



# Linking household members and defining relational networks using routine health data

Jeffrey I. Campbell, Ana Poblacion and Richard Sheward

## Purpose of review

The growth of rich electronic health record (EHR) data and large health databases has introduced new opportunities to link individuals together into households and relational networks. These 'linked relational networks' hold promise for providing family-level care and studying intergenerational epidemiology and clinical outcomes. However, as linked relational networks become more commonly available in EHRs and research databases, it is critical to understand their challenges and limitations.

## Recent findings

Matching algorithms are being used to create linked relational networks in EHR and health databases. Clinically, these algorithms have been most useful to provide dyadic maternal–newborn care. In research, studies using these algorithms investigate topics ranging from the pharmacoepidemiology of parental drug exposure on childhood health outcomes, to heritability of chronic conditions, to associations between parental and child healthcare access and service delivery. However, ethical and technical challenges continue to limit use of these algorithms. There is also a critical research gap in the external validity of these matching algorithms.

## Summary

Linked relational networks are in widespread use in pediatric clinical care and research. More research is needed to understand the scope, limitations, and biases inherent in existing matching strategies.

## Keywords

caregiver, electronic health record, family, household, pharmacoepidemiology, relational network

## INTRODUCTION

Delivering family-centered care has long been a goal of clinicians and health organizations. As increasing volumes of health data have been electronically recorded, health information technology (HIT) strategies emerged to create linked databases. These linked databases achieve matching by using sets of unique identifier data to relate individuals or groups across multiple databases or clinical records.

An extension of database linking is the ability to connect records and data of one individual to the records and/or data of other individuals. The network of linked individuals – what we call a 'linked relational network' – can exist within electronic health records (EHR), claims databases, and epidemiologic datasets. Generating linked relational networks has introduced new opportunities and challenges in pediatric and family-centered care. As linking strategies become more prevalent and sophisticated, it is important for clinicians and researchers to understand what these tools are, the expanding ways in which they are used, and their limitations.

For this review, we adopt a definition of a linked relational network as the set of individuals who can be linked by shared identifying attributes that: are routinely collected for the provision of healthcare, can specify a unique network of two or more individuals, and that plausibly mirror established real-world relationships. Surveying seminal and recent literature, we highlight algorithms that identify linked relational networks, and scientific advances that these algorithms have enabled. We discuss how these algorithms have been and likely will be used in clinical care, and their uses in epidemiology and health policy research (Table 1). Finally, we discuss important ethical considerations surrounding the use of these algorithms.

Department of Pediatrics, Boston University Chobanian & Avedisian School of Medicine, Boston, Massachusetts, USA

Correspondence to Jeffrey I. Campbell, MD, MPH, 670 Albany Street, 6th floor, Boston, MA 02118, USA. Tel: +1 617 414 5160; fax: +1 617 414 7230; e-mail: jeffrey.campbell@bmc.org

**Curr Opin Pediatr** 2024, 36:000–000

DOI:10.1097/MOP.0000000000001415

KEY POINTS

- Matching algorithms are widely used to link individuals to each other in electronic health records, and in research and administrative databases, forming what we call ‘linked relational networks’.
- Linked relational networks enable novel opportunities for family-level care, and to understand disease epidemiology and health service delivery within interpersonal networks at a large scale.
- As new applications of these algorithms emerge, it will be important for clinicians and researchers to understand the current limitations of linking, and the potential ethical challenges that use of these algorithms create.
- More research is needed to demonstrate the external validity of these matching algorithms.

GENERATING LINKED RELATIONAL NETWORKS

Methods to link individuals within health records and databases can be categorized as deterministic – requiring exact data matches to establish links – and probabilistic – using a computed probability that fully matching or partially matching data represent a link (Fig. 1) [1]. To date, most linking algorithms have used deterministic matches. In clinical practice, the certainty and uniqueness of deterministic matching is critical so that accurate information can be correctly routed to clinicians. For example, in the EHR, deterministic matching ensures that neonates

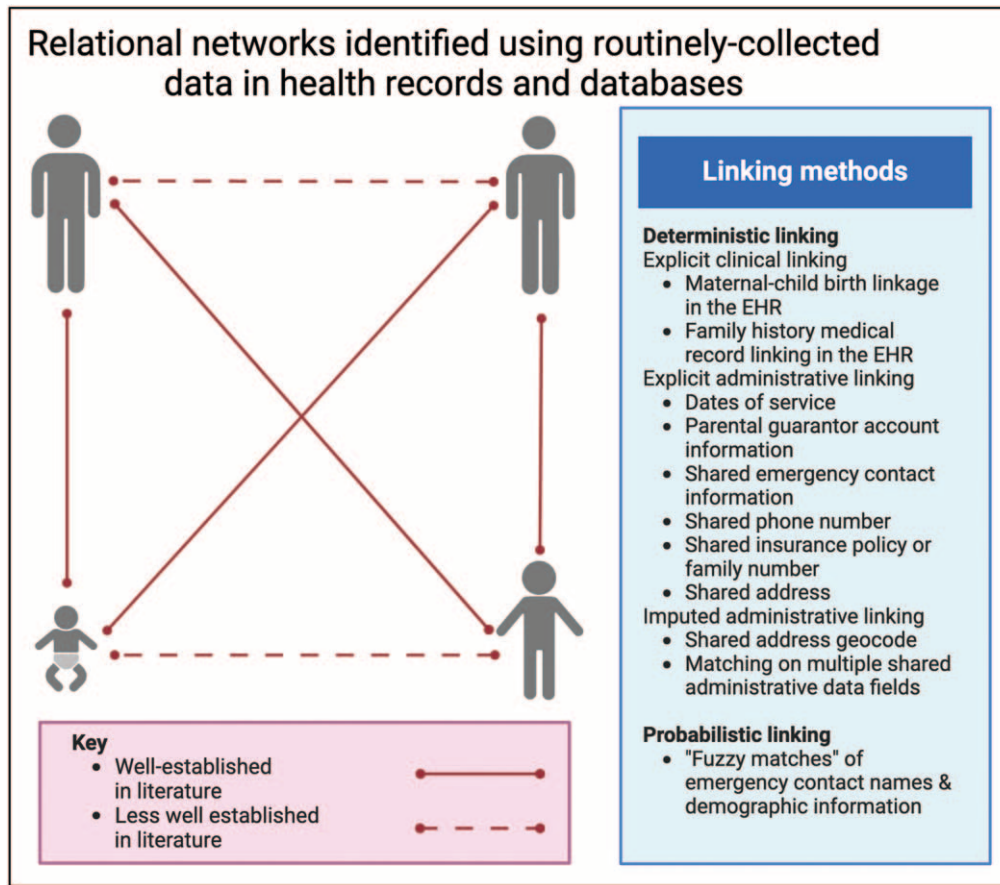
are accurately linked to their mothers to facilitate newborn care. In administrative and research databases, deterministic matching similarly ensures linking accuracy by using specific identifiers such as phone numbers, addresses [2,3], emergency contact, and/or payment guarantor information [4]. Within claims databases, these identifiers include shared insurance numbers and family identification numbers [2,4,5,6,7]. Computed data can also be deterministically matched to establish linked relational networks. For instance, address geocodes can be matched to ascertain linked relational networks based on residence. Recent literature has described how combinations of data elements (e.g. phone number plus address plus insurance information) can increase sensitivity of matching algorithms [2,8,9].

A limitation of deterministic matching is that relational ties between individuals in a database or clinical record may be overlooked because of discrepancies in data, such as misspellings of addresses or names. Thus, there has been increasing interest in using probabilistic matching algorithms to infer linked relational networks. A recent study of pregnancy-offspring outcomes conducted within the OCHIN Accelerating Data Value Across a National Community Health Center Network (ADVANCE) database leveraged free text data to establish probabilistic links between children’s and mothers’ records [9]. Here, the investigators matched free text emergency contact names within a database of candidate contacts, and added specificity by matching phone numbers, street addresses, and zip codes. Each step required only a partial data match, but the

**Table 1.** Applications of linked relational networks using health and administrative data in clinical care, policy research, and clinical and epidemiological research

|                                                                                                                                                 |
|-------------------------------------------------------------------------------------------------------------------------------------------------|
| Clinical care                                                                                                                                   |
| EHR maternal–neonatal dyad hyperlinks and data piping between clinical records enables newborn care                                             |
| Policy research                                                                                                                                 |
| Family insurance numbers enable evaluation of multiplier effect of enrolling individuals in insurance                                           |
| Clinical and epidemiology research                                                                                                              |
| Parent–child links enable evaluation of:                                                                                                        |
| Intergenerational prevalence and incidence of chronic conditions                                                                                |
| Association between parental preventive care and child well child care                                                                          |
| Evaluation of co-location of clinical care services                                                                                             |
| Associations between parent and child vaccine uptake                                                                                            |
| Effects of maternal and paternal antenatal medication exposure and neonatal/childhood outcomes                                                  |
| Multigenerational links among large family networks enable study of heritable conditions in clinical databases                                  |
| Family insurance identifiers facilitate study of effects of surgery and medication prescribing on spousal mental health and medication receipt  |
| Linkages between health and multigenerational demographic/administrative databases enables sibling comparison quasi-experimental cohort designs |

EHR, electronic health record.



**FIGURE 1.** Established and future relational network linkages using health information technology and data collected in health records and databases. Created in BioRender. Campbell, J. (2024) BioRender.com/k51o237.

overall strategy generated precise mother–child linkages that could be used for epidemiologic research.

The types of relationships that HIT-based linking strategies create are dependent upon available data. Algorithms using individuals' addresses will generate networks that mirror the classical definition of a 'household'. Meanwhile, links based on insurance guarantor or emergency contact phone numbers will resemble networks of social support, which do not need to be (and often are not) based on cohabitation.

## LINKED RELATIONAL NETWORKS IN CLINICAL CARE

To date, clinical applications of linking technologies have been primarily developed for maternal–neonatal care settings. Several large EHR systems contain perinatal care applications in which maternal and delivery data from maternal records are automatically piped to a neonate's record. Integrated clinical platforms that include linked maternal–neonatal records have shown to improve neonatal

care documentation and billing [10,11]. Meanwhile, the absence of linked maternal–neonatal records has been identified as a barrier to ensuring accurate postnatal diagnosis of congenital illnesses [12].

The routine linkage of maternal–neonatal records expands the amount and variety of data available for novel clinical tools in newborn care. For example, Gerard *et al.* [13] used linked maternal–infant records and machine learning to develop predictive models of neonatal hypoglycemia. In their analysis, an identifier was used to specify maternal–neonatal dyads, enabling incorporation of maternal medical history and comorbidity variables in candidate predictive models. These linked data will likely be valuable as machine learning, and similar methods are used to identify patterns in large clinical datasets that can be translated to the bedside.

Scant literature describes implementation of linkage techniques for clinical care beyond the neonatal period. Clinical tools that link caregiver–child dyads as well as siblings in the EHR have several potential applications. One potential application is a 'family health record', which would share clinical information of family members between multiple

clinicians and locations, thereby improving family-centered care [14]. In public health, linked relational networks could facilitate provision of infection control and prophylaxis to at risk household members after exposure to transmissible infectious diseases. Finally, if embedded widely in the EHR, linked relational networks might decrease patient and care navigator burden when households, not individuals, are screened for household-level social needs. Despite benefits, technical and ethical hurdles may hinder use of health data-defined relational networks beyond the newborn period, as it will be discussed in the following. To mitigate challenges, it will be important for patients and clinicians to provide input and ideally co-create clinically linked relational network tools to ensure acceptability.

### LINKED RELATIONAL NETWORKS IN POLICY RESEARCH

The use of family identifier numbers in claims databases has enabled evaluation of the effects of health policy interventions within linked relational networks. In a seminal study, DeVoe *et al.* used household identification numbers within the Oregon Health Plan (Oregon's Medicaid program) and emergency contact and guarantor linkages within the OCHIN EHR database [4] to generate a cohort of caregiver and linked children [15]. Then, investigators examined the association between caregiver insurance enrollment and subsequent enrollment of their children in the health plan, finding increased enrollment of children when their parents had been selected to enroll in the health plan. This finding indicates that policies targeting adult coverage can have ripple effects on children's access to healthcare, while illustrating how linked relational networks can be used to study health policy effects within families.

### INNOVATIVE USES OF LINKED RELATIONAL NETWORKS IN CLINICAL RESEARCH AND EPIDEMIOLOGY

The most common use of linked relational networks in research databases has been to study clinical and epidemiologic relationships within network members. Many of these studies use data elements intentionally designed to specify family units – in particular, family identification numbers within insurance databases. Other studies have used innovative linkages of data elements that were not initially designed to specify relational networks, such as emergency contact information, address geocodes, and insurance guarantor data.

### Research using unique, linked identification numbers in administrative databases

Studies have used established family identifiers in insurance and administrative databases to examine clinical outcomes and epidemiology within relational networks. For instance, several insurance databases include a 'family identifier number' that can be used to link family members who share insurance plans (notably, Medicaid removed this identifier number in 2014) [9<sup>\*\*\*</sup>]. The archetypical use of these unique identifiers are studies linking mothers to neonates [16–18]. Recent research has sought to expand the sample sizes of these studies to enable population-level assessments. A key example is a recent population-wide study on the effects of maternal disability on children's receipt of well child care using linked clinical data from dyads mother–child born in-hospital in the Ontario (Canada) MOMBABY dataset [19]. Likewise, strategies to create maternal–neonatal relational network links within the widely used Pediatric Health Information System database have recently been published [20].

Another key advance in this field has been the use of a broad array of relational ties that can be defined using family identification numbers, expanding linked relational networks beyond maternal–neonatal dyads. In a retrospective, claims-based analysis using the OptumLabs Data Warehouse, Meserve *et al.* [21] used an insurance family identifier number to link expectant fathers to neonates. The investigators examined whether paternal receipt of immunomodulators for immune-mediated inflammatory diseases was associated with adverse birth outcomes, finding no association. Other studies have linked adults to other adults [22,23], illustrating how family identifier numbers in insurance databases can be used to understand family-level clinical and epidemiologic outcomes.

Unique identifier numbers have also enabled creation of linked relational networks across multiple health-related, administrative, and demographic databases. In a systematic review of the association between paternal medication prescriptions and linked neonatal outcomes, Gaitonde *et al.* used national patient registries that enabled deterministic linkages with birth records and other administrative databases [24<sup>\*\*\*</sup>]. Similarly, Chen *et al.* [25<sup>\*\*\*</sup>] analyzed the relationship between experiencing a traumatic event and subsequent development of a psychiatric disorder, using a comparison cohort comprised of siblings who had not experienced a traumatic event. This sibling network enabled the investigators to control for potential confounding by unmeasurable familial factors. To create the sibling comparison cohort, the authors linked individual and family-level identifiers across the Swedish

Patient Register and Swedish Multi-Generation Register. This study illustrates how a sibling comparison quasi-experimental study design can be implemented at a population level when data in registries can be linked [26].

### Novel linkage strategies to define relational networks in clinical research and epidemiology

Although unique identifier numbers facilitate creation of relational networks across clinical and research databases, many databases do not contain such numbers. This data limitation has led investigators to explore the use of other deterministic and probabilistic linking methods, often using more than one identifier common to members of the linked relational network.

One of the first research databases to implement these linkages for clinical and epidemiologic studies was the OCHIN research database [4,8]. Within the database, investigators determined caregiver–child relational networks using common emergency contact records and/or guarantor fields of patients who were labeled in the relevant field as the children’s mother or father and who were 12–55 years older than the child to examine health service delivery within the specified networks [27,28<sup>■</sup>,29]. Notably, in one study in the OCHIN database, researchers were only able to link 33% of children with a parent [8]. Of these, 82% were linked to a mother only, 14% to a father only, and 4% to both. Most families consisted of only one linked child (61%). These statistics indicate potential limitations and gaps in using health record data to define relational networks that comprehensively reflect family or household membership.

Others have added data elements to matching algorithms to improve the yield of linkage generation. The ongoing PReventing Obesity through healthy Maternal gestational weight gain In the Safety nEt (PROMISE) Study uses a combination of deterministic and probabilistic matching to capture maternal–neonatal dyads, which it leverages to examine relationships between antenatal weight gain and neonatal outcomes in the OCHIN database [9<sup>■</sup>]. This study uses deterministic guarantor, emergency contact, geocode, and phone number matching, plus probabilistic ‘fuzzy matching’ of a variety of data elements to establish linkages. Similarly, to investigate the impact of prenatal drug exposure on neonatal and pediatric outcomes, Weaver *et al.* developed an algorithm using mother’s sex, age, pregnancy with a live birth, children’s age, child’s birthdate, insurance identifier number, and insurance coverage time to link mothers and neonates in

two large claims datasets [30<sup>■</sup>]. The use of multiple variables resulted in a high linkage rate. Linked mothers and infants comprised 73.6% of all mothers and 49.1% of all infants in these databases, respectively. Heerman *et al.* [31] used insurance (common plan identification), child’s identification (listed on maternal records), address (common geocoded location), home phone, and emergency contact information variables to link mother–child dyads while investigating the intergenerational impact of antibiotic treatment during pregnancy on child obesity at 5 years.

Although most relational networks in this literature have been used to study clinical outcomes and epidemiology, investigators in New York City leveraged multigenerational networks established with health record linking to study disease heritability and simulate large-scale observational genetics studies [32]. Investigators developed a method to create linked relational networks for a subset of patients within two large healthcare systems. This algorithm – named relationship inference from the electronic health record (RIFTEHR) – used emergency contact name, ZIP code, and phone number matching, along with relationships specified in emergency contact fields in the EHR. This method was used to infer 7.4 million relationships in this clinical database. This approach has been subsequently adapted to other healthcare systems to understand multigenerational patterns of obesity [33].

### ETHICAL CHALLENGES IN GENERATING AND USING LINKED RELATIONAL NETWORKS

The principal challenge that linkage strategies raise is potential breaches of privacy. In 2020, passage of the 21st Century Cures Act required health systems to provide patients with access to their EHR information [34]. When data from linked relational network members are piped between clinical charts, private health information may be unintentionally shared between patients. Lamar *et al.* [35<sup>■</sup>] report an instance in which a history of intimate partner violence, legal information, and a safety plan documented in a mother’s chart was automatically transferred to the chart of her newborn and thereby became visible to the newborn’s father, who was also the perpetrator. EHR systems may contain private information-blocking features, but in this case, these systems had not been activated. Other potential protections could include restricting access to relational network data to clinicians only. Several questions surrounding consent have also not been addressed because clinical applications to date have focused on maternal–newborn dyads. If clinical

applications expand to other relationships (e.g. connecting adults to other adults), should individuals who could be linked be informed and given the option to opt in or out? Should consent be reviewed periodically in case relationships have changed? Or should these linkages be automatic or covered under blanket privacy clauses signed by patients when they enroll in clinical care? Addressing these questions will be important to ensure ethical standards are maintained and that potential clinical tools are acceptable to patients.

When used in research, linked relational networks may exacerbate disparities in cases of incomplete or inaccurate data. English *et al.* developed a linkage algorithm based on insurance identification numbers, phone numbers, and addresses, and found that patients with lower socioeconomic status had lower chances of having a linkage [2]. This difference may have been because of higher absence of family identifier numbers in Medicaid insurance plans than commercial insurance, due to frequent changes in phone numbers, multiple moves, or due to residence in multiunit buildings where a shared addresses were flagged as false-positive linkages.

To date, the sensitivity and specificity of health data-linked relational networks has been validated using other health record data (e.g. using maternal-child dyadic links in an EHR to understand performance of multiple-variable linkage strategies). However, no external gold standard, such as direct patient interviews, has been developed to understand potential limitations or systemic errors in these algorithms. Performing external validations will be critical for understanding the scope and limitations of linked relational networks. External validation will also be important for uncovering systemic biases in these algorithms' abilities to define clinically and scientifically relevant networks. In pediatric care and research, it is important to perform these external validations to understand how well linked relational networks in health data map to real-world relationships that are important for child health and development, such as caregivers, parents, and siblings.

## CONCLUSION AND IMPLICATIONS FOR PRACTICE

Linked relational networks are widely used in pediatric and family-centered research and clinical care. They demonstrate how health information technology and data science have leveraged large and complex health databases, offering new opportunities for family-centered care, epidemiological research, and policy recommendations. As these technologies evolve, there is potential for broader applications in

family-centered care, such as streamlining social needs screening or enabling household-level infectious disease control. Linked relational also enable novel research designs, such as sibling comparison studies and multigenerational analyses of health outcomes. Key challenges include potential privacy concerns, questions surrounding consent, and lack of external validation of linkage algorithms. Those utilizing these networks should be aware of their strengths and limitations, and more research is needed to understand the scope, limitations, and biases inherent in existing matching strategies.

## Acknowledgements

None.

## Financial support and sponsorship

J.I.C. was supported by NIH/NIAID K23AI186597 and by the Boston University Clinical and Translational Sciences Institute (NIH/NCATS UL1TR001430).

## Conflicts of interest

There are no conflicts of interest.

## REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Cuccaro-Alamin S, Eastman AL, Foust R, *et al.* Strategies for constructing household and family units with linked administrative records. *Child Youth Serv Rev* 2021; 120:105706.
2. English TM, Kinney RL, Davis MJ, *et al.* Identification of relationships between patients through elements in a data warehouse using the Familial, Associational, and Incidental Relationship (FAIR) Initiative: a pilot study. *JMIR Med Inform* 2015; 3:e9.
3. Brady KJS, Legler A, Adams WG. Exploring opportunities for household-level chronic care management using linked electronic health records of adults and children: a retrospective cohort study. *Matern Child Health J* 2020; 24:829–836.
4. Angier H, Gold R, Crawford C, *et al.* Linkage methods for connecting children with parents in electronic health record and state public health insurance data. *Matern Child Health J* 2014; 18:2025–2033.
5. Bertioia ML, Phiri K, Clifford CR, *et al.* Identification of pregnancies and infants within a US commercial healthcare administrative claims database. *Pharmacoepidemiol Drug Saf* 2022; 31:863–874.
6. Chung MK, Hart B, Santillana M, Patel CJ. Pediatric and young adult household transmission of the initial waves of SARS-CoV-2 in the United States: Administrative Claims Study. *J Med Internet Res* 2024; 26:e44249.
7. Smith HC, Petersen I, Schartau P. Association of recent fatherhood with antidepressant treatment initiation among men in the United Kingdom. *JAMA Network Open* 2023; 6:e2316105.
8. Angier H, Giebultowicz S, Kaufmann J, *et al.* Creation of a linked cohort of children and their parents in a large, national electronic health record dataset. *Medicine (Baltimore)* 2021; 100:e26950.
9. Boone-Heinonen J, Lyon-Scott K, Springer R, *et al.* Pregnancy health in a multistate U.S. population of systemically underserved patients and their children: PROMISE cohort design and baseline characteristics. *BMC Public Health* 2024; 24:886.

This preliminary analysis of the PROMISE study describes implementation of a multiple-variable linking strategy within the OCHIN database. This strategy enables robust linking of mothers and children to study effects of antenatal weight gain on neonatal outcomes.

10. Liu W, Walsh T. The impact of implementation of a clinically integrated problem-based neonatal electronic health record on documentation metrics, provider satisfaction, and hospital reimbursement: a quality improvement project. *JMIR Med Inform* 2018; 6:e40.
  11. Rehman Durrani NU, Cowstill C, Krishnappa SB. Improving the quality and timeliness of neonatal intensive care unit discharge note: a quality improvement project. *J Clin Neonatol* 2022; 11:124–135.
  12. Combs CA, Hameed AB, Friedman AM, Hoskins IA. Special statement: proposed quality metrics to assess accuracy of prenatal detection of congenital heart defects. *Am J Obstet Gynecol* 2020; 222:B2–B9.
  13. Gerard J, Stuebe A, Sweeney A, *et al.* Using machine learning to develop a predictive model of infant hypoglycemia based on maternal and infant variables in an electronic health record. *Proceedings of the International Symposium on Human Factors and Ergonomics in Healthcare* 2023; 12:94–100.
- In this study, the authors used machine learning to examine the role of multiple infant and maternal variables to predict neonatal hypoglycemia, leveraging health record maternal-neonatal network links.
14. Angier H, Garvey B, DeVoe JE. Focus on families to improve child health during the COVID-19 pandemic and beyond. *JAMA Health Forum* 2021; 2: e210238.
  15. DeVoe JE, Marino M, Angier H, *et al.* Effect of expanding Medicaid for parents on children's health insurance coverage: lessons from the Oregon experiment. *JAMA Pediatr* 2015; 169:e143145.
  16. Figueroa R. Use of antidepressants during pregnancy and risk of attention-deficit/hyperactivity disorder in the offspring. *J Dev Behav Pediatr* 2010; 31:641–648.
  17. Whitley R, Davis EA, Suppappanya N. Incidence of neonatal herpes simplex virus infections in a managed-care population. *Sex Transm Dis* 2007; 34:704–708.
  18. Chang S, Begier EM, Schech SD, *et al.* Perinatal hepatitis B transmission and vaccination timing in a managed care cohort: assessment of the temporary delay in newborn hepatitis B vaccination due to thimerosal content. *Pediatr Infect Dis J* 2007; 26:329–333.
  19. Brown HK, Taylor C, Camden A, *et al.* Maternal disability and early child preventive care. *Pediatrics* 2024; 154:e2024066069.
  20. Hahn PD, Melvin P, Graham DA, Milliren CE. A methodology to create mother-baby dyads using data from the pediatric health information system. *Hosp Pediatr* 2022; 12:884–892.
  21. Meserve J, Luo J, Zhu W, *et al.* Paternal exposure to immunosuppressive and/or biologic agents and birth outcomes in patients with immune-mediated inflammatory diseases. *Gastroenterology* 2021; 161:107.e3–115.e3.
  22. Varady NH, Chen AF, Niu R, *et al.* Association between surgical opioid prescriptions and opioid initiation by opioid-naïve spouses. *Ann Surg* 2023; 277:e1218–e1224.
  23. Khalil M, Woldesenbet S, Iyer S, *et al.* Impact of spousal mental illness on healthcare utilization among patients with gastrointestinal cancer. *J Surg Oncol* 2024; 1–8.
  24. Gaitonde S, Moride Y, Suarez E, Lopez-Leon S. Where are the data to assess the safety of paternal drug exposure? A systematic review of secondary databases: a contribution from IMI concePTION. *Pharmacoepidemiol Drug Saf* 2024; 33:e5764.
- In this study, the authors review literature on paternal medication prescriptions and neonatal outcomes, highlighting several databases from Nordic countries that enable linkage of fathers to newborns at the population level.
25. Chen Y, Shen Q, Lichtenstein P, *et al.* Incidence trajectories of psychiatric disorders after assault, injury, and bereavement. *JAMA Psychiatry* 2024; 81:374–385.
- This study demonstrates how linked relational networks can be used to enable quasi-experimental study designs that use sibling controls at the population level. In doing so, this study is able to control for family-level confounding in the incidence of mental health diagnoses following traumatic events.
26. D'Onofrio BM, Lahey BB, Turkheimer E, Lichtenstein P. Critical need for family-based, quasi-experimental designs in integrating genetic and social science research. *Am J Public Health* 2013; 103(Suppl 1):S46–S55.
  27. Angier H, Kaufmann J, Heintzman J, *et al.* Association of parent preventive care with their child's recommended well child visits. *Acad Pediatr* 2022; 22:1422–1428.
  28. Martwick J, Kaufmann J, Bailey S, *et al.* Impact of healthcare location concordance on receipt of preventive care among children whose parents have a substance use and/or mental health diagnosis. *J Prim Care Community Health* 2024; 15:21501319241229925.
- This article demonstrates the utility of linked relational networks to study how healthcare delivery location affects well child care adherence among a population of children whose parents have mental health and/or substance use diagnoses.
29. Kaufmann J, DeVoe JE, Angier H, *et al.* Association of parent influenza vaccination and early childhood vaccinations using linked electronic health record data. *Vaccine* 2022; 40:7097–7107.
  30. Weaver J, Hardin JH, Blacketer C, *et al.* Development and evaluation of an algorithm to link mothers and infants in two US commercial healthcare claims databases for pharmacoepidemiology research. *BMC Med Res Methodol* 2023; 23:246.
- This article validates a methodology to link parents and children in administrative healthcare claims databases using IBM MarketScan Commercial Database (CCA) and Optum's de-identified Clinformatics Data Mart Database (Clinformatics) of almost 3.5 million mothers to more than 4.1 million children. This methodology is used in several subsequent articles because of its utility for pharmacoepidemiologic research within these large databases.
31. Heerman WJ, Daley MF, Boone-Heinonen J, *et al.* Maternal antibiotic use during pregnancy and childhood obesity at age 5 years. *Int J Obes* 2019; 43:1202–1209.
  32. Polubriaginof FCG, Vanguri R, Quinnes K, *et al.* Disease heritability inferred from familial relationships reported in medical records. *Cell* 2018; 173:1692.e11–1704.e11.
  33. Krefman AE, Ghamsari F, Turner DR, *et al.* Using electronic health record data to link families: an illustrative example using intergenerational patterns of obesity. *J Am Med Inform Assoc* 2023; 30:915–922.
  34. Department of Health and Human Services. 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program. 2020. Available at: <https://www.federalregister.gov/documents/2020/05/01/2020-07419/21st-century-cures-act-interoperability-information-blocking-and-the-onc-health-it-certification>. [Accessed 7 October 2024]
  35. Lamar CE, Jain D, Schmitz KH. Perils of auto-population of newborn EHR with maternal intimate partner violence history. *Pediatrics* 2023; 152: e2022060189.
- This case report identifies a potential harm arising from automatically sharing information between linked network members. It proposes potential safeguards against breach of privacy in the setting of linked records and data sharing mandated in the 21st Century Cures Act.