Strategies for constructing household and family units with linked administrative records

Stephanie Cuccaro-Alamin [a,b,*], Andrea Lane Eastman [a], Regan Foust [a], Jacquelyn McCroskey [a], Huy Tran Nghiem [a], Emily Putnam-Hornstein [a,b,c]

[a] *Children's Data Network, Suzanne Dworak-Peck School of Social Work, School of Social Work, University of Southern California, United States*
[b] *California Child Welfare Indicators Project, School of Social Welfare, University of California, Berkeley, United States*
[c] *School of Social Work, University of North Carolina at Chapel Hill, United States*

ABSTRACT

Children's circumstances are inextricably tied to those of their parents. Two-generation intervention strategies recognize that improving the capacity and circumstances of parents will yield benefits in this generation and the next. Administrative data from health and human services agencies present a unique source of information concerning these interventions designed to buffer child and parent risk. Those records, however, often exist in silos, collected by discrete programs for administration and primarily focused on the client served. Breaking down those silos by linking parents and children, as well as other family or household members, makes it possible to effectively develop, coordinate, and evaluate two-generation intervention strategies. The objective of this review is to outline conceptual strategies for constructing household and family units with linked administrative records for research purposes. Specifically, this paper: (1) provides an understanding of how households and family units have historically been conceptualized in the United States and illustrate the limitations of this approach for studying complex families; (2) examines the limitations of this approach for studying contemporary families; and (3) explores strategies for organizing administrative data into household and family units for this purpose.

1. Overview

Children's circumstances are inextricably tied to those of their parents. Two-generation intervention strategies recognize that improving the capacity and circumstances of parents will yield benefits in this generation and the next (Chase-Lansdale et al., 2019; Eastman & Putnam-Hornstein, 2019). Health and human services agencies administer essential services to vulnerable individuals in at-risk families that aim to enhance family safety and well-being. These agencies also collect administrative data related to these individual clients and the services delivered. As such, these data present a unique source of information that may inform programs designed to offset family economic insecurity and buffer child and parent risk in the short and long term (e.g., health, mental health, social programs, financial assistance).

Advances in machine learning and computing technology have facilitated increasingly rigorous and accurate probabilistic methods for linking administrative records across service programs (Culhane &
Metraux, 1997; Hotz et al., 1998; Putnam-Hornstein, Ghaly, & Wilkening, 2020), enhancing our ability to understand the timing, sequencing, and outcomes that may result from multiple program encounters. Most linkage efforts, however, focus on individuals rather than households or families (Manning et al., 2014; O'Hara et al., 2017). To effectively develop, coordinate, and evaluate two-generation interventions, the composition of households and family relationships must also be represented in datasets. The construction of household and family identifiers creates opportunities to study child outcomes in the context of important and influential household or family dynamics, including the characteristics, experiences, and service encounters of other individuals in the household or family unit. Additionally, such identifiers can facilitate the incorporation of potentially valuable information from other household or family members into our individual-level records linkage algorithms (e.g., improving child-specific matches using parent information).

1.1. The current paper

The objective of this review is to outline conceptual strategies that could be adopted for organizing frequently encountered health and human services agency records into households or family units. The strategies identified were informed by an examination of relevant research in which we sought to provide an understanding of how households or family units have historically been conceptualized for research in the United States; examine the limitations of this approach for studying contemporary family structures; and explore different strategies that have been adopted for organizing administrative data into household and family units for this purpose.

Drawing heavily on international methods, this review lays the foundation for the development of household or family unit identifiers using administrative records from health and human services agencies to more effectively develop, coordinate, and evaluate two-generation intervention strategies in the United States.

2. What is family?

Family is the basic unit of social and economic organization in society and the primary mechanism for human socialization, support, and development (Ginther & Pollak, 2004; Zill & Daily, 1993). "Families preserve and transmit cultural values, provide financial and material support essential to the growth and development of their members, particularly infants and children, and for the care of other dependents, including the elderly, disabled, and infirm" (United Nations, 1994, p.1). Research shows that household and family structure play a critical role in children's health and safety, economic well-being, access to services, educational persistence and attainment, and behavior later in life (Carlson & Corcoran, 2001; Hill, Yeung, & Duncan, 2001; Krueger, Jutte, Franzini, Elo, & Hayward, 2015; McLanahan & Percheski, 2008).

2.1. Identifying family units

As the primary system influencing a child's development, the family is considered a basic unit of analysis in research on child well-being (Bronfenbrenner, 1979, 1986). Determining who counts in these units has important implications for our understanding of family functioning, particularly the measurement of household income, resource distribution, social support, and service utilization (Heggeness, Alexander, & Stern, 2012). In the United States, the most widely used and accessible information regarding families, their living arrangements, and economic well-being comes from the U.S. Census Bureau. Since the first decennial census in 1790, the Bureau and its associated surveys have helped define our collective concept of family and how it is typically operationalized in research (McCarthy, Doolittle, & Sclater, 2012).

The census approach for defining family units is rooted in neoclassical economics and revolves around the household or residence (Amato, 2014; Oya, 2015). The terms *family* and *household* are often used interchangeably. Although both are basic units of analysis in demography, they are not the same thing. In short, households can contain multiple families, and families can exist across multiple households (Hill, 1995; McFalls, 2003).

2.1.1. Household

According to the Bureau, "a household includes the related family members and all the unrelated people, if any, such as lodgers, foster children, wards, or employees who share the housing unit. A person living alone in a housing unit, or a group of unrelated people sharing a housing unit such as partners or roomers, is also counted as a household" (U.S. Census Bureau, 2019, para. 23). The census examines the relationship of each individual in the household to a "householder," who is identified as the household head to distinguish between family and nonfamily households (U.S. Census Bureau, 2019).

2.1.2. Family/family structure

In the family household, the Bureau then identifies individual families. For statistical purposes, a family is defined as: "a group of two people or more (one of whom is the householder) related by birth, marriage, or adoption and residing together" (U.S. Census Bureau, 2019, para. 14). All people in a household who are related to the householder are regarded as members of the family. A family household, however, may also include related and unrelated subfamilies, married couples, or a parent–child group that does not include the householder.

The census further classifies individual families in the household by their structure. Family structure refers to the "composition and membership of the family and the organization and patterning of relationships among individual family members" (O'Toole, 2013, p. 676). Structural characteristics often "refer to who is in the family (with respect to number, gender, and generation) and the basis of family membership (blood, marriage, or adoption)" (Amato, 2014, p. 185).

For families with children, several characteristics including the number of parents, the marital or union status of parents, and the biological relationship of parents to children interact to determine how family structure is classified in the census, e.g., two married parents, single parent, etc. (Hill, 1995; U.S. Census Bureau, 2019). Although distinct from the household, the degree to which the American definition of "family" has been shaped by the U.S. Census Bureau and its family household conceptualization cannot be overstated (Amato, 2014). Recent rapid demographic shifts, however, have revealed important limitations in this dominant methodological approach.

3. The changing family

In the United States, "family" has traditionally been defined quite narrowly as the nuclear two-parent family (Bernardes, 1985; Harter & Bertrand, 1977; Hill, 1995; Wetzel, 1990). During the last half-century, however, changes in gender roles, work, marriage, and childbearing have brought about rapid demographic shifts that have resulted in a decline in children being raised in "traditional" families (Bianchi & Casper, 2000; Brown & Manning, 2009; Cancian & Reed, 2009; Cherlin, 1992, 2004, 2010; Coontz, 2006; Miller, 1992; Payne, 2018; Waldfogel et al., 2010). As of 2014, 46% of all children in the United States under age 18 lived in a home with two married heterosexual parents in their first marriage, compared to 73% in 1960 (Livingston, 2014). Since the 1960s, an increased rate of divorce, declines in the marriage rate, and increases in nonmarital fertility have resulted in a decoupling /separation of marriage and childbearing (Allred, 2018; Lundberg & Pollak, 2007; Payne, 2018; Raley & Wildsmith, 2004). With children increasingly living in more diverse family structures, demographic shifts have exposed limitations in conventional methods of measuring families and households (Allred, 2018; Brown & Manning, 2009; Payne, 2018; Raley & Wildsmith, 2004).

3.1. The complex contemporary family

Contemporary families are more dynamic and tend to have more ambiguous boundaries (Brown & Manning, 2009). Many groups that operate as families from the perspective of the child are not classified as such under the "family household" construct used in the census (Raley & Wildsmith, 2004). Such static definitions are often oversimplified and ignore variation in children's living arrangements and relational ties (Madhavan, Clark, Beguy, Kabiru, & Gross, 2017).

3.1.1. Dynamic

Families are geographically dynamic in that they may reside across several households, cities, states, or even continents. For that reason, measurement based on common residence alone may exclude important economic and social contributions from kin who do not reside in the home (Akresh & Edmonds, 2010; Hill, 1995). Conversely, people living in the same household can belong to different families. Therefore,

household-based measurement may include non-kin who reside in the home but don't contribute economically to a child's development (Cherlin & Furstenberg, 1994; Seltzer, 2000; Widmer, 1999). Such is the case in households where families are doubled up (Mykyta & Macartney, 2011). Although the census differentiation between household and family household accounts for this, less sophisticated methodologies often employed in family research may not accurately reflect the reality of children's everyday lives (Brown & Manning, 2009).

Families are also temporally dynamic because they experience transitions such as births, deaths, changes in marital status, or living arrangements like cohabitation (Brown, 2004; Hill et al., 2001; Magnuson & Berger, 2009; Wu & Martinson, 1993). Changes in the composition of the household and the timing of such transitions during a child's development can influence child well-being (Cavanagh & Huston, 2008; Fomby & Cherlin, 2007; Raley & Wildsmith, 2004; Smith, Crosnoe, & Cavanagh, 2017; Wu & Martinson, 1993). Specifically, research has shown that children experiencing multiple transitions (i.e., divorce, births, marriages) may have poor social adjustment (Cavanagh & Huston, 2008).

Non-traditional or complex family forms have been shown to be more unstable than traditional ones (Guzzo, 2014b; Kennedy & Ruggles, 2014; Magnuson & Berger, 2009; Raley & Bumpass, 2003; Raley & Wildsmith, 2004). Children in these families are more likely to experience instability such as divorce or separation (Bumpass & Lu, 2000; Raley & Wildsmith, 2004). Static measures of family structure often cannot capture these important dynamic features of contemporary families.

3.1.2. Ambiguous boundaries

The concept of family boundaries derives from family systems theory and refers to the processes or rules and roles by which members participate in family life (Carroll, Olson, & Buckmiller, 2007). Boundaries in families are often more flexible and permeable than rigid household constructs can measure (Brown & Manning, 2009; Sweeting & Seaman, 2005). Complex families, in particular, are often characterized by a higher degree of boundary ambiguity (Boss & Greenberg, 1984). As a result, when measuring families, who counts depends on social, legal, and political factors. Increasingly, family researchers must take a more subjective perspective using an emic approach, which allows families to define their own perceptions of family memberships and relationships as opposed to using fixed classifications (Amato, 2014).

Moving beyond the parent–child relationship is also critical. Increasingly, siblings have been recognized as a critical component of the child's family experience (Björklund, Ginther, Sundström, & Sundstrom, 2007; Gennetian, 2005; Mostafa, Gambaro, & Joshi, 2018). Growth in multiparter fertility has increased attention to the sibling composition of families (Sweeney, 2010). More than three quarters (78%) of children live with siblings, and more than one third (36%) live with siblings who do not share the same biological parents (Guzzo, 2014a; Kreider & Ellis, 2011). Research has suggested sibling relationships, whether full or nonfull (half or step), can greatly affect family functioning and children's experiences (Fomby, Goode, & Mollborn, 2016; Gennetian, 2005; Halpern-Meekin & Tach, 2008; Hofferth, 2006).

Additionally, stepparents, parents' cohabiting partners, and extended family members such as grandparents can also play important roles (Eggebeen, 1992; Hogan, Eggebeen, & Clogg, 1993; Mollborn et al., 2011). Household-level data often does not capture the social and emotional connections or influence across and between generations, which may be important to the development of a child (Chatters, Taylor, & Jayakody, 1994; Eggebeen, 1992; Hogan et al., 1993; Mostafa et al., 2018; Sharma, 2013).

4. Methods for measuring complex families

The growth of complex family structures means that our taxonomies must be adapted to reflect them (Beauregard, Ozbilgin, & Bell, 2009;

Coontz, 1998, 2013; Kennedy & Bumpass, 2008; Kreider & Ellis, 2011; Raley & Bumpass, 2003; Shorter, 1975). Experts agree that data collection and research must extend beyond the child and parents (or parent figures) to incorporate broader indicators of family membership, including sibling composition and multigenerational families (Gennetian, 2005; Ginther & Pollak, 2004; Halpern-Meekin & Tach, 2008; Mostafa et al., 2018). Additionally, new approaches must address the dynamic nature of family structure by accounting for transitions (Cherlin, 2010). Several promising approaches have emerged to more effectively capture the dynamic features of contemporary families.

4.1. Census changes

In response to the rapid demographic shifts of the last half century, the U.S. Census Bureau has made several important methodological modifications to more accurately capture the living arrangements of complex families (Brown, 2010). For example, to better measure cohabitating unions, "unmarried partner" was added as a "relationship to the householder" option to the census in 1990 (Kennedy & Fitch, 2012; Kreider, 2008; Lewis, 2013). The 2020 census will also include same-sex husband, wife, spouse, and unmarried partner options (Kennedy & Fitch, 2012; Manning & Smock, 2005; Ortman, 2007; U.S. Census Bureau, 2018; Wang, 2018).

Starting in 2007, several federal surveys added household relationship matrixes that identify how each person in a household is related to others, not just the head of household (Kreider, 2008), as well as parental pointers that allow for the identification of two parents for each child and classify the type of parental relationship (biological, adoptive, or step) (Kennedy & Fitch, 2012; Kreider, 2008). Despite these improvements, there remains considerable variation in the way cohabitation and family relationships are measured across federal surveys (Manning, Joyner, Hemez, & Cupka, 2016).

4.2. Family complexity

Addressing the important role of sibling relationship in children's well-being, the family complexity approach enhances family measurement by merging measures of family structure and sibling composition (Björklund et al., 2007; Gennetian, 2005; Hofferth, 2006; Manning, Brown, & Stykes, 2014). As such, the approach can "distinguish between simple two-biological-parent families, families with complex-sibling (half or stepsiblings) arrangements, and complex-parent (stepparent, single-parent) families" (Manning et al., 2014, p. 1). Including information regarding siblings in family classifications can also enable researchers to infer unobserved processes around family structure (Mostafa et al., 2018).

4.3. Network approaches

Although the notion that families extend beyond households is widely recognized, previous research has not been clear about how to study families as distinct from households. Used primarily in qualitative studies of social capital, the network approach views families "as overlapping networks that extend across multiple households" (Amato, 2014, p. 192). The approach uses a focal individual at the center of the network as the unit of analysis (Bessière & Gollac, 2018; Widmer, 2006).

Recognizing the dynamic and ambiguous nature of families, Bonvalet and Lelièvre (2016) have proposed the concept of the "entourage." The entourage or contact circle includes the key members of the family network to which an individual belongs as well as the successive households or residential systems to which the individual has belonged over a lifetime (Bonvalet & Lelièvre, 2016). Widmer (1999) developed an approach to analyze multiple relationships in a single model. The method considers family contexts as cognitive networks and utilizes egocentric focal mapping of networks. Using this method, Widmer (2006) demonstrated that individuals' significant family configurations

extend well beyond the nuclear family unit. Although used primarily in qualitative analysis, recent technological advances network analysis make this approach well suited to the study of increasingly complex and unbounded family contexts (Widmer, 1999).

4.4. Longitudinal

Given the importance of family transitions in child-level outcomes, researchers have also developed methods to capture the temporally dynamic nature of household and family systems (Citro, Hernandex, & Herriot, 1986; Citro & Watts, 1987; Duncan & Hill, 1985; Hill, 1992; Hill et al., 2001; Ruggles, 1998). Economists and poverty researchers have experimented using longitudinal methods to capture how changes in family structure affect economic well-being.

Measuring longitudinal changes in households or families is complex (Citro et al., 1986; Citro & Watts, 1987; Ruggles, 1998). Generally, researchers examine "at what age?, for how long?, and with whom?" individuals are connected to household and family units (Schumm, 2012. p. 1358). Measurements of household characteristics, however, can lose power when examined longitudinally because the changes being monitored (e.g., marriage, birth, death) often result in formation of new units (Doyle & Long, 1998). To address this loss of precision, some have advocated for an attribute-based approach that applies aggregate conditions of the household (e.g., composition of the household and the individual relationships to other coresidents) measured in discrete time intervals (such as months) to the individual (Duncan & Hill, 1985; Ruggles, 1998). Using this method, dynamic changes can be incorporated as explanatory variables (Duncan & Hill, 1985).

Hill (1992) argued for a multidimensional approach that enables researchers to aggregate individuals into distinct longitudinal analysis units (subfamilies, household and family) depending on the research context. The approach measures household or family formation (i.e., initial formation), and tracks changes over time, including entries and exits due to formation or dissolution of marital or cohabiting unions; entry of children due to childbirth or adoption, temporary, or long-term entries or exits by other household members, including aging parents, college students, nonrelatives (e.g., schooling, incarceration, setting up own household); and exits due to death. To provide a more complete picture, other characteristics must also be tracked, including age and gender of members, relationships among members, and information regarding pooling and distribution of resources.

5. Administrative records

Although these new approaches enable researchers to more effectively characterize the dynamic features of contemporary families, their application has been limited to smaller research studies or specific national surveys. Large-scale national surveys, such as the U.S. Census, have not only struggled to keep pace with the changing nature of the American family, but also have become increasingly expensive and difficult to administer and are often plagued by high non-response rates (Meyer, Mok, & Sullivan, 2015). Social researchers and policy makers are increasingly turning to administrative data sources to help fill these gaps (Harron, 2016; O'Hara, Shattuck, & Goerge, 2017).

Administrative records can be generally described as data derived from the operation of systems charged with determining eligibility and delivering services. They are most often collected by government agencies for the purposes of registration, transaction, and record keeping. Examples of administrative data include records relating to births and deaths, earnings and taxes, vehicle licensing and registration, medical conditions and procedures, receipt of public benefits, educational persistence and attainment, criminal justice, and child protective services contacts (Card, Chetty, Feldstein, & Saez, 2010).

5.1. Benefits

Administrative data have emerged as an important data source for research on family and child well-being. Data collected from agencies and programs that deliver services to children and families offer some advantages over smaller sample-based studies and nationally representative surveys. First, administrative data are widely available and considerably less expensive than research involving fieldwork or complex sampling techniques (Christensen, 1958; Dunn, 1946; Roos et al., 2008). Second, administrative datasets involve well-defined service populations. For that reason, sampling error is not an issue and findings have high levels of external validity and applicability (Harron et al., 2017). Third, although not error-free, administrative data are highly reliable because they involve records of events like birth, marriage, and divorce that are officially gathered for legal purposes. They are also relatively uniform across geographies and clients (Christensen, 1958). Fourth, administrative datasets are typically dynamic. Specifically, they are event-based encounter data that can be configured longitudinally to examine changes and outcomes over time (Connelly, Playford, Gayle, & Dibben, 2016). And because they are longitudinal, nonresponse or subject attrition at follow-up, which is common in survey research, is not an issue (Christensen, 1958; Meyer et al., 2015). Last, administrative data can be used to study hard-to-reach or underrepresented populations and sensitive or stigmatized topics that are often difficult to examine in surveys (Bell, Bayliss, Glauert, & Ohan, 2018; Brownell & Jutte, 2013; Christensen, 1958; Harron et al., 2017).

5.2. Limitations

Administrative data also have limitations, most stemming from the fact that they are "found" data not explicitly designed for research purposes (Connelly et al., 2016; Harron et al., 2017). As such, researchers are restricted to the information that is collected and to the format in which it was collected (Christensen, 1958). First, data quality can be an issue. Administrative data often are not collected in a uniform format, and missing data issues are common. In addition, with respect to family research, health and human services agency taxonomies do not always capture complex family forms (Beauregard et al., 2009). Second, datasets are often complex and require extensive cleaning and organizing to be utilized (Goerge & Lee, 2002). Codebooks are often not available, which means that researchers must consult with agency administrators to determine how data are generated and identify potential sources of error and bias (Lin, Maxwell, & Forry, 2017). Third, working with these datasets often requires a higher level of database management and coding expertise. For instance, the databases may be multidimensional or relational and require sophisticated linkages to utilize (Connelly et al., 2016). Last, there are important confidentiality and privacy concerns involved in working with individual-level administrative data because they often involve clients' personally identifiable information (Hotz et al., 1998; Künn, 2015; Stevens & Laurie, 2014).

Despite these limitations, advances in computing technologies have emerged to facilitate linkage of records in administrative datasets with one another or with survey data to more accurately detail individual service trajectories and outcomes over time (Culhane & Metraux, 1997; Hotz et al., 1998). These methods offer researchers and policy makers a powerful tool to help understand individuals and complex families in the 21st century.

5.3. Linked administrative data

Linked administrative data have been utilized to examine many aspects of child well-being and are an important tool for studying the public service trajectories of vulnerable children (Brownell & Jutte, 2013; Connelly et al., 2016; Currie, 2013; Harron et al., 2017; O'Hara et al., 2017; Putnam-Hornstein, Needell, & Rhodes, 2013). Increasingly, researchers are linking administrative data sources to construct more

complete records of family structure and outcomes (Culhane & Metraux, 1997; Harron, 2016; O'Hara et al., 2017).

Although the challenges associated with administrative data extend to linked administrative data, linked data also present their own specific challenges. First, the unit of analysis can vary across data sources, because the information collected is determined by agency statutes and the specific program's needs. Second, the focal individual can vary among programs, which has implications for the information available for nonfocal clients. Third, even when the focal individual is the same across data sources, these sources often lack common identifiers. Therefore, linkage often requires access to personally identifiable information (Harron et al., 2017; O'Hara et al., 2017; Stevens & Laurie, 2014), which is typically protected and its use limited for reasons of privacy and confidentiality. Researchers must have institutional review board and human subjects committee approval. Often, data are linked in a secure location and then stripped of personal information once linkage keys are created to decrease the risk of privacy or confidentiality breaches (Hotz et al., 1998; Stevens & Laurie, 2014).

There are two broad approaches to the linkage of administrative records using personally identifiable information: deterministic and probabilistic (Harron, 2016). Both methods utilize common individual identifiers such as name, birth date, place of birth, gender, social security number, agency client identifiers, and addresses to link records. Deterministic matches rely on direct matches between these fields, whereas probabilistic matching utilizes algorithms that assign weights to matches between fields and probabilities to record matches (Harron, 2016). Probabilistic methods are better suited for linkage of large-scale public service datasets, which may have missing or incomplete fields and lack unique identifiers common to both data sources (Bentley, Ford, Taylor, Irvine, & Roberts, 2012; Hagger-Johnson, Harron, Goldstein, Aldridge, & Gilbert, 2017). Unlike deterministic linkages, probabilistic linkage must balance false negatives (i.e., failing to link records belonging to the same individual) with false positives (i.e., erroneously linking records belonging to different individuals) (Zhu, Matsuyama, Ohashi, & Setoguchi, 2015). Probabilistic matching is often iterative, and matches can be improved with the addition of new data (Harville & Moore, 1999). Improvements in record-matching techniques and software have made this method increasingly accessible (Enamorado, Fifield, & Imai, 2019; Victor & Mera, 2001).

Probabilistically linked data from various administrative sources can be used to augment one another. Not only can this method be used to add relevant fields, populate missing or incomplete fields, validate existing information (such as self-reported data), and extend outcome measurement beyond the initial data collection period, it also can be configured to examine changes in status or outcomes over time (Card, Hildreth, & Shore-Sheppard, 2004; Künn, 2015; Meyer et al., 2015; O'Hara et al., 2017; Simon, 2014). Algorithm performance and match rates can also be improved through machine learning training procedures that incorporate information from additional datasets (Luque & Wagner, 2015; Massey, 2014; Millett, Quint, De Stavola, Smeeth, & Thomas, 2016). Finally, when "records are successfully linked at the individual level, a broader range of relevant factors and outcomes can be examined longitudinally for entire populations" (Culhane, Fantuzzo, Hill, & Burnett, 2018, p. 222).

6. Constructing household and family units with linked administrative data

Although most of the research to date using linked administrative data has focused on individuals, improvements in data collection and linkage techniques have made this an increasingly promising method for household and family research (Manning et al., 2014; O'Hara et al., 2017). Constructing household- and family-level records from linked administrative data is complex because the unit of analysis often varies across data sources and administrative data rarely contain complete household identifiers (Goldschmidt, Klosterhuber, & Schmieder, 2017).

Despite these barriers, a targeted review of the literature identified several methodologies that have been developed to link individual service records and construct household- and family-level data (Harron, 2016; Putnam-Hornstein et al., 2013).

Many of the relevant strategies documented here represent the work of international researchers operating in the context of very different social security systems and data infrastructures. The review is designed to provide illustration of possible techniques, but it must be recognized that currently the United States does not offer the same level of data integration.

6.1. Households

International researchers have developed techniques for constructing household identifiers from address fields found in administrative data. Additionally, they have shown how these identifiers can be configured longitudinally. In the US, researchers have also shown that when multiple health and human service data sources are utilized, benefit units can be used to reliably identify households. This section illustrates some of the challenges of this address-based approach and highlights the importance of data quality and coverage.

6.1.1. Address data

Researchers have experimented with different approaches of identifying households using address information from administrative records and compared it to national census sources for validation (Gath & Bycroft, 2018; Harper & Mayhew, 2016). Particularly notable is the work of the United Kingdom's Office for National Statistics, which has developed methodologies to capture household structures using administrative data sources in an effort to replace traditional census enumeration (Office for National Statistics, 2017).

Ongoing work by researchers in Nordic countries has explored the challenges of using address data for identification of household structure (Christiansen & Keilman, 2013). Using address data as a proxy for household requires redefining households as a group of individuals who reside at the same address at a particular time (Gibb & Das, 2015). Address quality issues, however, can limit the ability to group individuals into correct households (Gath & Bycroft, 2018; Gibb & Das, 2015). Although address fields such number, street, city, state, and ZIP code are common across most administrative files, agencies may combine some fields and not include others. Data entry errors in address fields or missing address data are also common and can affect linkage rates (Ansolabehere & Hersh, 2017; Dusetzina et al., 2014). Research has suggested this issue may be more prevalent among people with complex family forms (Gath & Bycroft, 2018; Gibb & Das, 2015). The method is also limited by the fact that it cannot distinguish between multiple families living at the same address, which is also more common among complex families.

Working in New Zealand, Gibb and Das (2015) set out to examine the quality of address information in the administrative data sources held in Statistics New Zealand's Integrated Data Infrastructure. Using tax, birth, and migration data to create a spine of all individuals, they linked address data from five administrative sources (taxes, health services, general practitioner, social assistance, and education) to create household units. They then compared measures of household location, size, and composition to their nation's 2013 census for validation. Results showed that 55% of households identified using administrative data had the same household size as the census and 48% contained the same set of household members. Larger households were more difficult to accurately classify.

Given the problems associated with missing or incomplete address data, Harper and Mayhew outlined a bottom-up approach that uses locally available (and potentially richer) administrative data to classify household types in London, England (Harper & Mayhew, 2012a, 2012b, 2016). Specifically, point-in-time, person-level population registry data (including addresses) from six boroughs were aggregated by age, sex, or

other attributes, resulting in eight core household types that can be defined depending on the level of specificity required for the analysis. These include family household with children, single adult with children, older cohabiting couple, older person living alone, three-generation household, cohabiting adult household, adult living alone, and other households such as split generation or children living alone with no adult such as students. Additional subtypes can also be specified to account for age and size combinations. For example, for "family households with children" subtypes can be delineated to record the specific number of children in the household. Age categories are flexible and could be modified, if appropriate, for the specific research question. For example, for studies involving young children, family types could be defined based on the number of children under age 5. Although a full population registry is required to complete this approach, it provides a potential method defining household and family structure definitions using individual-level data.

6.1.2. Longitudinal approaches

Clearly, constructing a cross-sectional database of households using administrative data is complicated; leveraging administrative data to do so in a longitudinal format is even more complex. Tingay, Roberts, and Musselwhite (2018) outlined a method for establishing and tracking households for longitudinal research using the large-scale integrated Secure Anonymous Information Linkage health services database at Swansea University.

Using demographic and geographic data, the research team created a method of nested linkages that allows for the identification of households (groups of people who share accommodations) and the family relationships in them. The database includes individual-level anonymized linkage fields (ALFs) that allow for probabilistic linkage between datasets. Each ALF is also associated with a residential anonymized record linkage field (RALF), which is a unique registered property address. RALFs include dates of moving in and out based on available data. Using RALFs, individuals living at the same residence during the same time period are assigned household anonymized linkage fields (HALFs). Relationship data are used to categorize families. Specifically, birth data are used to apply the birth mother's ALF to each child record so that households can also be divided into biological (or partly biological) and nonbiological. Although biological fathers cannot be identified in the data, an adult male living in the same HALF as children either since or prior to the birth of the children could be used to infer a cohabiting relationship with other adults and children. The approach is promising in that it enables researchers "to track household members across addresses and detect changes in households over time" (Tingay et al., 2018. p. 5).

6.1.3. Benefit or assistance units

In addition to address data, U.S. researchers have explored the use of benefit or assistance units found in social program and health data to approximate households and families. Although the individuals covered under the assistance and benefit unit depend on the particular service program, when multiple program units are triangulated, approximate household rosters can be defined. For example, using data on client eligibility for Temporary Assistance for Needy Families, Basic Food/Supplemental Nutrition Assistance Program (SNAP), Medicaid, and other services from its integrated client databases, the Washington State Department of Social and Health Services developed an algorithm to identify households using program-level assistance units. It showed that when linked with other data sources Basic Food/SNAP data are a plausible input source for the creation of a household database (Sprague, 2017).

The use of SNAP benefit units holds particular promise for household approximation because a SNAP household or unit is defined as "everyone who lives together and purchases and prepares meals together is grouped together" (U.S. Department of Agriculture, 2018). SNAP data are used by Medicaid programs to renew Medicaid eligibility or enroll children and adults younger than 65 in Medicaid (U.S. Department of Agriculture, 2019) allowing for linkage to other programs. Although, simulation research has demonstrated that SNAP units alone often undercount household members (Czajka & Cunnyngham, 2016; Czjaka, Cunnyngham, & Rosso, n.d.). With more than 40 million Americans receiving SNAP benefits, the family benefit data may still serve as an effective resource for linkage purposes.

6.2. Families

Constructing accurate family units from administrative data requires detailed and complete information on relationships between individuals. Because administrative data are typically collected only for individuals receiving program services, enumerating accurate family rosters based on these data sources alone likely excludes other family members not eligible or not receiving services. Recent research highlighted this issue.

Using relationship information found in New Zealand administrative data, Gath and Bycroft (2018) attempted to create family units as of March 5, 2013, and compared data to their nation's 2013 census for validation. To do so, all known partnerships and parent–child relationships (regardless of residence) that could be determined from administrative sources were recorded at the individual level. Address information was then used to identify "unique family nuclei" in households. "A family (or family nucleus) is defined as a couple, with or without child(ren), or one parent and their child(ren), all of whom usually reside together in the same household" (Gath & Bycroft, 2018, p. 8).

The results indicated that the family relationship information available in the administrative data were insufficient to create accurate family units. When information was available, it tended to match the family information reported in the census. Adults aged older than 60 years and those in their mid-20s were more likely to have missing family information. Despite these limitations, research has shown that several important partial family units, including parent–child relationships, couples, and sibling relationships, can be discerned using linked administrative data.

6.2.1. Parent and child

The most common family-level administrative data linkage is between parents and children. These dyads or triads are often used in both program administration and research. For example, the U.S. Internal Revenue Service routinely uses social security data for children linked to parental social security numbers to validate earned income tax credit filings. Both deterministic and probabilistic parent–child linkages are also common across census surveys (Johnson, Massey, & O'Hara, 2015; O'Hara, Genadek, Medalia, & Alexander, 2018). Linkages of federal data have been used successfully to study intergenerational mobility, earnings, inequality, and use of public benefits (Chetty, Hendren, Kline, Saez, & Turner, 2014).

Parent–child linkages can most reliably be determined through vital statistics birth or adoption records. Although methods vary, "near universal" mother–child linkages have been achieved using vital birth records (Jutte, Roos, & Brownell, 2011, p. 94). A majority of fathers (~85%) can also be identified with this method (Jutte et al., 2011). Missing paternal information on birth registrations, particularly age and ethnicity, is common in vital statistics registries worldwide (Whitley, Deary, Der, Batty, & Benzeval, 2012) and more common among vulnerable populations (Tan, Wen, Walker, & Demissie, 2004). However, paternal information can be improved when multiple administrative sources are linked (Sims & O'Donnell, 2015).

Birth records have been used extensively to identify parents and children and then linked with other administrative data sources to study child well-being outcomes (Arthur, Lucenko, Sharkova, Xing, & Mangione-Smith, 2018; Eastman & Putnam-Hornstein, 2019; Finno-Velasquez, Palmer, Prindle, Tam, & Putnam-Hornstein, 2017;

Maloney, Jiang, Putnam-Hornstein, Dalton, & Vaithianathan, 2017). In Allegheny County, Pennsylvania, Maloney et al. (2017) used both parent and child identifiers available in birth records and probabilistically linked births to administrative records in the county's human services data warehouse, including records from child protective services, juvenile probation, alcohol and drug rehabilitation, behavioral health services, and county jail.

Although birth records can be used to reliably identify biological parent–child relationships, other parental types such as foster and adoptive may be more difficult to discern. Single-source administrative data sources such as child protective services, education, or health services often offer detailed relationship indicators between clients on a case or benefit unit that can be used for this purpose. In specific datasets, more nuanced parent–child relationships may be recorded, including biological, adoptive, foster, step, presumptive, etc. Once identified in a single source, these relationships can be validated by linkages to additional data sources.

When detailed relationship indicators are not available, researchers can also use other data sources to link parents and children. For instance, researchers successfully created child–parent links in the state's public health care insurance program using household identification numbers and in a community health information network database using "guarantor" and "emergency contact" relationship fields (Angier et al., 2014). Research has suggested that the more diverse administrative data are mined the more accurate these linkages can be (Ong, Mannino, Schilling, & Kahn, 2014), and the more complete a picture of children's family networks can emerge (Angier et al., 2014).

6.2.2. Couples

Identifying couples in administrative data is challenging. Theoretically, vital statistics marriage and divorce registries can be used to track formation and dissolution of formal (legalized) unions. However, research has shown that state marriage and divorce records are not as complete or reliable as state birth and death registries (The Lewin Group, 2008). Although improvements in data quality and electronic data collection have been made in recent years, overall declines in marriage and the emergence of more complex family forms also limits the usefulness of such data (The Lewin Group, 2008).

In the absence of complete marriage registries, some researchers have linked other administrative data sources to identify married couples. For example, in New Zealand, Suei (2016) compared legal relationships reported in the New Zealand census with their status constructed from administrative data sources and found them to be highly consistent for married couples, but less so for those in other relationship categories, particularly those separated or widowed. Similarly, in a study using Swedish Censuses, Wisselgren, Edvinsson, Berggren, and Larsson (2014) found more complex or dynamic family structures were more difficult to link using administrative data. For example, families in which the father is no longer present and the mother is either a widow or has remarried are more difficult to link.

Goldschmidt et al. (2017) developed a method to impute married couple household identifiers from administrative employment records in Germany. Criteria for married pairs included same home location, uniquely matching last name, and one male and one female with an age difference of less than 15 years. Other research has shown that higher linkage rates can be achieved by applying constructed name variables and household links (Wisselgren et al., 2014). Cohabitating couples who share biological children may also be identifiable via birth records and status can be triangulated with address data.

6.2.3. Siblings

When parent–child linkages, and particularly mother–child linkages, are achieved, sibling mapping can occur. High rates of biological sibling matches have been achieved via birth and other service records (Chen, Chen, & Liu, 2009; Lery, Shaw, & Magruder, 2005). High rates of half-siblingship can also be identified using birth records, although such linkages tend to be biased toward maternal half-siblings because paternity is less often captured in vital records. Identifying stepsiblings is more complicated because it requires marriage or address data.

Using child protective services data from California, Lery et al. (2005) compared the effectiveness of four methods of sibling linkages: child relationship, maternal, paternal, and removal address (in cases of foster care placement) matching. Identified links overlapped significantly among methods; however, analysis of removal address yielded many possible siblings not identified using any of the other three methods. Using multiple methods yielded more matches than any method alone. Such methods provide a promising approach for linking more complex family structures including identifying half- and stepsiblings or fictive kin.

7. Conclusion

Constructing household and family identifiers is critical to effectively develop, coordinate, and evaluate two-generation intervention strategies. Doing so creates opportunities to study child outcomes in the context of important and influential dynamics, including the characteristics, experiences, and service encounters of other individuals in the household or family unit.

As this review has illustrated, linked administrative data provides a unique opportunity to group individuals at the household and family level, thereby improving policy and program research. Utilizing information collected by health and human services agencies in this new way will allow for a more complete understanding of the effects of programs offered by characterizing the interactions and trajectories of the household and family context.

Recent advances in machine learning and probabilistic matching techniques have facilitated increasingly rigorous and accurate methods for connecting administrative records across programs (Culhane & Metraux, 1997; Hotz et al., 1998). This review details the complexity in constructing household- and family-level records from linked administrative data and highlights the importance of access to reliable address and family relationship information (Gibb & Das, 2015). Maximizing the number of linked data sources has been shown to enhance data quality and improve linkage rates.

Despite these barriers, we identified methodologies that have been developed to construct household- and family-level data identifiers. Moving forward with the creation and validation these analysis units is an important next step, given that linked administrative data potentially offer a cost-effective and flexible strategy for measuring the role of dynamic household and family structure and program participation on child well-being. The ability to evaluate the long-term impact and cost effectiveness of programs for children and families is essential given the social and economic impact of the COVID-19 pandemic. Not only will the U.S. government be tasked with providing more services despite growing budget deficits, informed policy making also will be more crucial than ever before.

References

Akresh, R., & Edmonds, E. (2010). The analytical returns to measuring a detailed household roster (No. IZA Discussion Paper No. 4759). Retrieved from https://ideas.repec.org/p/iza/izadps/dp4759.html.

Allred, C. (2018). Marriage: More than a century of change, 1900-2016. Bowling Green, OH: National Center for Family and Marriage Research. https://doi.org/10.25035/ncfmr/fp-18-17.

Amato, P. (2014). Approaches to measuring families. In S. M. McHale, P. Amato, & A. Booth (Eds.), *Emerging methods in family research* (pp. 179–193). New York, NY: Springer.

Angier, H., Gold, R., Crawford, C., O'Malley, J. P., Tillotson, C. J., Marino, M., & DeVoe, J. E. (2014). Linkage methods for connecting children with parents in electronic health record and state public health insurance data. *Maternal Child Health Journal, 18*(9), 2025–2033. https://doi.org/10.1007/s10995-014-1453-8

Ansolabehere, S., & Hersh, E. D. (2017). ADGN: An algorithm for record linkage using address, date of birth, gender, and name. *Statistics and Public Policy, 4*(1), 1–10. https://doi.org/10.1080/2330443X.2017.1389620

Arthur, K. C., Lucenko, B. A., Sharkova, I. V., Xing, J., & Mangione-Smith, R. (2018). Using state administrative data to identify social complexity risk factors for children. *The Annals of Family Medicine, 16*(1), 62–69. https://doi.org/10.1370/afm.2134

Beauregard, T. A., Ozbilgin, M., & Bell, M. P. (2009). Revisiting the social construction of family in the context of work. *Journal of Managerial Psychology, 24*(1), 46–65. https://doi.org/10.1108/02683940910922537

Bell, M. F., Bayliss, D. M., Glauert, R., & Ohan, J. L. (2018). Using linked data to investigate developmental vulnerabilities in children of convicted parents. *Developmental Psychology, 54*(7), 1219–1231. https://doi.org/10.1037/dev0000521

Bentley, J. P., Ford, J. B., Taylor, L. K., Irvine, K. A., & Roberts, C. L. (2012). Investigating linkage rates among probabilistically linked birth and hospitalization records. *BMC Medical Research Methodology, 12*(1), 149. https://doi.org/10.1186/1471-2288-12-149

Bernardes, J. (1985). Do we really know what 'the family' is? In P. Close, & R. Collins (Eds.), *Family and economy in modern society* (pp. 192–211). London, England: Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-17795-0_11.

Bessière, C., & Gollac, S. (2018). Is social network analysis useful for studying the family economy? *Economic Sociology, 19*(3), 4–10.

Bianchi, S. M., & Casper, L. (2000). *American families*. Washington, DC: Population Reference Bureau.

Björklund, A., Ginther, D., Sundström, M., & Sundstrom, M. (2007). Family structure and child outcomes in the USA and Sweden. *Journal of Population Economics, 20*, 183–201. https://doi.org/10.1007/s00148-006-0094-7

Bonvalet, C., & Lelièvre, É. (2016). A conceptual shift from "household" to "entourage": Redefining the scope of the family. In C. Bonvalet, & E. Lelièvre (Eds.), *Family beyond household and kin: Life event histories and entourage, a French survey* (pp. 57–72). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-24684-0_3.

Boss, P., & Greenberg, J. (1984). Family boundary ambiguity: A new variable in family stress theory. *Family Process, 23*(4), 535–546. https://doi.org/10.1111/j.1545-5300.1984.00535.x

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychology, 22*(6), 723–742. https://doi.org/10.1037/0012-1649.22.6.723

Brown, S. L. (2004). Family structure and child well-being: The significance of parental cohabitation. *Journal of Marriage and Family, 66*(2), 351–367. https://doi.org/10.1111/j.1741-3737.2004.00025.x

Brown, S. L. (2010). Marriage and child well-being: Research and policy perspectives. *Journal of Marriage and the Family, 72*(5), 1059–1077. https://doi.org/10.1111/j.1741-3737.2010.00750.x

Brown, S. L., & Manning, W. D. (2009). Family boundary ambiguity and the measurement of family structure: The significance of cohabitation. *Demography, 46*(1), 85–101. https://doi.org/10.1353/dem.0.0043

Brownell, M. D., & Jutte, D. P. (2013). Administrative data linkage as a tool for child maltreatment research. *Child Abuse & Neglect, 37*(2-3), 120–124. https://doi.org/10.1016/j.chiabu.2012.09.013

Bumpass, L., & Lu, H.-H. (2000). Trends in cohabitation and implications for children s family contexts in the United States. *Population Studies, 54*(1), 29–41. https://doi.org/10.1080/713779060

Cancian, M., & Reed, D. (2009). *Family structure, childbearing, and parental employment: Implications for the level and trend in poverty*. Princeton, NJ: Mathematica Policy Research.

Card, D. E., Chetty, R., Feldstein, M. S., & Saez, E. (2010). Expanding access to administrative data for research in the United States. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1888586

Card, D., Hildreth, A. K. G., & Shore-Sheppard, L. D. (2004). The measurement of medicaid coverage in the SIPP: Evidence from a comparison of matched records. *Journal of Business & Economic Statistics, 22*(4), 410–420. https://doi.org/10.1198/073500104000000208

Carlson, M. J., & Corcoran, M. E. (2001). Family structure and children's behavioral and cognitive outcomes. *Journal of Marriage and Family, 63*(3), 779–792. https://doi.org/10.1111/j.1741-3737.2001.00779.x

Carroll, J. S., Olson, C. D., & Buckmiller, N. (2007). Family boundary ambiguity: A 30-year review of theory, research, and measurement. *Family Relations, 56*(2), 210–230. https://doi.org/10.1111/j.1741-3729.2007.00453.x

Cavanagh, S. E., & Huston, A. C. (2008). The Timing of Family Instability and Children's Social Development. *Journal of Marriage and Family, 70*(5), 1258–1270. https://doi.org/10.1111/j.1741-3737.2008.00564.x

Chase-Lansdale, P. L., Sabol, T. J., Sommer, T. E., Chor, E., Cooperman, A. W., Brooks-Gunn, J., … Morris, A. (2019). Effects of a two-generation human capital program on low-income parents' education, employment, and psychological wellbeing. *Journal of Family Psychology, 33*(4), 433–443. https://doi.org/10.1037/fam0000517

Chatters, L. M., Taylor, R. J., & Jayakody, R. (1994). Fictive Kinship Relations in black extended families. *Journal of Comparative Family Studies, 25*(3), 297–312. https://doi.org/10.3138/jcfs.25.3.297

Chen, S. H., Chen, Y.-C., & Liu, J.-T. (2009). The impact of unexpected maternal death on education: First evidence from three national administrative data links. *American Economic Review, 99*(2), 149–153. https://doi.org/10.1257/aer.99.2.149

Cherlin, A. J. (1992). *Marriage, divorce, remarriage* (2nd ed.). Cambridge, MA: Harvard University Press.

Cherlin, A. J. (2004). The deinstitutionalization of American marriage. *J Marriage and Family, 66*(4), 848–861. https://doi.org/10.1111/j.0022-2445.2004.00058.x

Cherlin, A. J. (2010). Demographic trends in the United States: A review of research in the 2000s. *Journal of Marriage and Family, 72*(3), 403–419. https://doi.org/10.1111/j.1741-3737.2010.00710.x

Cherlin, A. J., & Furstenberg, F. F., Jr (1994). Stepfamilies in the United States: A reconsideration. *Annual Review of Sociology, 20*(1), 359–381. https://doi.org/10.1146/annurev.so.20.080194.002043

Chetty, R., Hendren, N., Kline, P., Saez, E., & Turner, N. (2014). Is the United States still a land of opportunity? Recent trends in intergenerational mobility (NBER Working Paper No. 19844). Retrieved from http://www.nber.org/papers/w19844.

Christensen, H. T. (1958). The method of record linkage applied to family data. *Marriage and Family Living, 20*, 38–43. https://doi.org/10.2307/347362

Christiansen, S., & Keilman, N. (2013). Probabilistic household forecasts based on register data- the case of Denmark and Finland. *Demographic Research, 28*, 1263–1302. https://doi.org/10.4054/DemRes.2013.28.43

Citro, C. F., Hernandex, D. J., & Herriot, R. A. (1986). Longitudinal household concepts in SIPP: Preliminary results. Retrieved from https://www2.census.gov/prod2/sipp/wp/SIPP_WP_019.pdf.

Citro, C. F., & Watts, H. W. (1987). Patterns of household composition and family status change: Survey of income and program participation. Retrieved from https://www.census.gov/library/working-papers/1986/demo/SEHSD-WP1986-03.html.

Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research, 59*, 1–12. https://doi.org/10.1016/j.ssresearch.2016.04.015

Coontz, S. (1998). *The way we really are: Coming to terms with America's changing families*. New York, NY: Basic Books.

Coontz, S. (2006). *Marriage, a history: How love conquered marriage*. New York, NY: Penguin Books.

Coontz, S. (2013). *American families: A multicultural reader*. New York, NY: Routledge.

Culhane, D. P., Fantuzzo, J., Hill, M., & Burnett, T. (2018). Maximizing the use of integrated data systems: Understanding the challenges and advancing solutions. *The Annals of the American Academy of Political and Social Science, 675*, 221–239. https://doi.org/10.1177/0002716217743441

Culhane, D. P., & Metraux, S. (1997). Where to from here? A policy research agenda based on the analysis of administrative data. In D. P. Culhane (Ed.), *Understanding homelessness: New policy and research perspectives* (pp. 341–360). Washington, DC: Fannie Mae Foundation.

Currie, J. (2013). "Big data" versus "big brother": On the appropriate use of large-scale data collections in pediatrics. *Pediatrics, 131*, S127–S132. https://doi.org/10.1542/peds.2013-0252c

Czajka, J. L., & Cunnyngham, K. (2016, November 2). Supplemental nutrition assistance program household composition: Issues and findings from linked survey and administrative data. In Presented at the 6 Census/ERS Administrative Records Conference.

Czajka, J. L., Cunnyngham, K., & Rosso, R. (n.d.). Simulated Versus Actual SNAP Unit Composition in Survey Households in Two States. 18.

Doyle, P., & Long, S. E. (1998). The impact of the unit of analysis on measures of serial multiple program participation. Retrieved from https://census.gov/library/working-papers/1988/demo/SIPP-WP-48.html.

Duncan, G., & Hill, M. (1985). Conceptions of longitudinal households: Fertile or futile? *Journal of Economic and Social Measurement, 13*, 361–375.

Dunn, H. L. (1946). Record linkage. *American Journal of Public Health, 36*, 1412–1416. https://doi.org/10.2105/AJPH.36.12.1412

Dusetzina, S. B., Tyree, S., Meyer, A.-M., Meyer, A., Green, L., & Carpenter, W. R. (2014). An overview of record linkage methods. *Linking Data for Health Services Research: A Framework and Instructional Guide [Internet]*. Agency for Healthcare Research and Quality (US). https://www.ncbi.nlm.nih.gov/books/NBK253312/.

Eastman, A. L., & Putnam-Hornstein, E. (2019). An examination of child protective service involvement among children born to mothers in foster care. *Child Abuse & Neglect, 88*, 317–325. https://doi.org/10.1016/j.chiabu.2018.11.002

Eggebeen, D. J. (1992). Family structure and intergenerational exchanges. *Research on Aging, 14*, 427–447. https://doi.org/10.1177/0164027592144001

Enamorado, T., Fifield, B., & Imai, K. (2019). Using a probabilistic model to assist merging of large-scale administrative records. *American Political Science Review, 113*, 353–371. https://doi.org/10.1017/S0003055418000783

Finno-Velasquez, M., Palmer, L., Prindle, J., Tam, C. C., & Putnam-Hornstein, E. (2017). A birth cohort study of Asian and Pacific Islander children reported for abuse or neglect by maternal nativity and ethnic origin. *Child Abuse & Neglect, 72*, 54–65.

Fomby, P., & Cherlin, A. J. (2007). Family instability and child well-being. *American Sociological Review, 72*, 181–204.

Fomby, P., Goode, J. A., & Mollborn, S. (2016). Family complexity, siblings, and children's aggressive behavior at school entry. *Demography, 53*, 1–26. https://doi.org/10.1007/s13524-015-0443-9

Gath, M., & Bycroft, C. (2018). Potential for linked administrative data to provide household and family information. Retrieved from https://www.stats.govt.nz/assets/Uploads/Reports/The-potential-for-linked-administrative-data-to-pr

ovide-household-and-family-information/the-potential-linked-admin-data-househ old-family-information.pdf.

Gennetian, L. A. (2005). One or two parents? Half or step siblings? The effect of family structure on young children's achievement. *Journal of Population Economics, 18*, 415–436. https://doi.org/10.1007/s00148-004-0215-0

Gibb, S., & Das, S. (2015). Quality of geographic information in the integrated data infrastructure. Retrieved from https://www.stats.govt.nz/research/quality-of-g eographic-information-in-the-integrated-data-infrastructure.

Ginther, D. K., & Pollak, R. A. (2004). Family structure and children's educational outcomes: Blended families, stylized facts, and descriptive regressions. *Demography, 41*, 671–696. https://doi.org/10.1353/dem.2004.0031

Goerge, R. M., & Lee, B. J. (2002). Matching and cleaning administrative data. *New Zealand Economic Papers, 36*, 63–64. https://doi.org/10.1080/00779950209544351

Goldschmidt, D., Klosterhuber, W., & Schmieder, J. F. (2017). Identifying couples in administrative data. *Journal for Labour Market Research, 50*, 29–43. https://doi.org/ 10.1007/s12651-017-0218-4

Guzzo, K. B. (2014a). New partners, more kids: Multiple-partner fertility in the United States. *The Annals of the American Academy of Political and Social Science, 654*, 66–86. https://doi.org/10.1177/0002716214525571

Guzzo, K. B. (2014b). Trends in cohabitation outcomes: Compositional changes and engagement among never-married young adults. *Journal of Marriage and the Family, 76*, 826–842. https://doi.org/10.1111/jomf.12123

Hagger-Johnson, G., Harron, K., Goldstein, H., Aldridge, R., & Gilbert, R. (2017). Probabilistic linking to enhance deterministic algorithms and reduce linkage errors in hospital administrative data. *Journal of Innovation in Health Informatics, 24*(2), 891. https://doi.org/10.14236/jhi.v24i2.891

Halpern-Meekin, S., & Tach, L. (2008). Heterogeneity in two-parent families and adolescent well-being. *Journal of Marriage and Family, 70*, 435–451. https://doi.org/ 10.1111/j.1741-3737.2008.00492.x

Harper, G., & Mayhew, L. (2012a). Applications of population counts based on administrative data at local level. *Applied Spatial Analysis and Policy, 5*, 183–209. https://doi.org/10.1007/s12061-011-9062-z

Harper, G., & Mayhew, L. (2012b). Using administrative data to count local populations. *Applied Spatial Analysis and Policy, 5*, 97–122. https://doi.org/10.1007/s12061-011-9063-y

Harper, G., & Mayhew, L. (2016). Using administrative data to count and classify households with local applications. *Applied Spatial Analysis and Policy, 9*, 433–462. https://doi.org/10.1007/s12061-015-9162-2

Harron, K. (2016). Introduction to data linkage. Retrieved from https://adrn.ac.uk /media/174200/data_linkage_katieharron_2016.pdf.

Harron, K., Dibben, C., Boyd, J., Hjern, A., Azimaee, M., Barreto, M. L., & Goldstein, H. (2017). Challenges in administrative data linkage for research. *Big Data & Society, 4*, 1–12. https://doi.org/10.1177/2053951717745678

Harter, C. L., & Bertrand, W. E. (1977). A methodology for classifying household family structures. *Journal of Comparative Family Studies, 8*, 401–413. https://doi.org/ 10.3138/jcfs.8.3.401

Harville, D. S., & Moore, R. A. (1999). Determining record linkage parameters using iterative logistic regression. Retrieved from http://www.stat.cmu.edu/NCRN/PUBL IC/RLClassFiles/Lectures/Harville-Moore-FS-Logistic-Regression.pdf.

Heggeness, M., Alexander, J., & Stern, S. (2012). Alternative strategies for grouping people into resource units: Measuring poverty in the American Community Survey. Retrieved from https://census.gov/content/dam/Census/library/working-papers/ 2012/demo/SEHSD-WP2012-09.pdf.

Hill, M. S. (1992). The future of SIPP for analyzing interactions of family composition and income change. *Journal of Economic and Social Measurement, 18*, 67–89. https:// doi.org/10.3233/JEM-1992-181-405

Hill, M. S. (1995). When is a family a family? Evidence from survey data and implications for family policy. *Journal of Family and Economic Issues, 16*, 35–64. https://doi.org/10.1007/BF02353666

Hill, M, S., Yeung, W.-J.-J., & Duncan, G. J. (2001). Childhood family structure and young adult behaviors. *Journal of Population Economics, 14*, 271–299. https://doi. org/10.1007/s001480000039

Hofferth, S. L. (2006). Residential father family type and child well-being: Investment versus selection. *Demography, 43*, 53–77. https://doi.org/10.1353/dem.2006.0006

Hogan, D. P., Eggebeen, D. J., & Clogg, C. C. (1993). The structure of intergenerational exchanges in American families. *American Journal of Sociology, 98*, 1428–1458.

Hotz, V. J., Goerge, R., Balzekas, J., Margolin, F., Hotz, V. J., Balzekas, J. D., … Wiseman, M. (1998). *Administrative data for policy-relevant research: Assessment of current utility and recommendations for development*. Evanston, IL: Joint Center for Poverty Research.

Johnson, D. S., Massey, C., & O'Hara, A. (2015). The opportunities and challenges of using administrative data linkages to evaluate mobility. *The Annals of the American Academy of Political and Social Science, 657*, 247–264. https://doi.org/10.1177/ 0002716214552780

Jutte, D. P., Roos, L. L., & Brownell, M. D. (2011). Administrative record linkage as a tool for public health research. *Annual Review of Public Health, 32*, 91–108. https://doi. org/10.1146/annurev-publhealth-031210-100700

Kennedy, S., & Bumpass, L. L. (2008). Cohabitation and children's living arrangements: New estimates from the United States. *Demographic Research, 19*(47), 1663–1692. https://doi.org/10.4054/DemRes.2008.19.47

Kennedy, S., & Fitch, C. A. (2012). Measuring cohabitation and family structure in the United States: Assessing the impact of new data from the Current Population Survey. *Demography, 49*, 1479–1498. https://doi.org/10.1007/s13524-012-0126-8

Kennedy, S., & Ruggles, S. (2014). Breaking up is hard to count: The rise of divorce in the United States, 1980–2010. *Demography, 51*, 587–598. https://doi.org/10.1007/ s13524-013-0270-9

Kreider, R. M. (2008). Improvements to demographic household data in the Current Population Survey. Retrieved from https://www.census.gov/population/www/do cumentation/twps08/twps08.pdf.

Kreider, R. M., & Ellis, R. (2011). Living arrangements of children: 2009. Retrieved from https://www.census.gov/prod/2011pubs/p70-126.pdf.

Krueger, P. M., Jutte, D. P., Franzini, L., Elo, I., & Hayward, M. D. (2015). Family structure and multiple domains of child well-being in the United States: A cross-sectional study. *Population Health Metrics, 13*, 6. https://doi.org/10.1186/s12963-015-0038-0

Künn, S. (2015). The challenges of linking survey and administrative data. IZA World of Labor. https://doi.org/10.15185/izawol.214.

Lery, B., Shaw, T. V., & Magruder, J. (2005). Using administrative child welfare data to identify sibling groups. *Children and Youth Services Review, 27*, 783–791. https://doi. org/10.1016/j.childyouth.2004.12.012

Lewis, J. M. (2013). Married couples are not the only type of two-parent family. Retrieved from U.S. Census Bureau website: https://www.census.gov/newsroom/blo gs/random-samplings/2013/11/married-couples-are-not-the-only-type-of-two-pare nt-family.html.

Lin, V.-K., Maxwell, K., & Forry, N. (2017). Considerations in preparing to analyze administrative data to address child care and early education research questions. Retrieved from https://www.acf.hhs.gov/sites/default/files/opre/cceepra_considera tionsforanalyzingadmindata_508compliant.pdf.

Livingston, G. (2014). Fewer than half of US kids live in "traditional" family. Retrieved from Pew Research Center website: https://www.pewresearch.org/fact-tank/2014 /12/22/less-than-half-of-u-s-kids-today-live-in-a-traditional-family/.

Lundberg, S., & Pollak, R. A. (2007). The American family and family economics (NBER Working Paper No. 12908). https://doi.org/10.3386/w12908.

Luque, A., & Wagner, D. (2015). Assessing coverage and quality of the 2007 Prototype Census Kidlink Database. Retrieved from https://www.census.gov/content/dam/C ensus/library/working-papers/2015/adrm/carra-wp-2015-07.pdf.

Madhavan, S., Clark, S., Beguy, D., Kabiru, C. W., & Gross, M. (2017). Moving beyond the household: Innovations in data collection on kinship. *Population Studies, 71*, 117–132. https://doi.org/10.1080/00324728.2016.1262965

Magnuson, K., & Berger, L. M. (2009). Family structure states and transitions: Associations with children's wellbeing during middle childhood. *Journal of Marriage and the Family, 71*, 575–591. https://doi.org/10.1111/j.1741-3737.2009.00620.x

Maloney, T., Jiang, N., Putnam-Hornstein, E., Dalton, E., & Vaithianathan, R. (2017). Black-White differences in child maltreatment reports and foster care placements: A statistical decomposition using linked administrative data. *Maternal and Child Health Journal, 21*, 414–420. https://doi.org/10.1007/s10995-016-2242-3

Manning, W. D., Brown, S. L., & Stykes, B. (2014). Family complexity among children in the United States. *The Annals of the American Academy of Political and Social Science, 654*, 48–65.

Manning, W. D., Joyner, K., Hemez, P., & Cupka, C. (2016). Measuring cohabitation in national surveys. Retrieved from https://paa.confex.com/paa/2017/mediafile/Exte ndedAbstract/Paper15168/wdm-paa-Measuring%20Cohabitation.pdf.

Manning, W. D., & Smock, P. J. (2005). Measuring and modeling cohabitation: New perspectives from qualitative data. *Journal of Marriage and Family, 67*, 989–1002. https://doi.org/10.1111/j.1741-3737.2005.00189.x

Massey, C. G. (2014). Creating linked historical data: An assessment of the Census Bureau's ability to assign protected identification keys to the 1960 Census. Retrieved from https://www.census.gov/content/dam/Census/library/working-papers/20 14/adrm/carra-wp-2014-12.pdf.

McCarthy, J. R., Doolittle, M., & Sclater, S. D. (2012). Understanding family and household through quantitative evidence. In J. Ribbens McCarthy, M. Doolittle, & S. Day Sclater (Eds.), *Understanding family meanings: A reflective text* (pp. 87–118). Bristol, England: Policy Press.

McFalls, J. A. (2003). What's a household? What's a family? Retrieved from htt ps://www.prb.org/whatsahouseholdwhatsafamily/.

McLanahan, S., & Percheski, C. (2008). Family structure and the reproduction of inequalities. *Annual Review of Sociology, 34*, 257–276. https://doi.org/10.1146/ annurev.soc.34.040507.134549

Meyer, B. D., Mok, W. K. C., & Sullivan, J. X. (2015). Household surveys in crisis. *Journal of Economic Perspectives, 29*, 199–226. https://doi.org/10.1257/jep.29.4.199

Miller, N. (1992). *Single parents by choice: A growing trend in family life*. New York, NY: Plenum Press.

Millett, E. R. C., Quint, J. K., De Stavola, B. L., Smeeth, L., & Thomas, S. L. (2016). Improved incidence estimates from linked vs. Stand-alone electronic health records. *Journal of Clinical Epidemiology, 75*, 66–69. https://doi.org/10.1016/j. jclinepi.2016.01.005

Mollborn, S., Fomby, P., & Dennis, J. A. (2011). Who matters for children's early development? Race/ethnicity and extended household structures in the United States. *Child Indicators Research, 4*, 389–411. https://doi.org/10.1007/s12187-010-9090-2

Mostafa, T., Gambaro, L., & Joshi, H. (2018). The impact of complex family structure on child well-being: Evidence from siblings. *Journal of Marriage and Family, 80*, 902–918. https://doi.org/10.1111/jomf.12456

Mykyta, L., & Macartney, S. (2011). The effects of recession on household composition: "Doubling up" and economic well-being (SEHSD Working Paper No. 2011–4). Retrieved from https://www.census.gov/content/dam/.../working-papers/.../SEHS D-WP2011-04.docx.

Office for National Statistics. (2017). Occupied address (household) estimates from administrative data: 2011 and 2015. Retrieved from https://www.ons.gov.uk/c ensus/censustransformationprogramme/administrativedatacensusproject/admin istrativedatacensusresearchoutputs/householdsandfamilies/occupiedaddresshouseh oldestimatesfromadministrativedata2011and2015.

O'Hara, A., Genadek, K., Medalia, C., & Alexander, T. (2018). Reconciling parent-child relationships across US administrative datasets. *International Journal of Population Data Science, 3*, 273. https://doi.org/10.23889/ijpds.v3i4.864

O'Hara, A., Shattuck, R. M., & Goerge, R. M. (2017). Linking federal surveys with administrative data to improve research on families. *The Annals of the American Academy of Political and Social Science, 669*, 63–74. https://doi.org/10.1177/0002716216678391

Ong, T. C., Mannino, M. V., Schilling, L. M., & Kahn, M. G. (2014). Improving record linkage performance in the presence of missing linkage data. *Journal of Biomedical Informatic, 52*, 43–54. https://doi.org/10.1016/j.jbi.2014.01.016

Ortman, J. M. (2007). Changes to the household relationship data in the Current Population Survey. Presented at the 2017 Applied Demography Conference, San Antonio, TX. Retrieved from https://www.census.gov/content/dam/Census/library/working-papers/2017/demo/SEHSD-WP2017-40.pdf.

O'Toole, M. T. (Ed.). (2013). *Mosby's medical dictionary* (9th ed.). St. Louis, MO: Elsevier.

Oya, C. (2015). Who counts? Challenges and biases in defining 'households' in research on poverty. *Journal of Development Effectiveness, 7*, 336–345. https://doi.org/10.1080/19439342.2015.1068358

Payne, K. (2018). Change in the U.S. remarriage rate, 2008 and 2016. https://doi.org/10.25035/ncfmr/fp-18-16.

Putnam-Hornstein, E., Ghaly, M., & Wilkening, M. (2020). Integrating data to advance research, operations, and client-centered services in California. *Health Affairs, 39*, 655–661. https://doi.org/10.1377/hlthaff.2019.01752

Putnam-Hornstein, E., Needell, B., & Rhodes, A. E. (2013). Understanding risk and protective factors for child maltreatment: The value of integrated, population-based data. *Child Abuse & Neglect, 37*, 116–119. https://doi.org/10.1016/j.chiabu.2012.08.006

Raley, R. K., & Bumpass, L. L. (2003). The topography of the divorce plateau: Levels and trends in union stability in the United States after 1980. *Demographic Research, 8*, 245–260. https://doi.org/10.4054/DemRes.2003.8.8

Raley, R. K., & Wildsmith, E. (2004). Cohabitation and children's family instability. *Journal of Marriage and Family, 66*, 210–219. https://doi.org/10.1111/j.0022-2445.2004.00014.x-i1

Roos, L. L., Brownell, M., Lix, L., Roos, N. P., Walld, R., & MacWilliam, L. (2008). From health research to social research: Privacy, methods, approaches. *Social Science & Medicine, 66*, 117–129. https://doi.org/10.1016/j.socscimed.2007.08.017

Ruggles, P. (1998). Longitudinal analysis of federal survey data (Joint Economic Committee Report No. 112). Retrieved from https://www2.census.gov/prod2/sipp/wp/SIPP_WP_112.pdf.

Schumm, W. R. (2012). Methodological decisions and the evaluation of possible effects of different family structures on children: The new family structures survey (NFSS). *Social Science Research, 41*, 1357–1366. https://doi.org/10.1016/j.ssresearch.2012.08.011

Seltzer, J. A. (2000). Families formed outside of marriage. *Journal of Marriage and Family, 62*, 1247–1268.

Sharma, R. (2013). The family and family structure classification redefined for the current times. *Journal of Family Medicine and Primary Care, 2*, 306–310. https://doi.org/10.4103/2249-4863.123774

Shorter, E. (1975). *The making of the modern family.* New York, NY: Basic Books.

Simon, A. (2014). Using administrative data for constructing sampling frames and replacing data collected through surveys. *International Journal of Social Research Methodology, 17*, 185–196. https://doi.org/10.1080/13645579.2012.733176

Sims, S., & O'Donnell, M. (2015). Identifying families most likely to have missing paternal details in birth registrations using linked data. *Journal of Family Studies, 21*, 163–177. https://doi.org/10.1080/13229400.2015.1020983

Smith, C., Crosnoe, R., & Cavanagh, S. (2017). Family instability and children's health. *Family Relations, 66*, 601–613. https://doi.org/10.1111/fare.12272

Sprague, W. W. (2017). *Constructing Measures of Household Characteristics from Administrative Data Sources and ACS: Similarities and Differences in Methods and Findings.* American Community Survey Data Users Group.

Stevens, L. A., & Laurie, G. (2014). The Administrative Data Research Centre Scotland: A scoping report on the legal & ethical issues arising from access & linkage of administrative data (Edinburgh School of Law Research Paper No. 2014/35).

Retrieved from https://www.phrn.org.au/media/80636/the-adrc-scotland-a-scoping-report-on-the-legal-and-ethical-issues.pdf.

Suei, S. (2016). Comparing legally registered relationship status information from census and admin data [Unpublished document].

Sweeney, M. M. (2010). Remarriage and stepfamilies: Strategic sites for family scholarship in the 21st century. *Journal of Marriage and Family, 72*, 667–684.

Sweeting, H., & Seaman, P. (2005). Family within and beyond the household boundary: Children's constructions of who they live with. In L. McKie, & S. Cunningham-Burley (Eds.), *Families in society: Boundaries and relationships* (pp. 95–110). Bristol, England: Policy Press.

Tan, H., Wen, S. W., Walker, M., & Demissie, K. (2004). Missing paternal demographics: A novel indicator for identifying high risk population of adverse pregnancy outcomes. *BMC Pregnancy and Childbirth, 4*, 21. https://doi.org/10.1186/1471-2393-4-21

The Lewin Group (2008). Collection of marriage and divorce, statistics by states: Final report. Retrieved from https://www.acf.hhs.gov/sites/default/files/opre/coll_stats.pdf.

Tingay, K., Anastasiou, A., & Musselwhite, C. (2018). A longitudinal, historical model of households using linked Welsh administrative data. *International Journal of Population Data Science, 3*, 367. https://doi.org/10.23889/ijpds.v3i4.962

United Nations (1994). The family, its roles, composition and structure. Retrieved from https://www.un.org/popin/icpd/prepcomm/official/rap/RAP4.html.

U.S. Census Bureau (2018). Questions planned for the 2020 Census and American Community Survey. Retrieved from https://www2.census.gov/library/publications/decennial/2020/operations/planned-questions-2020-acs.pdf.

U.S. Census Bureau (2019). Subject definitions. Retrieved from https://www.census.gov/programs-surveys/cps/technical-documentation/subject-definitions.html.

U.S. Department of Agriculture (2018). Supplemental Nutrition Assistance Program (SNAP) FY15-FY 18 annual view summary. Retrieved from https://www.fns.usda.gov/pd/supplemental-nutritio-assistance-program-snap.

U.S. Department of Agriculture (2019). SNAP eligibility. Retrieved March 10, 2020, from https://www.fns.usda.gov/snap/recipient/eligibility.

Victor, T. W., & Mera, R. M. (2001). Record linkage of health care insurance claims. *Journal of the American Medical Informatics Association, 8*, 281–288. https://doi.org/10.1136/jamia.2001.0080281

Waldfogel, J., Craigie, T.-A., & Brooks-Gunn, J. (2010). Fragile families and child wellbeing. *The Future of Children, 20*, 87–112. https://doi.org/10.1353/foc.2010.0002

Wang, H. L. (2018). 2020 Census will ask about same-sex relationships. NPR. Retrieved from https://www.npr.org/2018/03/30/598192154/2020-census-will-ask-about-same-sex-relationships.

Wetzel, J. R. (1990). American families: 75 years of change. *Monthly Labor Review, 113*, 4–13.

Whitley, E., Deary, I. J., Der, G., Batty, G. D., & Benzeval, M. (2012). Paternal age in relation to offspring intelligence in the West of Scotland Twenty-07 prospective cohort study. *PLOS One, 7*, Article e52112. https://doi.org/10.1371/journal.pone.0052112

Widmer, E. D. (1999). Family contexts as cognitive networks: A structural approach of family relationships. *Personal Relationships, 6*, 487–503. https://doi.org/10.1111/j.1475-6811.1999.tb00205.x

Widmer, E. D. (2006). Who are my family members? Bridging and binding social capital in family configurations. *Journal of Social and Personal Relationships, 23*, 979–998. https://doi.org/10.1177/0265407506070482

Wisselgren, M. J., Edvinsson, S., Berggren, M., & Larsson, M. (2014). Testing methods of record linkage on Swedish censuses. *Historical Methods, 47*, 138–151. https://doi.org/10.1080/01615440.2014.913967

Wu, L. L., & Martinson, B. C. (1993). Family structure and the risk of a premarital birth. *American Sociological Review, 58*, 210–232. https://doi.org/10.2307/2095967

Zhu, Y., Matsuyama, Y., Ohashi, Y., & Setoguchi, S. (2015). When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *Journal of Biomedical Informatics, 56*, 80–86. https://doi.org/10.1016/j.jbi.2015.05.012

Zill, N., & Daily, M. (1993). *Researching the family: A guide to survey and statistical data on U.S. families.* Washington, DC: Child Trends.