HHS Public Access

Author manuscript

J Midwifery Womens Health. Author manuscript; available in PMC 2019 February 11.

Published in final edited form as:

J Midwifery Womens Health. 2017 September ; 62(5): 545–561. doi:10.1111/jmwh.12640.

UTILIZING DATASETS TO ADVANCE PERINATAL RESEARCH

Julia C. Phillippi, CNM, PhD, FACNM,
Assistant Professor of Nursing at Vanderbilt University School of Nursing in Nashville

Jeremy L. Neal, CNM, PhD,
Assistant Professor at Vanderbilt University School of Nursing, Nashville

Nicole S. Carlson, CNM, PhD,
Assistant Professor of Nursing at Emory University School of Nursing in Atlanta GA

Frances M. Biel, MPH, MS,
Research Associate in the Department of Obstetrics and Gynecology at Oregon Health & Science University in Portland, OR

Jonathan M. Snowden, PhD, and
Assistant Professor in the Department of Obstetrics & Gynecology and in the OHSU/Portland State University School of Public Health

Ellen L. Tilden, CNM, PhD
Assistant Professor in the Nurse-Midwifery Department in Portland Oregon

Abstract

Many organizations collect and make available perinatal data for research and quality improvement initiatives. Analysis of existing data and use of retrospective study design has many advantages for perinatal researchers. These advantages include large samples, inclusion of women from diverse groups, data reflective of actual clinical processes and outcomes, and decreased risk of direct maternal and fetal harm. We review 11 publicly available datasets relevant to perinatal research and quality improvement, detail the availability of interactive websites, and discuss strategies to locate additional datasets. While analysis of existing data has limitations, it may provide statistical power to study rare perinatal outcomes, support research applicable to diverse populations, and facilitate timely and ethical well-woman research immediately relevant to clinical care.

INTRODUCTION

Conduct of research and quality improvement is increasingly a professional expectation for midwives and other perinatal care providers. Over 7% of full-time midwives in the United States (US) conduct research,[1] and many more conduct quality improvement and benchmarking. The Core Competencies for Basic Midwifery practice specify that research

CORRESPONDING AUTHOR: Julia Phillippi, Julia.c.phillippi@vanderbilt.edu, 517 Godchaux Hall, 461 21st Avenue South, Nashville, TN 37240.

CONFLICT OF INTEREST DISCLOSURE
The authors have no conflicts of interest to disclose.

and quality improvement skills are entry-level professional expectations.[2 p.2] Additionally, many certified nurse-midwives (CNMs), certified midwives (CMs), and advanced practice nurses pursue doctoral training to facilitate data use for assessment and/or outcomes improvement. Though only 4.1% of CNMs/CMs had a doctoral degree in 2001, 11.6% of CNMS/CMs were doctorally-prepared in 2012, and numbers continue to increase.[1, 3, 4]

There is clear need for more and increasingly rigorous science to guide improvements in perinatal care. Neonatal outcomes in the United States have shown only modest improvements in the past 5 years [5, 6] and are worse than those in resource-comparable nations.[7] Importantly, US maternal morbidity and mortality is increasing.[8] These factors and striking health disparities in maternity outcomes between racial and ethnic groups signal a need for stronger perinatal research and effective quality improvement.[9, 10]

National organizations such as the American College of Nurse-Midwives (ACNM), the American College of Obstetricians and Gynecologist (ACOG), and the Agency for Healthcare Research and Quality (AHRQ) have called for improvement in the provision of high-quality, evidence-based perinatal care.[11–13] High-quality perinatal data are essential to reaching this goal. Though experimental trials have many strengths for use in quality improvement and research, and remain the scientific gold standard, they frequently are limited by small samples lacking statistical power to detect changes in important maternal, fetal, or neonatal outcomes. In addition, it is sometimes not feasible to incorporate randomization or blinding in research involving pregnant women. These limitations may be especially problematic in the conduct of well-women perinatal research because this population may be disinclined to participate and severe outcomes are rare.

Strengths of Dataset Analysis for Perinatal Research

Large, existing datasets offer a wealth of information about the effects of healthcare services and treatments on maternal-child health outcomes. Arguably, the most important strength of utilizing large datasets for well-woman perinatal research is that these datasets allow for study of rare outcomes (eg, cord prolapse) infrequently encountered in a single location or trial. Since this research does not change care routinely provided to pregnant women, fewer ethical considerations may emerge, potentially facilitating expedited institutional review board (IRB) approval and shortening time between identifying a scientific gap and disseminating research findings. Further, datasets enable the aggregation of information from diverse women in various settings, ideally using clearly-defined operational definitions for each measure.[8] In experimental studies, the intervention is often tightly controlled and provided to a homogenous sample in order to ensure treatment fidelity and boost internal validity; however, strict protocols limit external validity and generalizability.[14] As well, large databases may include greater variations in treatment, practice patterns, and care settings, which more accurately capture the variability of pregnancy, labor, and birth processes and outcomes as well as women’s choices. This may enable broader and more accurate insight into the effects of maternity care systems and interventions in existing clinical settings.[15]

Analysis of such large existing datasets frequently provides sample sizes with the statistical power to assess care using fewer resources than randomized controlled trials (RCTs).

Existing data have such value that the National Institutes of Health (NIH) requires researchers receiving more than $500,000 in federal funding to make their data available for secondary analysis.[16]

Limitations of Dataset Analysis for Perinatal Research

Analysis of large, existing maternal-child health datasets also has several limitations. While data analysis may produce information that accurately reflects maternity care processes and outcomes, data are often affected by confounding bias (participants and/or providers select treatment or non-treatment in non-random ways), introducing differences into treatment groups and bias into effect estimation. Also important, large existing datasets constrain research questions. For example, a researcher may wish to study the pregnancy outcomes of nulliparous women carrying a term, single, vertex presenting fetus (NTSV), but a dataset only captures whether women are nulliparous, term and with a singleton pregnancy and does not include fetal presentation. The researchers must then choose to study nulliparous, term, singleton pregnancies or to pursue the question with a different dataset. The conduct of secondary data research is characterized by many similar branching decision points. For this reason, researchers must become thoroughly familiar with a dataset: its provenance, contents, validity, and completeness of all relevant data elements. The investigators should use this knowledge to frame research questions. In our experience, this process is an iterative and, ideally, team-based endeavor. Although this process is recursive and defines the questions that can be answered with a data resource, it does not supplant the scientific method (eg, observation, hypothesis formulation and testing) central to any science.

Collection and entry errors are a universal feature of datasets. Primary investigators may have access to the personnel, equipment, or original data (eg, an electronic health record) involved in the error, enabling correction of errors. In contrast, with secondary analysis or retrospective research, it may be impossible to obtain further information, including how errors arose. These limitations must be thoughtfully considered and weighed against the strengths of secondary analysis.

The purpose of this article is to review publically-available datasets for study of perinatal care in the United States. We will describe current databases with perinatal-health content, detailing the benefits and limitations of using these data for quality improvement and research. In addition, we will highlight new analytical approaches that increase the validity of dataset analysis. Additional datasets and analytic techniques are available; we highlight those most relevant to perinatal sciences.

APPROACH

We identified potentially-relevant perinatal databases for review and invited input from maternity-care clinical scientists and epidemiologists across the country. Informed by this comprehensive list, we discussed the merits of inclusion of each dataset as a group; through this iterative process, we refined the selection of databases. The entire author team determined which databases were most pertinent for perinatal and health services research, quality, and benchmarking. Each dataset was assigned to an initial author who conducted the preliminary research, obtaining the information for each database including: content, value,

timeframe of data available, steps for access, contact information, and published examples of dataset use. A second author then confirmed information for accuracy. Differences were reconciled through author discussion and additional research. Data were sent directly between authors until aggregated by the primary author. Seeking to maximize utility and efficiency for the reader, databases were placed into perinatal and non-perinatal categories. The datasets most pertinent to perinatal research and quality improvement are organized alphabetically in the next section.

Terms used to describe analysis of existing data are defined in Table 1. These terms differentiate research by study design and origin of the dataset. Research can be further differentiated within each of these broad terms (eg, administrative data research is a form of secondary data analysis). Though terms are not interchangeable, definitions overlap. Data analysis may use a retrospectively-collected dataset (for example, a retrospective cohort study using routinely-collected birth records), or use a prospectively-collected dataset (for example, a study using Community Child Health Research network data).

DATABASES FOR PERINATAL RESEARCH

Agency for Healthcare Research and Quality

This agency within the US Department of Health & Human Services is focused on improving the quality, safety, effectiveness, and efficiency of healthcare.[19] One part of the AHRQ mission is to generate measures and data to track and improve US systems-level performance and progress. Available data from hospitals include: the Healthcare Cost and Utilization Project (HCUP), the Nationwide Inpatient Sample, and State Inpatient Databases. The AHRQ database also includes information collected from individuals and families, like the Medical Expenditure Panel Survey from individual states and from data standard experts (US Healthcare Information Knowledgebase).

Perinatal researchers might be especially interested in the HCUP data. This is one of the most comprehensive sources of US hospital data, including information on in-patient care, ambulatory care, and emergency department visits in both nationwide and state-specific databases.[20] Data capture health care delivery and patient outcomes variables over time or by region, state, or community. This information was aggregated from federal, state, and industry sources and includes patient demographics and outcomes by diagnosis and procedure codes. Among the nationwide HCUP databases, the National Inpatient Sample (abbreviated as NIS) is of particular interest for perinatal researchers as it is the largest publically available, all-payer patient health care database in the United States. The National Inpatient Sample approximates a 20% stratified sample of all hospital discharges in US community hospitals. The stratification is to ensure that a range of geographic locations and birth locations are included. Also included in HCUP are state-specific databases; 48 states currently participate in the State Inpatient Database, and 31 states currently make this information available to the public for a reasonable fee. Detailed information is available on the AHRQ Healthcare Cost and Utilization Project website. The challenges of this dataset (eg, absence of key variables such as parity and gestational age; lack of linkage between mothers and neonates) are counterbalanced by the unique sample, which is representative of the US population and includes detailed information from many states. For this reason,

HCUP data have been used to generate high-impact perinatal outcomes and care delivery systems science, such as Kozhimannil's research on variation in cesarean rates across US hospitals.[21]

American Association of Birth Centers Perinatal Data Registry

The American Association of Birth Centers (AABC) has collected data on birth outcomes since 1996. Originally known as the Uniform Data Set, it was created for birth center use. The dataset, now known as the Perinatal Data Registry, has been expanded to include variables relevant to any birth location, and information can be entered by any maternity care practice. Data from 1997 forward are available for analysis (Susan Stapleton CNM, DNP, written communication, September 2016). Following validation,[22, 23] this dataset has been used to study birth center outcomes in the United States.[24, 25]

While the focus of the dataset is on physiologic birth,[26] data are entered following the initial prenatal visit, and maternal and newborn health are followed until 6 weeks postpartum. Therefore, these data include information about the history and prenatal, intrapartum, and postpartum care of women and their children with a range of birth outcomes.[26] A list of variables is available on the website.[26] With its prospective collection, the Perinatal Data Registry contains a wealth of information about women initially planning birth center birth. However, birth center practice varies by site, especially in eligibility criteria for intrapartum admission and provider-type,[25] and site-level practice information is not paired with perinatal outcomes. Another limitation is that definitions within this dataset may not match those used in other databases and, since individual-level detail is not available, this prevents meta-analysis using data from multiple registries.

Requests for registry data are differentiated by whether only summary statistics are needed or if a dataset is required. Statistical summaries of data can be obtained through a routine data request, and IRB approval is not needed, as individuals and facilities are de-identified. If a dataset is requested, AABC requires a formal data request and agreement be accompanied by an IRB letter verifying that the study is exempt from review. Prior to results dissemination, researchers must submit any proposed presentation, abstract, or manuscript for approval by the AABC Board of Directors and Research Committee.

Centers for Disease Control and Prevention Databases

Pregnancy Risk Assessment Monitoring System—The Pregnancy Risk Assessment Monitoring System (PRAMS), collects state-specific, population-based data on maternal attitudes and experiences before, during, and shortly after pregnancy. This US surveillance project was developed in 1987 by state-level departments of health and the Centers for Disease Control and Prevention (CDC), Division of Reproductive Health. Forty-seven states, New York City, Puerto Rico, the District of Columbia, and the Great Plains Tribal Chairmen's Health Board currently participate; two other states (California and Ohio) previously participated. The current jurisdictions capture 83% of US live births. To obtain data, participating states or locations use birth certificate information to sample 1,300–3,400 women who recently birthed in that locale. Most states oversample women who had low

birthweight infants, and many states stratify by mother's race or ethnicity. State-level information is subsequently aggregated to form the full dataset.

The PRAMS questionnaire is composed of 'core' questions asked in all states Participants respond to a maximum of 52 questions on current PRAMS Core Questionnaire (Phase 8; released June 2016), if all topical areas are applicable to the responder. Core questions capture maternal attitudes and perceptions about the most recent pregnancy, content and source of prenatal care, maternal alcohol and tobacco consumption, physical abuse before and during pregnancy, pregnancy-related morbidity, contraceptive use, women's knowledge of pregnancy-related issues, and infant care. States may also select from among a few hundred additional questions developed by the CDC, which reflect additional topics of interest such as mental health, injury prevention, and social support, or add their own questions. As a result, each state's questionnaire is unique, but some information is available across several or all states. The questionnaire is available in English and Spanish.

This database enables analysis of state-level or aggregate national data. Data can be used to estimate changes in population-health status, measure progress toward maternal-child health goals, and identify groups at high risk for health problems. In particular, the data contain more information on a woman's social environment, family situation, and socioeconomic position than other datasets, making it a valuable resource to study social determinants of perinatal health especially in relationship to health bevhaviors.[27, 28] Although PRAMS data are self-reported, data reliability and validity is reportedly high when compared with other population-based data-collection systems, such as US birth certificates.[29]

If the investigator simply wants statistics or basic analysis, the PRAMStat interactive data portal permits data queries without requiring IRB approval or dataset download. Seven different versions of the PRAMS questionnaires exist (Table 2). Researchers may request the analytic research file by submitting an application to the CDC using the steps outlined in Table 2. [30]

US National Vital Statistics Reports—The CDC's National Center for Health Statistics (NCHS) determines the format and content of data collected via the US Standard Certificate of Live Birth, Death, and Fetal Death. Each state is required to collect and report the data elements contained in US Standard Certificate to the NCHS, although individual states issue their own vital record certificates and may choose to collect additional data. Datasets are divided by geographic area (United States and US territories), but the publicly available online datasets do not provide location details, such as state. These datasets are available via the CDC website without a data use agreement, IRB approval, or fee.[31] For investigators whose research questions have geographic components, access to vital statistics data with geographic information can be requested either from the National Association for Public Health Statistics and Information Systems or through the NCHS Research Data Center. Requests require a data use agreement and a processing fee and are reviewed to ensure participant confidentiality and data security measures.

A variety of datasets and years are available. Each dataset has a corresponding user's guide including definitions for terms and variables. Simpler datasets, such as birth and death data

files, are available approximately 2 years after the calendar year ends. Linked datasets (eg, birth data files) joined to infant death data files enable more nuanced analysis than a lone dataset. For instance, perinatal death may be of interest only in the context of unplanned out-of-hospital births, but infant death is not on the birth certificate. In this case, a linked dataset is required, in which each record is connected between the birth dataset and the death dataset, providing the ability to analyze data from each source. Period and cohort-linked infant death data files have a longer release time, sometimes over 5 years.

Choice of dataset and time span are important considerations as there is a tradeoff between timeliness and completeness. Period-linked data are available for more recent years than cohort-linked but are cross-sectional and less preferred for detailed analytic questions. Alternately, cohort-linked data follow the group of infants born in one year for an entire year (ie, capturing death in the first year of life), optimal for detailed infant-level analyses.

While it is possible to compare outcomes across time or by region of the country, data are not uniformly collected. The most recent changes to the US Standard Certificate occurred in 2003; however, revised certificate uptake has been gradual. As of January 1, 2014, 3 states had not yet implemented the 2003 revision of the Standard Certificate of Live Birth.[32] (Supporting Information: Appendix S1). For analyses requiring data from 2003 revision, the proportion of the total births using the revised certificate affects statistical power. These limitations must be weighed against the key advantage of the datasets; they contain information on every birth in the United States. Because of this advantage, NCHS data have been widely used in US population-based perinatal research and surveillance.[8]

Centers for Medicare & Medicaid Services

The US Centers for Medicare & Medicaid Services (CMMS) collects data on enrollment and utilization of these healthcare payment systems.[33] Both participant data and information on physician and professional provider characteristics, prescription drugs, and facility characteristics are available from this rich data source. Some datasets do not include identifiable information and are publically available, while other identifiable datasets require approval. Of particular interest to perinatal researchers are the research-identifiable files, which include information on claim records for a variety of Medicaid charges, including physician services, laboratory or X-ray, and clinic services linked by diagnosis, length of stay, and payment amount. For example, a recent study using CMMS data found that approximately 20% of reproductive-aged women in New York's Medicaid program received opioid prescriptions from outpatient settings from 2008–2013.[34]

Researchers can also access personal summary data for every individual enrolled in Medicaid or Medicare during specific years and in specific states, including demographic, eligibility, and utilization information. Strengths of these datasets from the CMMS are their inclusion of rich beneficiary-level protected health information for inclusion in research analyses. Limitations of these datasets are that not all states participate in all datasets, with participation varying from year to year. The approval process to access research-identifiable information requires 3–4 months and can involve a fee; however, this fee is waived for student use.[35] Researchers must apply for use of identified and de-identified datasets

through the Research Data Assistance Center and complete a data use agreement for access to identifiable or limited-use datasets.[36]

Eunice Kennedy Shriver National Institute of Child Health and Human Development Data and Specimen Hub

The *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) has established the Data and Specimen Hub (DASH) as a mechanism for institute funded investigators to comply with NIH data-sharing policies. This repository enables other investigators to access data from the Institute's funded studies for secondary research.[16]

There are 24 studies in the specimen hub that can be searched via topic (eg, high-risk pregnancy, HIV/AIDS, infant health, labor and birth), by study (eg, type, clinical research network, enrollment dates), dataset (type and format), or document type (eg, study protocol, codebook and/or variable dictionary).[16] Researchers can create an account to request data.

Access to individual-level data requires the investigator to submit forms as outlined in Table 2. An IRB approval may be required, but the dataset contains only de-identified data that are coded according to standards set in the Health and Human Services Regulations for the Protection of Human Subjects and the Health Insurance Portability and Accountability Privacy Rule (HIPAA). The specimen hub's data access committee reviews data requests to determine whether the proposed research use is scientifically and ethically appropriate and congruent with data use limitations. Some studies receive automatic approval. Instructions are on the National Institute of Child Health and Human Development Data and Specimen Hub website.[37]

Examples of datasets within the Eunice Kennedy Shriver National Institute of Child Health and Human Development Data and Specimen Hub

Community Child Health Research Network: The Community Child Health Research Network was a multi-site, prospective cohort study conducted from 2004–2009 to examine how community-, family-, and individual-level stressors influence and interact with biological factors to affect maternal and child health.[38] The study examined how these variables are associated with disparities in pregnancy outcomes and in infant or early childhood mortality and morbidity. The research blended social, behavioral, and biomedical approaches into a community-linked study. This de-identified dataset includes 4,837 people; 3,079 of them postpartum women following a livebirth at ≥ 20 weeks gestation and 1,758 spouses. Most participants were from lower socioeconomic levels, living in African American, Latina, or white communities in 5 regions of the United States. The dataset sampled 3 urban regions, one mixed urban-suburban region, and one rural region. Examples of uses of Community Child Health Research Network data include examining racial disparities in postpartum depression,[39] testing associations between psychosocial stress and C-reactive protein in women during the first postpartum year,[40] and clarifying causal pathways between sleep and postpartum depression.[41]

Consortium on Safe Labor Dataset: The Consortium on Safe Labor (CSL) was a multi-site prospective observational cohort study conducted between 2002–2008 in 12 clinical

institutions (19 hospitals) located in all 9 districts of the American College of Obstetricians and Gynecologists in the United States.[42] The CSL was formed to describe contemporary labor progression, identify the most appropriate time for a cesarean birth among women with labor protraction or arrest, and explore causes of the high US cesarean birth rate. This de-identified dataset includes detailed information from electronic medical records on 228,562 births as well as additional variables regarding the woman, her labor and birth, and her neonate.[43] The CSL dataset includes information on maternal demographics, pregnancy details, labor interventions, labor outcomes, and newborn interventions and outcomes as well as information on the hospital and provider for each woman. In addition, the CSL includes repeated measures on oxytocin dosing and cervical examinations for each participant throughout labor. The CSL cohort has been analyzed in seminal studies on a variety of perinatal topics[44–47] (eg, labor progression, obstetric procedure use), and is publicly-available for further secondary analysis.

Screening for Risk Factors for Spontaneous Preterm Delivery: Maternal-Fetal Medicine Unit Network: This multi-site, prospective, observational study was conducted from 1992–1995 to evaluate tests that define a group of women with at least a 2-fold increase in risk for spontaneous preterm birth. The tests evaluated included a demographic, behavioral, psychological, anthropometric and historical profile, serum and plasma levels of various proteins such as C-reactive protein, major basic protein and alpha-fetoprotein, vaginal ultrasound evaluation of the cervix, a cervical digital examination, the presence of bacterial vaginosis and trichomonas, and both vaginal and cervical fetal fibronectin. This de-identified dataset includes 3,073 women with singleton pregnancies. The primary outcome was birth at 23–34 weeks gestation following spontaneous preterm labor or premature rupture of membranes.

Listening to Mothers

The Maternity Center Association's Listening to Mothers initiative aims to understand the experiences and views of US childbearing women delivering a single child in a hospital setting. Multiple variables are captured on participating women who complete online surveys about their maternal care experiences. Five national *Listening to Mothers* surveys have been conducted including 3 initial pregnancy and birth surveys and 2 follow-up postpartum surveys. Results reveal gaps between women's preferences for care and the care they receive. Such detailed data on women's perinatal care experiences and perspectives are not available in other datasets and can be used by clinicians, educators, and researchers to inform maternity care practice, education, and policy.[48, 49] The datasets from these surveys are publically available in the Odum Institute Dataverse Network Data Archive.[50] *Listening to Mothers* data have been used to examine the effects of the perceived support during labor and delivery on women's positive and negative evaluations of their birth experiences,[51] assess racial and ethnic disparities in patient-reported communication problems and perceived discrimination in maternity care,[49] and examine the association between workplace accommodations for pregnant employees (eg, availability of paid and unpaid maternity leave) and changes in women's health insurance coverage postpartum,[48] among many other works.

Midwives Alliance of North America Statistics Project

The Midwives Alliance of North America Statistics Project (MANA Stats) datasets include data on pregnancy as well as labor and early postpartum processes and outcomes. The majority of providers contributing to this dataset are US certified professional midwives (CPMs), practicing in home or birth center settings. Approximately 30% of all US CPMs contribute to MANA Stats data collection efforts.[52] Participation is voluntary; however, Oregon, Washington, and Vermont mandate CPM participation, and other states are considering this approach (M. Cheyney, PhD, oral communication October, 2016). A smaller proportion of data is collected by CNMs/CMs.

MANA Stats comprises 3 datasets, each with different granularity for variables that predominantly capture lower-intervention care processes and outcomes of the childbearing cycle. The relatively low rate of obstetric or hospital-based interventions within this dataset makes it particularly relevant for the study of physiologic labor and birth processes. Care outcomes are also captured for the approximately 12% of women or neonates who begin care in the out-of-hospital setting but transfer to the hospital.[53]

When women initiate prenatal care with midwives who participate in MANA Stats data collection, those willing are consented for participation (>95% of eligible women consent), [52] and data about early prenatal care are entered. Additional data are entered during pregnancy, birth, and the early postpartum period. The online system alerts MANA Division of Research leadership when longitudinal data of a particular participant are missing, and processes exist to directly engage the responsible midwife to prompt completion.[52] This prospective structure and well-organized system for accountability improve data quality and validity.[52] The steps for using MANA Stats data for research are in Table 2.

Strong Start for Mothers and Newborns Initiative Data

The Strong Start for Mothers and Newborns initiative is a project by the US Department of Health and Human Services. Its goal is to improve maternal and infant outcomes among pregnant women with insurance coverage through Medicaid and the Children's Health Insurance Program (CHIP).[54] This 5-year initiative (2013–2018) compares women with Medicaid coverage receiving standard care versus those enrolled in 1 of 3 types of enhanced prenatal care: 1) maternity care homes (62.4% of sites); 2) group prenatal care (19.7% of sites); 3) and birth center care (17.8% of sites). Comparisons aim to discover if these enhanced prenatal models reduce preterm births, improve outcomes, and/or decrease cost of care during the first year of life.

The Strong Start project includes 27 awardees and 213 sites in 30 states, the District of Columbia, and Puerto Rico with an estimated sample of 80,000 women.[54] The Strong Start dataset includes participant- and program-level data collected quarterly from each site. Participant-level data include type of enhanced prenatal care model, maternal demographic characteristics, pregnancy risk characteristics, pregnancy outcomes, costs for women and infants over first year of life, and maternal satisfaction with care. Program-level data include baseline and ongoing aggregate information on pregnancy outcomes and model descriptions. In addition, CMS is working with states to collect linked vital statistics data obtained from

birth certificates, Medicaid eligibility, Medicaid claims, and encounter data on mothers and infants in the comparison group receiving standard care. Quality checks led by CMS are conducted annually. Initial aggregate data for project years 1–2 are available at the Strong Start website;[54] the full Strong Start dataset should be completed in 2018.

Local or Regional Databases

Local or regional research communities and clinical practices are further data sources. Many research datasets are owned by the primary investigator and may be underutilized after primary findings are published. Investigators may not have time or funding to completely query an existing dataset. To locate these datasets, researchers can search publications in their area of interest, with attention to investigators' data collection methods, then contact authors to inquire about secondary analysis opportunities.

Researchers might also locate datasets through local or regional clinical practice organizations. For example, clinical benchmarking or quality improvement activities require providers to collect patient demographic, pregnancy, labor, and birth information. Regional health entities, such as departments of health and hospitals, may also collect clinical data from electronic health records or vital statistics data. Although these local datasets typically include identifiers and require IRB approvals to access, they can be a rich source of data for secondary analysis. Administrative approval for access may require time and data management, and dataset cleaning is often required. However, cost for access to these datasets can be low compared to national datasets. Researchers seeking local datasets should work with their state or county to identify available variables, necessary regulatory approval, and required costs.

Additional Health-Related Datasets

There are a variety of datasets focused on alternative topics that include maternal-child health information. These databases may contain information directly applicable to perinatal health or provide context of larger societal trends affecting health outcomes. For instance, nutrition datasets may include information on breast-feeding or use of nutrition programs by pregnant and lactating women. To provide another example, states may aggregate vital statistics and medical claims data, allowing for questions to be asked of the data beyond what is captured in vital statistics records. Other longitudinal datasets (eg, the Behavioral Risk Factor Surveillance System and the National Longitudinal Survey of Youth) have been used for perinatal or women's health analyses. Names and relevant content of non-perinatal databases with potentially useful information to maternal-child health researchers are provided in Table 3. Other databases may contain useful information, and researchers should assess their value against strictures surrounding access.

Non Health-Related Datasets

Non-health related data can also be used in maternal-child research. Significant quantities of data regarding specific locations and/or populations are collected or aggregated by governments and companies, and this information can provide context for perinatal health trends. One application of non-health data in perinatal research is the use of geographic information systems (GIS) to overlay various location-tagged data.[55] Using computer

software, researchers integrate census demographic information (eg, race, ethnicity or population density) with other spatially-linked data such as healthcare facility locations or perinatal outcomes by zip code.[55] Geographic information systems technology has been most commonly used in maternal-child health research to determine health service access and analysis of risk factor distribution.[55] This approach has also been used to assess the effect of toxins on perinatal outcomes[56] and guide population-based interventions.[55]

Even without sophisticated modeling, researchers or clinicians can use non-health related information to capture context. Census data and many other non-health-related datasets are de-identified and do not require IRB approval for use. However, combining datasets might permit identification of individuals and is controversial.[57] While full IRB approval may not be universally needed even when using multiple datasets, researchers should work with the IRB and relevant data-granting bodies to ensure participant protections.

Interactive Websites to Access Merged and De-Identified Datasets

While some research requires a full dataset for statistical manipulation, clinicians or those engaging in quality improvement may wish to query a database to obtain basic information or statistics. Several interactive websites facilitate data access (Table 4). Since the data are de-identified and only available at the state level, a data use agreement or IRB approval are not required, and there is no charge for data access. The CDC's PRAMStat was previously discussed. Two additional useful sites for perinatal statistics are Peristats and CDC WONDER (Wide-ranging Online Data for Epidemiologic Research). These websites merge multiple databases, enabling researchers to perform statistical queries and generate maps or charts. Another example of a website facilitating interactive use of multiple datasets is the Interactive Public Use Microdata Series which permits queries of census data from 1850 onward down to the level of the individual or family.[58]

Title V Information System (TVIS) also provides access to perinatal health data reported by the US states, territories, and jurisdictions receiving Title V funding.[59] This interactive website also includes information on the distribution of federal funding and each state's maternal-child health action plan. Data from 2000 onward can be used for website-based statistical queries and downloaded for analysis. However, definitions of variables have changed over time, which may prevent analysis with the entire dataset.[60]

DISCUSSION

Each of the aforementioned data resources has applications for perinatal research and quality improvement. Researchers conducting secondary analysis benefit from an understanding of the relative merits of each. Perhaps even more important for such research is an understanding of the overall utility, power, and limitations of secondary analysis of perinatal databases.

Limitations of Prospective Randomized Controlled Trials for Perinatal Research

Randomized controlled trials have long been promoted as the ideal method for systematically determining the effect of interventions on outcomes through distributing all potential confounding influences equally between groups.[61] Experimental research design

has distinct advantages and is the optimal design for many clinical questions; however, there are several attributes of prospective and randomized trials that may limit the utility of experimental design when conducting low-risk maternal-child health research. Maternity care experimental research design may not generate results appropriate for clinical translation and may prevent generation of knowledge relevant to vulnerable populations. One major criticism of clinical RCTs is that rigid treatment protocols differ from real-world application of processes, treatments, and medications. Thus, while a clinical RCT may produce extremely relevant information for well-women who are willing and able to precisely follow a protocol or whose pregnancy and labor course exactly match the study design, this information may be inappropriate or irrelevant for women who are unable or unwilling to follow the delineated protocol or whose perinatal events unfold differently.

Given the high degree of variability in women's preferences during pregnancy and birth as well as the wide variation in events of healthy pregnancy and birth, we propose that RCTs may not be the optimal study design for addressing certain questions about well-women during the childbearing cycle. Because well women frequently may not consent to, persist in, or remain eligible for RCTs, final sample sizes can be small and results can be distorted by selection bias due to differential dropout. This is especially problematic for research analyzing important morbidity and mortality outcomes as severe perinatal events are thankfully rare. Moreover, racial and ethnic minority women, women of low socioeconomic position, immigrant women, and women with less access to healthcare are under-represented in prospective research, limiting generalizability of findings.[62]

Further concern relates to selective inclusion and exclusion criteria for prospective or randomized trials that often exclude pregnant women from participation, resulting in a paucity of literature on how interventions or medications affect health in pregnancy.[63] Notably, pregnant women were unequivocally excluded from participation in NIH trials from 1977 to 1993. Despite recent loosening of these restrictions, few clinical drug trials include pregnant women due to fear of fetal harm.[64] Approval of prospective research involving pregnant participants often involves lengthy justifications, extremely-rigid protections, and intensive external oversight, making it difficult or impossible for researchers to conduct timely studies with immediate clinical relevance.[63]

Secondary research using existing datasets can analyze system or intervention effects on the health of pregnant women using data generated during actual medical care, public health surveillance, or administrative recordkeeping. For all of these reasons, the scientific gold standard of experimental research design may not be the optimal design for the conduct of well-woman perinatal outcomes research and quality assessment questions. With full awareness of the strengths and limitations of both experimental and observational study design, each perinatal research team can determine which approach is superior for addressing the specific question they seek to answer.

Causal Inference

There are a number of strategies to address limitations encountered when conducting research utilizing existing data. Causal inference approaches have applications in RCTs and are not exclusive to secondary data,[65, 66] but have received increasing attention as one

overarching framework and set of methods to strengthen retrospective design and secondary data analysis, to ensure that effects estimated from such studies are valid.[67] Because existing data often reflect actual care processes of women and clinicians (including their preferences for treatment options), it is likely that the differences between groups are not due solely to the effect of the intervention or treatment. Causal inference approaches vary in the specifics (with a lack of consensus as to what constitutes a "causal model"), but all invoke causal assumptions which, if satisfied, enable the researcher to attribute an association to causation rather than some competing explanation (eg, bias).[15, 68] These causal assumptions are distinct from the statistical assumptions required for unbiased estimation from a regression model.[69] By explicitly identifying and engaging with these assumptions (only some of which are testable), the investigator can minimize the chance that bias (eg, confounding bias, selection bias, over-adjustment bias) explains the calculated findings. Some approaches for causal inference utilize the counterfactual framework to improve estimation of the association between cause and effect;[70] such approaches have multi- disciplinary roots in philosophy, statistics, epidemiology, economics, and computer science.[15, 65, 68, 70]

Causal inference approaches provide the tools and theory to move beyond the dictum that 'correlation does not equal causation,' and determine under which circumstances a causal association may underlie correlation. An explicit causal framework formalizes the researchers' knowledge of relationships between exposure, outcome, confounding versus mediating variables, selection variables, and in recent years, missingness and measurement error as well.[71, 72] Then, the investigator may use existing data and appropriate methods to estimate a causal quantity of interest.

Some approaches (eg, propensity score matching) generate a control or comparison group to provide information about what might have happened to individuals had they not been exposed to the intervention.[73] An example of this approach is the use of propensity score analysis to enhance balance between samples of women who chose group or individual prenatal care.[74, 75] Other approaches (such as instrumental variables and regression discontinuity analyses) exploit exogenous changes in the exposure variable (ie, changes that are unrelated to confounding variables or any other variable in the data system) to explore how exposure affects the outcome under study.[76, 77] What these approaches and others share is that they explicitly state the investigator assumptions to infer that a calculated association represents a causal effect, rather than a spurious association resulting from bias. More thorough exploration of causal inference is beyond the scope of this paper and has been addressed outside of the maternal child outcomes literature.[15, 70] Application of the causal inference framework and methods to well-woman childbearing science will be an important area for future consideration.

CONCLUSION

Secondary analysis of existing datasets has many advantages for scientific and quality improvement activities. Randomized controlled trials involving pregnant and childbearing-age women are challenging to conduct, and RCT results may not accurately reflect clinical outcomes with more diverse women or treatment styles nor capture rare perinatal outcomes. There is a wide variety of publically available datasets for study of US perinatal outcomes.

While research using existing datasets must address important limitations, these datasets, coupled with appropriate analytical methodology, can be used to feasibly and ethically generate adequately-powered, clinically-relevant, and immediately-translational maternal-child healthcare outcomes science. The causal inference framework for research design and analysis includes approaches that strengthen secondary data analysis research. Increased utilization of the multiple existing maternal child health data sources, paired with use of appropriate analytical techniques, holds promise for accelerating well-woman perinatal outcomes investigation and strengthening the evidence base for perinatal care of low-risk women and their children.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

During manuscript production, Dr. Julia C. Phillippi received support from a Vanderbilt University Medical Center Faculty Research Scholars Award, and was supported by grant number K08HS024733 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality.

Dr. Ellen L. Tilden receives support from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and National Institutes of Health Office of Research on Women's Health, Oregon BIRCWH Scholars in Women's Health Research Across the Lifespan (K12HD043488-14)

Jonathan M. Snowden and Frances M. Biel are supported by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (grant number R00 HD079658-03).

References

1. Fullerton J, Sipe TA, Hastings-Tolsma M, et al. The Midwifery Workforce: ACNM 2012 and AMCB 2013 Core Data. Journal of Midwifery & Women's Health. 2015; 60(6):751–761.
2. American College of Nurse-Midwives. Core Competencies for Basic Midwifery Practice. Silver Spring, MD: American College of Nurse-Midwives; 2012. http://www.midwife.org/ACNM/files/ACNMLibraryData/UPLOADFILENAME/000000000050/Core%20Comptencies%20Dec%202012.pdf. Accessed September 30, 2016
3. Schuiling KD, Sipe TA, Fullerton J. Findings from the analysis of the American College of Nurse-Midwives' membership surveys: 2000–2003. Journal of Midwifery & Women's Health. 2005; 50(1):8–15.
4. Schuiling KD, Sipe TA, Fullerton J. Findings from the analysis of the American College of Nurse-Midwives' membership surveys: 2006–2008. J Midwifery Womens Health. 2010; 55(4):299–307. [PubMed: 20630356]
5. MacDorman MF, Reddy UM, Silver RM. Trends in stillbirth by gestational age in the United States, 2006–2012. Obstetrics & Gynecology. 2015; 126(6):1146–1150. [PubMed: 26551188]
6. Osterman MJ, Kochanek KD, MacDorman MF, Strobino DM, Guyer B. Annual summary of vital statistics: 2012–2013. Pediatrics. 2015; 135(6):1115–1125. [PubMed: 25941306]
7. Miller S, Abalos E, Chamillard M, et al. Beyond too little, too late and too much, too soon: a pathway towards evidence-based, respectful maternity care worldwide. The Lancet. 2016; 388(10056):2176–2192.
8. MacDorman MF, Declercq E, Cabral H, Morton C. Recent increases in the US maternal mortality rate: disentangling trends from measurement issues. Obstetrics & Gynecology. 2016; 128(3):447–455. [PubMed: 27500333]

9. Kennedy HP, Yoshida S, Costello A, et al. Asking different questions: research priorities to improve the quality of care for every woman, every child. The Lancet Global Health. 2016; 4(11):e777–e779. [PubMed: 27663682]
10. Graham W, Woodd S, Byass P, et al. Diversity and divergence: the dynamic burden of poor maternal health. The Lancet. 2016; 388(10056):2164–2175.
11. Agency for Healthcare Research and Quality. 2014 National Healthcare Quality and Disparities Report. Rockville, MD: Agency for Healthcare Research and Quality US Department of Health and Human Services; 2015. http://www.ahrq.gov/research/findings/nhqrdr/nhqdr14/index.html. Report No.: AHRQ 15-0007. Accessed September 30, 2016
12. American College of Obstetricians and Gynecologists' Task Force on Collaborative Practice. Collaboration in practice: implementing team-based care: Report of the American College of Obstetricians and Gynecologists' task force on collaborative practice. Obstet Gynecol. 2016; 127(3):612–617. [PubMed: 26901331]
13. American College of Nurse-Midwives, Midwives Alliance of North America, National Association of Certified Professional Midwives. Supporting Healthy and Normal Physiologic Childbirth: A Consensus Statement by ACNM, MANA, and NACPM. 2012. http://midwife.org/ACNM/files/ACNMLibraryData/UPLOADFILENAME/000000000272/Physiological%20Birth%20Consensus%20Statement-%20FINAL%20May%2018%202012%20FINAL.pdf. Accessed September 30, 2016
14. Victora CG, Habicht J-P, Bryce J. Evidence-based public health: moving beyond randomized trials. Am J Public Health. 2004; 94(3):400–405. [PubMed: 14998803]
15. Imbens, GW, Rubin, DB. Causal Inference in Statistics, Social, and Biomedical Sciences. New York: Cambridge University Press; 2015.
16. Eunice Kennedy Shriver National Institute of Child Health and Human Development. NICHD DASH. 2016. https://dash.nichd.nih.gov/. Accessed September 13, 2016
17. Rothman, KJ, Lash, TL, Greenland, S. Modern Epidemiology. 3rd. Philadelphia, PA: Wolters Kluwer; 2013.
18. Boslaugh, S. Secondary Data Sources for Public Health: A Practical Guide. New York: Cambridge University Press; 2007.
19. Agency for Healthcare Research and Quality. Agency for Healthcare Research and Quality: A Profile | Agency for Healthcare Research & Quality. 2016. http://www.ahrq.gov/cpi/about/profile/index.html. Accessed September 13, 2016
20. Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project (HCUP) | Agency for Healthcare Research & Quality. 2016. http://www.ahrq.gov/research/data/hcup/index.html. Accessed September 13, 2016
21. Kozhimannil KB, Law MR, Virnig BA. Cesarean delivery rates vary tenfold among US hospitals; reducing variation may address quality and cost issues. Health Aff (Millwood). 2013; 32(3):527–535. [PubMed: 23459732]
22. Stapleton SR. Validation of an online data registry for midwifery practices: a pilot project. Journal of Midwifery & Women's Health. 2011; 56(5):452–460.
23. Riddell CA, Kaufman JS, Hutcheon JA, et al. Effect of Uterine Rupture on a Hospital's Future Rate of Vaginal Birth After Cesarean Delivery. Obstetrics & Gynecology. 2014; 124(6):1175–1181. [PubMed: 25415169]
24. Stapleton SR, Osborne C, Illuzzi J. Outcomes of care in birth centers: demonstration of a durable model. Journal of Midwifery & Women's Health. 2013; 58(1):3–14.
25. Thornton P, McFarlin BL, Park C, et al. Cesarean outcomes in US birth centers and collaborating hospitals: a cohort comparison. Journal of Midwifery & Women's Health. 2017; 62(1):40–48.
26. American Association of Birth Centers. AABC Perinatal Data Registry. American Association of Birth Centers; 2015. http://www.birthcenters.org/?PDR. Accessed September 9, 2016
27. Lahr MB, Rosenberg KD, Lapidus JA. Bedsharing and maternal smoking in a population-based survey of new mothers. Pediatrics. 2005; 116(4):e530–e542. [PubMed: 16199682]
28. Liu J, Rosenberg KD, Sandoval AP. Breastfeeding duration and perinatal cigarette smoking in a population-based cohort. Am J Public Health. 2006; 96(2):309–314. [PubMed: 16380564]

29. Ahluwalia IB, Helms K, Morrow B. Assessing the validity and reliability of three indicators self-reported on the pregnancy risk assessment monitoring system survey. Public Health Reports. 2013; 128(6):527–536. [PubMed: 24179264]
30. Centers for Disease Control and Prevention. Pregnancy Risk Assessment Monitoring System. 2016. https://www.cdc.gov/prams/ Last updated July 5, 2016. Accessed November 4, 2016
31. Centers for Disease Control and Prevention. Vital Statistics Data Available Online. Centers for Disease Control and Prevention; 2016. http://www.cdc.gov/nchs/data_access/vitalstatsonline.htm. Accessed September 16, 2016
32. Hamilton BE, Martin JA, Osterman M, Curtin S, Matthews T. Births: final data for 2014. National Vital Statistics Reports. 2015; 64(12):1–64.
33. Centers for Medicare & Medicaid Services. Research, Statistics, Data & Systems. 2015. https://www.cms.gov/research-statistics-data-and-systems/research-statistics-data-and-systems.html Last updated December 21, 2015. Accessed September 13, 2016
34. Gallagher BK. Opioid Prescriptions Among Women of Reproductive Age Enrolled in Medicaid—New York, 2008–2013. MMWR Morb Mortal Wkly Rep. 2016; 65
35. Centers for Medicare and Medicaid Services Research Data Assistance Center. Data Request for Student Research | ResDAC. 2016. http://www.resdac.org/data-request-student-research. Accessed September 13, 2016
36. Centers for Medicare and Medicaid Services Research Data Assistance Center. About ResDAC | ResDAC. 2016. http://www.resdac.org/about-resdac. Accessed September 30, 2016
37. Eunice Kennedy Shriver National Institute of Child Health and Human Development. NICHD DASH – Tutorial. 2016. https://dash.nichd.nih.gov/Resource/Tutorial. Accessed September 30, 2016
38. Eunice Kennedy Shriver National Institute of Child Health and Human Development. Community Child Health Research Network. 2016. https://www.nichd.nih.gov/research/supported/Pages/cchn.aspx. Accessed February 28, 2017
39. Dolbier CL, Rush TE, Sahadeo LS, et al. Relationships of race and socioeconomic status to postpartum depressive symptoms in rural African American and Non-Hispanic White women. Matern Child Health J. 2013; 17(7):1277–1287. [PubMed: 22961387]
40. Guardino CM, Schetter CD, Hobel CJ, et al. Chronic Stress and C-Reactive Protein in Mothers During the First Postpartum Year. Psychosomatic medicine. 2017
41. Saxbe DE, Schetter CD, Guardino CM, et al. Sleep Quality Predicts Persistence of Parental Postpartum Depressive Symptoms and Transmission of Depressive Symptoms from Mothers to Fathers. Annals of Behavioral Medicine. 2016; 50(6):862–875. [PubMed: 27492636]
42. Eunice Kennedy Shriver National Institute of Child Health and Human Development. Home Page – Consortium on Safe Labor. 2016. https://csl.nichd.nih.gov/. Accessed September 30, 2016
43. Eunice Kennedy Shriver National Institute of Child Health and Human Development. NICHD DASH Consortium on Safe Labor (CSL) Study Page. 2016. https://dash.nichd.nih.gov/Study/Study?id=2331. Accessed September 13, 2016
44. Zhang J, Landy HJ, Branch DW, et al. Contemporary patterns of spontaneous labor with normal neonatal outcomes. Obstet Gynecol. 2010; 116(6):1281. [PubMed: 21099592]
45. Zhang J, Troendle J, Reddy UM, et al. Contemporary cesarean delivery practice in the United States. Am J Obstet Gynecol. 2010; 203(4):326.e321–326.e310. [PubMed: 20708166]
46. Laughon SK, Zhang J, Grewal J, et al. Induction of labor in a contemporary obstetric cohort. Am J Obstet Gynecol. 2012; 206(6):486.e481–486.e489. [PubMed: 22520652]
47. Laughon SK, Berghella V, Reddy UM, et al. Neonatal and maternal outcomes with prolonged second stage of labor. Obstet Gynecol. 2014; 124(1):57. [PubMed: 24901265]
48. Jou J, Kozhimannil KB, Johnson PJ, Sakala C. Patient-perceived pressure from clinicians for labor induction and cesarean delivery: a population-based survey of US women. Health Services Research. 2015; 50(4):961–981. [PubMed: 25250981]
49. Attanasio L, Kozhimannil KB. Patient-reported communication quality and perceived discrimination in maternity care. Medical Care. 2015; 53(10):863–871. [PubMed: 26340663]
50. Odum Institute for Research in Social Science. UNC Dataverse v. 4.5 build 41-dc58ae1. 2016. https://dataverse.unc.edu/. Accessed November 4, 2016

51. Simon RM, Johnson KM, Liddell J. Amount, Source, and Quality of Support as Predictors of Women's Birth Evaluations. Birth. 2016; 43(3):226–232. [PubMed: 26991407]
52. Cheyney M, Bovbjerg M, Everson C, et al. Development and validation of a national data registry for midwife-led births: the Midwives Alliance of North America Statistics Project 2.0 dataset. Journal of Midwifery & Women's Health. 2014; 59(1):8–16.
53. Cheyney M, Bovbjerg M, Everson C, et al. Outcomes of care for 16,924 planned home births in the United States: The Midwives Alliance of North America Statistics Project, 2004 to 2009. Journal of Midwifery & Women's Health. 2014; 59(1):17–27.
54. Centers for Medicare & Medicaid Services. Strong Start for Mothers and Newborns Initiative: General Information | Center for Medicare & Medicaid Innovation. 2016. https://innovation.cms.gov/initiatives/strong-start/ Last updated May 18, 2016. Accessed September 30, 2016
55. Makanga PT, Schuurman N, von Dadelszen P, Firoz T. A scoping review of geographic information systems in maternal health. International Journal of Gynaecology & Obstetrics. 2016; 134(1):13–17. [PubMed: 27126906]
56. Blake SB. Spatial relationships among dairy farms, drinking water quality, and maternal-child health outcomes in the San Joaquin Valley. Public Health Nursing. 2014; 31(6):492–499. [PubMed: 25412694]
57. Gray EA, Thorpe JH. Comparative effectiveness research and big data: balancing potential with legal and ethical considerations. Journal of comparative effectiveness research. 2015; 4(1):61–74. [PubMed: 25565069]
58. Ruggles, S, Genadek, K, Goeken, R, Grover, J, Sobek, M. Integrated Public Use Microdata Series: Version 6.0. Minneapolis: University of Minnesota; 2015. Last updated February 16, 2017
59. US Department of Health & Human Services. Title V Information System (TVIS). US Department of Health & Human Services; 2016. https://mchb.tvisdata.hrsa.gov/. Accessed September 12, 2016
60. US Department of Health and Human Services. DataSpeak: Utilizing the Title V Information System Data and the Federally Available Data Resource Document. HRSA DataSpeak Webinar Series. 2016.
61. Bothwell LE, Greene JA, Podolsky SH, Jones DS. Assessing the gold standard—lessons from the history of RCTs. N Engl J Med. 2016; 374(22):2175–2181. [PubMed: 27248626]
62. Fisher JA, Kalbaugh CA. Challenging assumptions about minority participation in US clinical research. Am J Public Health. 2011; 101(12):2217–2222. [PubMed: 22021285]
63. Blehar MC, Spong C, Grady C, et al. Enrolling pregnant women: issues in clinical research. Women's Health Issues. 2013; 23(1):e39–e45. [PubMed: 23312713]
64. White A. Accelerating the paradigm shift toward inclusion of pregnant women in drug research: ethical and regulatory considerations. Seminars in perinatology. 2015; 39(7):537–540. [PubMed: 26385413]
65. Balzer LB, van der Laan MJ, Petersen ML. Adaptive pre-specification in randomized trials with and without pair-matching. Statistics in medicine. 2016; 35(25):4528–4545. [PubMed: 27436797]
66. Bellamy SL, Lin JY, Ten Have TR. An introduction to causal modeling in clinical trials. Clinical trials (London, England). 2007; 4(1):58–73.
67. Brookhart MA, Sturmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. Medical Care. 2010; 48(6 Suppl):S114–120. [PubMed: 20473199]
68. Greenland S, Brumback B. An overview of relations among causal modelling methods. International journal of epidemiology. 2002; 31(5):1030–1037. [PubMed: 12435780]
69. Petersen ML, van der Laan MJ. Causal models and learning from data: integrating causal modeling and statistical estimation. Epidemiology. 2014; 25(3):418–426. [PubMed: 24713881]
70. Morgan, SL, Winship, C. Counterfactuals and Causal Inference. New York: Cambridge University Press; 2015.
71. Daniel RM, Kenward MG, Cousens SN, De Stavola BL. Using causal diagrams to guide analysis in missing data problems. Statistical methods in medical research. 2012; 21(3):243–256. [PubMed: 21389091]

72. Edwards JK, Cole SR, Westreich D. All your data are always missing: incorporating bias due to measurement error into the potential outcomes framework. International journal of epidemiology. 2015; 44(4):1452–1459. [PubMed: 25921223]

73. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983; 70(1):41–55.

74. Tilden EL, Emeis CL, Caughey AB, et al. The Influence of Group Versus Individual Prenatal Care on Phase of Labor at Hospital Admission. Journal of Midwifery & Women's Health. 2016; 61(4): 427–434.

75. Carlson NS, Corwin EJ, Lowe NK. Labor Intervention and Outcomes in Women Who Are Nulliparous and Obese: Comparison of Nurse-Midwife to Obstetrician Intrapartum Care. Journal of Midwifery & Women's Health. 2017; 62(1):29–39.

76. Greenland S. An introduction to instrumental variables for epidemiologists. International journal of epidemiology. 2000; 29(4):722–729. [PubMed: 10922351]

77. Thistlethwaite DL, Campbell DT. Regression-discontinuity analysis: an alternative to the ex post facto experiment. Journal of Educational psychology. 1960; 51(6):309.

QUICK POINTS

- Existing datasets enable analyses regarding healthcare service system outcomes, treatment effects, and rare outcomes.
- Datasets aggregate information from prior research or routine medical care and administrative processes. These large datasets are available from local, state, and federal agencies, professional organizations, as well as medical institutions.
- There are a variety of existing datasets for perinatal research; each has unique data and limitations. Researchers should become familiar with a dataset's provenance, contents, validity, and completeness of data elements prior to use.
- Research using such datasets could accelerate well-woman perinatal outcomes investigation due, in part, to large sample sizes and inclusion of diverse participants.

Table 1

Common Types of Research using Existing Datasets

Research	Definition	Examples
Prospective study designs	Participants are enrolled prior to an event or exposure of interest and followed over time to determine effects or outcomes of the event or exposure.[17]	Clinical trials, data registries, and studies that follow individuals over time
Retrospective study designs	Exposure and covariate data are collected after the outcome has occurred (ie, the investigator is looking into the past to ascertain exposure status after collection of outcome data, rather than recruiting participants who are followed forward through time, as in a prospective study design).[17]	Case-control studies, retrospective cohort studies, studies using vital statistics from birth records
Secondary data analysis	Research using existing data to answer question(s) that are different from the original purposes or questions that motivated data collection.[18] Secondary data analysis may be conducted using prospectively-collected data or data not collected for research.	Analysis of data from a previous cohort study or randomized trial, or claims data analysis
Administrative data research	Studies analyzing data collected for administrative purposes rather than explicitly for research. Administrative data are collected routinely for legal, compliance, or official government or institutional purposes. Research using administrative data is a type of secondary data research.	Studies involving regulatory compliance measurement (eg, air quality measures), census data, information from vital records

Table 2

Publically Available Databases for Women's and Perinatal Health Services Research, Quality, and Benchmarking

Database Name	Content	Dates of data collection	Approval(s) needed to access data	Cost
Agency for Healthcare Research & Quality (AHRQ)	Hospital data (Healthcare Cost and Utilization Project, HCUP), individual healthcare expenditure data (Medical Expenditure Panel Survey), and state healthcare expense data (State Snapshots), as well as others.	Example: HCUP data on national inpatient services available for 1988–2011. Other datasets from AHRQ are available; dates vary by dataset.	Online and signed data use agreements and training course are required.	$50–150/ dataset for students $350–500 for researchers
American Association of Birth Centers (AABC) Perinatal Data Registry	The registry, originally designed for birth centers, now includes data from diverse perinatal settings and care providers. Women are enrolled at the first prenatal visit, and data entered at key time points in perinatal care, resulting in prospective enrollment ahead of perinatal outcomes.	Data from 2007- forward is available.	Scientific data request form must be submitted along with IRB approval. All abstracts, manuscripts, and publications must be approved by the AABC Research Committee and Board of Directors prior to submission.	Dependent on the amount of data and AABC membership status
Centers for Medicare & Medicaid Services	Enrollment and utilization data for Medicaid and Medicare participants from across the US available. Participant-level and facility-level data available.	Small number of states have data from 1980–present for Medicaid. Participation of all states mandated from 1999–present for Medicaid.	Contact the Research Data Assistance Center (ResDAC), the organization that assists researchers to discuss study plans and type of files needed. A fee and a data use agreement are required.	Up to $2000
Center for Disease Control and Prevention				
Pregnancy Risk Assessment Monitoring System	State-specific data on maternal attitudes and experiences before, during, and shortly after pregnancy obtained through English and Spanish surveys sent to a sample of women in each state.	Yearly data since 1987 Questionnaire versions: Phase 1 (1988–1989), Phase 2 (1990–1995), Phase 3 (1996–1999), Phase 4 (2000–2003), Phase 5 (2004–2008), Phase 6 (2009–2011), Phase 7 (2012–2015), Phase 8 (2016–present). Codebook defines variables.	Data sharing agreement is required. Researcher contact information, project abstract and data requested (states and years) should be submitted. Requests are approved by committee that meets monthly.	No Cost
US Vital Statistics Data	Birth, fetal death, and infant death data. Birth data cover 99% of all domestic births.	Annually; data type affects years available.	Check with internal IRB; generally non-human subjects research and exempt from IRB oversight; No permissions needed from CDC NCHS for data download and use.	Free
Eunice Kennedy Shriver National Institute of Child Health and Human Development Data and Specimen Hub				
Community Child Health Network	De-identified maternal and child health data from a multi-site, prospective cohort study of postpartum women and their spouses (n = 4,837).	2004–2009	Data use agreement signed by requestor and authorized representative of requestor's organization as well as a brief description of proposed secondary analysis should be submitted. The proposal is then approved by NICHD DASH data committee and/or study-specific approval entity.	No cost
Consortium on Safe Labor	Detailed, de-identified information from medical records of 228,562 women	2002–2008	Data use agreement signed by requestor and authorized representative of requestor's	No cost

Database Name	Content	Dates of data collection	Approval(s) needed to access data	Cost
	birthing in one of 19 US academic medical centers		organization as well as a brief description of proposed secondary analysis should be submitted. The proposal is then approved by NICHD DASH data committee and/or study-specific approval entity.	
Maternal-Fetal Medicine Unit Network: Screening for Risk Factors for Spontaneous Preterm Delivery	De-identified demographic, behavioral, psychological, anthropometric, biologic, ultrasound and physical examination data from a multi-site, observational study of 3,073 women with singleton pregnancies.	1992–1995	Data use agreement signed by requestor and authorized representative of requestor's organization as well as a brief description of proposed secondary analysis should be submitted. The proposal is then approved by NICHD DASH data committee and/or study-specific approval entity.	No cost
Listening to Mothers	National survey of women in the US who gave birth to a singleton baby. Includes information on pregnancy planning, care providers, labor support, labor onset, perinatal medical interventions, breastfeeding, feelings about care during labor and birth, treatment by caregivers, description of labor and birth experiences, general postpartum health, and emotional health after birth.	Listening to Mothers (2002; N = 1,583) Listening to Mothers II (2006; N = 1,573), Listening to Mothers III (2012; N = 2,400) Listening to Mothers II: New Mothers Speak Out (2006; n = 903), Listening to Mothers III: New Mothers Speak Out (2013; n = 1,072)	Check with internal IRB; generally non-human subjects research and exempt from IRB oversight. No permissions for data download from the Odum Institute's Dataverse.	No cost
Midwives Alliance of North America Statistics Project (MANA Stats)	Datasets include 3 iterations: 2.0, 3.0, and the current 4.0, each with different variables. Midwives enroll patients early in prenatal care and enter data as perinatal events unfold; thus, data are prospectively collected. Once entered, records cannot be deleted. Most data are from planned homebirth or birth center care; ~12% of study participants have intrapartum hospital transfer and data include outcomes after transfer.	2.0 (2004–2009; N=24,000), 3.0 (2009–2011; N=15,000), 4.0 (2011–current; N=61,000) with approximately 1100 added each month	First, propose research question to MANA Research Review Committee. If approved, complete MANA application includes: description of project and research questions, methods/procedures, risks/benefits assessment, dissemination plan, timeline, funding summary, references, list of variables requested, CVs and documentation of human research ethics training for all investigators, and IRB approval.	Students-$100 Faculty at large research institutions-$1000 Faculty at smaller research institutions-$250
Strong Start for Mothers and Newborns Initiative	Participant and program-level data on Strong Start initiative to improve Medicaid and CHIP maternal and infant outcomes through enhanced prenatal care. Participant-level data includes maternal demographics, characteristics and outcomes of pregnancy, costs, and maternal satisfaction.	2013–2018 (anticipated)	Currently, only aggregate data are available for first 2 years of study. Full study data anticipated in 2018.	No cost

Abbreviations: AABC, American Association of Birth Centers; AHRQ, Agency for Healthcare research & Quality; CHIP, Children's Health Insurance Program; DASH, Data and Specimen Hub; HCUP, Healthcare Cost and Utilization Project; MANA, Midwives Alliance of North America; NICHD, Eunice Kennedy Shriver National Institute of Child Health and Human Development; ResDac, Research Data Assistance Center

Table 3

Health-Related Databases Useful in Women's and Perinatal Health Services Research, Quality, and Benchmarking

Name	Content
Behavioral Risk Factor Surveillance System	State-based surveillance of preventative services use, health-related risk behaviors, and chronic conditions. This dataset includes information on women's preventative, reproductive and obstetric healthcare. Information, including surveys and interviews, is collected annually in all 50 states. Data are available from 1984 forward, making it one of the largest continuously-collected health datasets.
National Longitudinal Surveys from the Bureau of Labor Statistics	Datasets from longitudinal studies using national samples are available from the Bureau of Labor Statistics. These datasets include information from participants at several points to allow for study of individuals across time. For example, the National Longitudinal Survey of Youth enrolled two cohorts, one in 1979 and one in 1997, and continues to contact participants to measure multiple variables at regular intervals.
US Census Bureau	Every 10 years, there is a national census of the US population, including characteristics of individuals and households, and aggregate data is made publically available. The Census Bureau also conducts surveys on economic, workforce, income, and insurance coverage at a range of times and allow public access to de-identified data. Micro-level is available for researchers following approval. This information can provide valuable context to understand perinatal outcomes in relationship to local, regional, or national demographic trends.

Table 4

Interactive Websites to Obtain Perinatal Statistics from Datasets

Name	Data Sources	Website
Centers for Disease Control & Prevention's Wide-ranging Online Data for Epidemiologic Research (CDC WONDER)	The WONDER website provides access to over 53 databases on diverse topics. Birth and linked birth/infant death data are drawn from birth certificates and available in three discrete datasets related to changes in race categories. Sexually-transmitted infections data are from case reports provided as part of local, state and national-level sexually-transmitted infection programs in the United States, Puerto Rico, Guam, and the Virgin Islands. Information is also available on cancer, tuberculosis, adverse vaccine events, assisted reproductive technology, environmental statistics, and population projections. WONDER also provides access to other databases.	www.wonder.cdc.gov
Peristats from the March of Dimes	Combined data from March of Dimes survey and 12 US government agencies and organizations	www.marchofdimes.org/peristats
Pregnancy Risk Assessment Monitoring System – PRAMStat	Data from 2000 forward on over 250 perinatal indicators collected by the PRAMS surveys.	www.cdc.gov/prams/pramstat
Title V Information System (TVIS)	Provides data about the funding and effects of the Title V Maternal and Child Health Services Block Grant Program at national, state, and regional levels using data from grantees and national databases.	http://mchb.tvisdata.hrsa.gov

Abbreviations: CDC WONDER, Centers for Disease Control & Prevention's Wide-ranging Online Data for Epidemiologic Research; PRAMStat, Pregnancy Risk Assessment Monitoring System; TVIS, Title V Information System