percent, have also increased the pool of eligible beneficiaries in states that have chosen to undergo expansion.[50]

Despite these advantages, because administrative data are collected primarily for reporting and, in the case of Medicaid, reimbursement and rate-setting purposes, it is necessary to collect additional data directly from participants in MIHOPE-Strong Start in order to capture important information not found in administrative records. Information on parental relationship status and quality, prior life experiences, unmet need, and potential health and parenting risks, for example, is critical in order to understand the role that participant characteristics play in affecting outcomes. Linking data sources is an important component of many research designs, not just MIHOPE-Strong Start, as often no single data source is sufficient to address the types of questions of interest to a research study. Consider, for example, the administrative data sets in MIHOPE-Strong Start: Medicaid files do not contain information on birth outcomes, and thus cannot be the sole administrative data resource given the evaluation's emphasis on improving birth weight and reducing preterm births.

Based on the study team's experience with the process of acquiring administrative data from two different sources (Medicaid and birth certificate records) in 20 states, implications for other large-scale, national studies needing identifiable administrative data from multiple states are as follows:

- **States' administrative data are not usable in "real time."** Researchers need to learn about each agency's expected lag in data availability when considering how long it will take to complete the research. Birth certificate data are typically processed on a calendar-year basis and may not be available until 24 months after the start of the calendar year in which the birth occurred. Medicaid data are often available earlier, but the lag still may be 3 to 6 months from the time a service is rendered until it is recorded in a data agency's system.

[50]Kaiser Family Foundation (2014).

- **The data acquisition process is not always comparable across agencies within states or across states.** The team has had to systematically learn about and track many aspects of each agency's data system, such as ways the informed consent process has to be modified, if and when identifiable data will be available, how the agency will conduct the match to the study sample, which data security requirements are required, and how much the agency will charge for providing data. Furthermore, there is substantial variation in time across different agencies for each procedural step undertaken to gain access to administrative records.

- **Data acquisition efforts are likely to be easier for studies in which the agency providing data is also the funder.** Even if the agency is not funding the study, the data acquisition process can be streamlined if the agency has a stake in the research or will receive a product (such as a final report).

- **Most agencies have their own IRB (or other review committee), which will have to approve the research study.** In MIHOPE-Strong Start, these committees very rarely have considered the approval that the study already has received from the MDRC IRB, which has federalwide assurance for the protection of human subjects by HHS. Acquiring data from certain agencies even requires the approval of multiple review committees within the state.

- **A research study involving contact with human subjects and receiving informed consent will have to get approval from data agencies for these processes before collecting participants' data.** As such, the MIHOPE-Strong Start research team has had to strike the right balance between the timing of site recruitment and ensuring access to administrative data. Without confirming sites, a couple of data agencies would not review application material. Yet recruitment of individuals in these sites could not occur until the intake procedures for informed consent had been approved by the data agencies and access to data was ensured.

- **Researchers must have the staffing capabilities necessary to broker legal agreements for sharing data.** This often requires not only the guidance of an attorney with extensive knowledge about the research study, state statutes, and HIPAA, but also the capability of other staff members to thoughtfully navigate the varying state systems required for acquiring data from numerous state agencies.

The data acquisition process and requirements encountered in MIHOPE-Strong Start exemplify how the environment for accessing administrative data is changing. In addition to

acquiring Medicaid and vital statistics data, MDRC and its partners have experience collecting data from many other types of administrative sources for other studies, including those involving child welfare, child support, school records, criminal justice records, unemployment insurance, Temporary Assistance for Needy Families, food assistance, child care subsidies, and housing subsidies. Agencies' data security concerns are making the process of acquiring their data longer, more complex, and therefore more costly. State agencies are increasingly requiring detailed data security plans, written assurance (via legal agreement) that the organization will abide by the agency's own security requirements, copies of signed informed consent forms, and a lengthy review of the research study and the organization's right to access the agency's identifiable data. These detailed documents and procedures certainly delay the research process, but they are understandable given the importance of maintaining the privacy and protection of individuals.

Data availability lags may not be of concern for other research purposes, depending on the timeframe, the nature (descriptive versus evaluative), and the scope (negotiating with one agency versus multiple agencies, one state versus multiple states) of the study, or the need to gain access to identifiable administrative records. But as programs and policies become more complex, and the emphasis on quicker and more efficient evaluations increases, the findings detailed in this report based on the MIHOPE-Strong Start administrative data acquisition experience are likely to be relevant to similar research endeavors that aim to conduct large-scale analyses to address evaluative questions of current policy relevance.

References

Avellar, Sarah, and Diane Paulsell. 2011. *Lessons Learned from the Home Visiting Evidence of Effectiveness Review*. Princeton, NJ: Mathematica Policy Research, Inc.

Bhutta, Adnan T., Mario A. Cleves, Patrick H. Casey, Mary M. Cradock, and K. J. S. Anand. 2002. "Cognitive and Behavioral Outcomes of School-Aged Children Who Were Born Preterm: A Meta-Analysis." *Journal of the American Medical Association* 288, 6: 728-737.

Byrd, Vivian, and Allison H. Dodd. 2012. "Assessing the Usability of Encounter Data for Enrollees in Comprehensive Managed Care Across MAX, 2007-2009." Washington, DC: Centers for Medicare and Medicaid Services.

Byrd, Vivian, and James Verdier. 2011. *Collecting, Using, and Reporting Medicaid Encounter Data: A Primer for States. Final Report*. Washington, DC: Mathematica Policy Research.

Card, David, Raj Chetty, Martin Feldstein, and Emmanuel Saez. 2011. "Expanding Access to Administrative Data for Research in the United States." National Science Foundation 10-069 white paper. Website: http://eml.berkeley.edu/~saez/card-chetty-feldstein-saezNSF10dataaccess.pdf.

Centers for Disease Control and Prevention. 1994. *Model State Vital Statistics Act and Regulations*. 1992 Revision. DHHS Publication No. (PHS) 94-1115. Hyattsville, MD: National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services.

Centers for Medicare and Medicaid Services. 2010. "Medicaid Managed Care Enrollment as of July 1, 2010." Website: http://www.cms.gov/Research-Statistics-Data-and-Systems/Research-Statistics-Data-and-Systems.html.

Centers for Medicare and Medicaid Services Innovation Center. n.d. "Strong Start for Mothers and Newborns Initiative: General Information." Website: http://innovation.cms.gov/initiatives/strong-start/.

Cuevas, Katherine D., Debra R. Silver, Dorothy Brooten, JoAnne M. Youngblut, and Charles M. Bobo. 2005. "Hospital Charges at Birth and Frequency of Rehospitalizations and Acute Care Visits over the First Year of Life: A Comparison by Gestational Age and Birthweight." *The American Journal of Nursing* 105, 7: 56.

Filene, Jill H., Emily K. Snell, Helen Lee, Virginia Knox, Anne Duggan, and Charles Michalopoulos. 2013. *The Mother and Infant Home Visiting Program Evaluation-Strong Start: First Annual Report*. OPRE Report 2013-54. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Fine, Amy, Milton Kotelchuck, Nancy Adess, and Cheri Pies. 2009. "A New Agenda for MCH Policy and Programs: Integrating a Life Course Perspective." Policy brief. Martinez, CA: Contra Costa Health Services. Website: http://www.cchealth.org/groups/lifecourse/pdf/2009_10_policy_brief.pdf.

Hack, Maureen, Nancy K. Klein, and H. Gerry Taylor. 1995. "Long-Term Developmental Outcomes of Low Birth Weight Infants." *The Future of Children* 5, 1: 176-196.

Harwood, Robin, Stella Yu, and Laura Kavanagh. n.d. "Remembering Our Past, Building the Future: 100 Years of the Maternal and Child Health Research Program." Maternal and Child Health Bureau, Health Resources and Services Administration, U.S. Department of Health and Human Services. Website: mchb.hrsa.gov/research/.../100_Years_of_MCH_Research_Program.docx.

Hetzel, Alice M. 1997. *U.S. Vital Statistics System: Major Activities and Development, 1950-95*. DHHS Publication No. (PHS) 97-1003. Hyattsville, MD: National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services.

Hotz, Joseph, Robert Goerge, Julie Balzekas, Frances Margolin, eds. 1996. *Administrative Data for Policy-Relevant Research: Assessment of Current Utility and Recommendations for Development*. The Advisory Panel on Research Uses of Administrative Data from the Northwestern University/University of Chicago Joint Center for Poverty Research.

Institute of Medicine (U.S.) Committee on Understanding Premature Birth and Assuring Healthy Outcomes. 2007. *Preterm Birth: Causes, Consequences, and Prevention*. Washington, DC: National Academy of Sciences.

Kaiser Commission on Medicaid and the Uninsured. 2004. "A Brief Overview of Medicaid Data Sources." The Henry J. Kaiser Family Foundation. Website: http://kaiserfamilyfoundation.files.wordpress.com/2013/01/overview-of-differences-between-data-sources-cms-64-and-msis.pdf.

Kaiser Family Foundation. 2014. "How Will the Uninsured Fare Under the Affordable Care Act?" The Henry J. Kaiser Family Foundation. Website: http://kaiserfamilyfoundation.files.wordpress.com/2014/04/8531-how-will-the-uninsured-fare-under-the-aca.pdf.

Kitzman, Harriet, David L. Olds, Charles R. Henderson, Carole Hanks, Robert Cole, Robert Tatelbaum, Kenneth M. McConnochie, Kimberly Sidora, Dennis W. Luckey, David Shaver, Kay Engelhardt, David James, and Kathryn Barnard. 1997. "Effect of Prenatal and Infancy Home Visitation by Nurses on Pregnancy Outcomes, Childhood Injuries, and Repeated Childbearing: A Randomized Controlled Trial." *Journal of the American Medical Association* 278, 8: 644-652.

Klemm, John. 2000. "Medicaid Spending: A Brief History." *Health Care Financing Review* 22, 1: 105-112.

Kogan, Michael D., Wanda Barfield, and Charlan Kroelinger. 2014. "The Evolving Role of Leadership and Change in Maternal and Child Health Epidemiology." *Maternal and Child Health Journal*. Website: http://www.mchnavigator.org/documents/Kogan_MCHB-authored_2014_07_31.pdf.

Lackritz, Eve. 2004. "Meeting the Challenges of Prematurity: CDC Prevention Efforts," Centers for Disease Control and Prevention. Testimony before the Subcommittee on Children and Families, Committee on Health Education, Labor and Pensions of the U.S. Senate, May 12. Website: http://www.hhs.gov/asl/testify/t040512b.html.

Lee, Eunju, Susan D. Mitchell-Herzfeld, Ann A. Lowenfels, Rose Greene, Vajeera Dorabawila, and Kimberly A. DuMont. 2009. "Reducing Low Birth Weight Through Home Visitation: A Randomized Controlled Trial." *American Journal of Preventive Medicine* 36, 2: 154-160.

Lu, Michael C., and Neal Halfon. 2003. "Racial and Ethnic Disparities in Birth Outcomes: A Life-Course Perspective." *Maternal and Child Health Journal* 7, 1: 13-30.

March of Dimes. 2003. "Hospital Charges for Prematurity: 2001 Data." March of Dimes Perinatal Data Center.

March of Dimes, Partnership for Maternal, Newborn and Child Health, Save the Children, and World Health Organization. 2012. *Born Too Soon: The Global Action Report on Preterm Birth*. Geneva: World Health Organization.

Martin, Joyce A., Brady E. Hamilton, Stephanie J. Ventura, Michelle J. K. Osterman, Elizabeth C. Wilson, and T. J. Mathews. 2013a. *Births: Final Data for 2012*. National Vital Statistics Report 61, 1. Hyattsville, MD: National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services.

Martin, Joyce A., Elizabeth C. Wilson, Michelle J. K. Osterman, Elizabeth W. Saadi, Shae R. Sutton, and Brady E. Hamilton. 2013b. *Assessing the Quality of Medical and Health Data from the 2003 Birth Certificate Revision: Results from Two States.* National Vital Statistics Reports 62, 2. Hyattsville, MD: National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services.

Michalopoulos, Charles, Anne Duggan, Virginia Knox, Jill H. Filene, Helen Lee, Emily K. Snell, Sarah Crowne, Erika Lundquist, Phaedra S. Corso, and Justin B. Ingels. 2013. *Revised Design for the Mother and Infant Home Visiting Program Evaluation*. OPRE Report 2013-18. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Murphy, Sherry L., Jiaquan Xu, and Kenneth D. Kochanek. 2013. *Deaths: Final Data for 2010, National Vital Statistics Report*. Hyattsville, MD: National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services.

National Association for Public Health Statistics and Information Systems (NAPHSIS). 2013. *More, Better, Faster: Strategies for Improving the Timeliness of Vital Statistics*. Silver Spring, MD: NAPHSIS.

Office of Child Support Enforcement. 2012a. "Authorized Agencies Receiving National Directory of New Hires (NDNH) Information." Dated July 11. Washington, DC: Administration for Children and Families, U.S. Department of Health and Human Services.

Office of Child Support Enforcement. 2012b. "A Guide to the National Directory of New Hires (NDNH) Information." Revised July 11. Washington, DC: Administration for Children and Families, U.S. Department of Health and Human Services.

Office of Management and Budget. 2013. "Next Steps in the Evidence and Innovation Agenda." Memorandum to the Heads of Executive Departments and Agencies, July 26. M-13-17. Washington, DC: Executive Office of the President. Website: www.whitehouse.gov.

Office of Management and Budget. 2014. "Guidance for Providing and Using Administrative Data for Statistical Purposes." Memorandum to the Heads of Executive Departments and Agencies, February 14. M-14-06. Washington, DC: The Executive Office of the President. Website: www.whitehouse.gov.

Olds, David L., Charles R. Henderson, Robert Chamberlin, and Robert Tatelbaum. 1986. "Preventing Child Abuse and Neglect: A Randomized Trial of Nurse Home Visitation." *Pediatrics* 78, 1: 65-78.

Olds, David L., JoAnn Robinson, Ruth O'Brien, Dennis W. Luckey, Lisa M. Pettitt, Charles R. Henderson Jr., Rosanna K. Ng, Karen L. Sheff, Jon Korfmacher, Susan Hiatt, Ayelet Talmi. 2002. "Home Visiting by Paraprofessionals and by Nurses: A Randomized, Controlled Trial." *Pediatrics* 110, 3: 486-496.

Paneth, Nigel S. 1995. "The Problem of Low Birth Weight." *The Future of Children* 5, 1: 19-34.

Shiono, Patricia H., and Richard E. Behrman. 1995. "Low Birth Weight: Analysis and Recommendations." *The Future of Children* 5, 1: 4-18.

Solomon-Fears, Carmen. 2014. "The National Directory of New Hires." Report prepared for members and committees of Congress, February 24. Washington, DC: Congressional Research Service.

Wheaton, Laura, Christin Durham, and Pamela Loprest. 2012. *TANF and Related Administrative Data Project: Final Evaluation Report.* Washington, DC: The Urban Institute.