

# The Utah Population Database. The Legacy of Four Decades of Demographic Research

By Ken R. Smith and Geraldine P. Mineau

To cite this article: Smith, K. R. & Mineau, G. P. (2021). The Utah Population Database. The Legacy of Four Decades of Demographic Research. *Historical Life Course Studies*, 11, 48-73. <https://doi.org/10.51964/hlcs10916>

## HISTORICAL LIFE COURSE STUDIES

Major Databases with Historical Longitudinal Population Data:  
Development, Impact and Results

VOLUME 11, SPECIAL ISSUE 4,  
2020

GUEST EDITORS  
Sören Edvinsson  
Kees Mandemakers  
Ken Smith



## MISSION STATEMENT

# HISTORICAL LIFE COURSE STUDIES

*Historical Life Course Studies* is the electronic journal of the *European Historical Population Samples Network* (EHPS-Net). The journal is the primary publishing outlet for research involved in the conversion of existing European and non-European large historical demographic databases into a common format, the Intermediate Data Structure, and for studies based on these databases. The journal publishes both methodological and substantive research articles.

### Methodological Articles

This section includes methodological articles that describe all forms of data handling involving large historical databases, including extensive descriptions of new or existing databases, syntax, algorithms and extraction programs. Authors are encouraged to share their syntaxes, applications and other forms of software presented in their article, if pertinent, on the EHPS-Net website.

### Research articles

This section includes substantive articles reporting the results of comparative longitudinal studies that are demographic and historical in nature, and that are based on micro-data from large historical databases.

*Historical Life Course Studies* is a no-fee double-blind, peer-reviewed open-access journal supported by the European Science Foundation (ESF, <http://www.esf.org>), the Scientific Research Network of Historical Demography (FWO Flanders, <http://www.historicaldemography.be>) and the International Institute of Social History Amsterdam (IISH, <http://socialhistory.org/>). Manuscripts are reviewed by the editors, members of the editorial and scientific boards, and by external reviewers. All journal content is freely available on the internet at [hlcs.nl](http://hlcs.nl).

### Co-Editors-In-Chief:

Paul Puschmann (Radboud University) & Luciana Quaranta (Lund University)  
[hislives@kuleuven.be](mailto:hislives@kuleuven.be)

**The European Science Foundation** (ESF) provides a platform for its Member Organisations to advance science and explore new directions for research at the European level. Established in 1974 as an independent non-governmental organisation, the ESF currently serves 78 Member Organisations across 30 countries. EHPS-Net is an ESF Research Networking Programme.



**The European Historical Population Samples Network** (EHPS-net) brings together scholars to create a common format for databases containing non-aggregated information on persons, families and households. The aim is to form an integrated and joint interface between many European and non-European databases to stimulate comparative research on the micro-level.  
Visit: <http://www.ehps-net.eu>.



HISTORICAL LIFE COURSE STUDIES  
VOLUME 11 (2021), published 14-09-2021

# The Utah Population Database

## The Legacy of Four Decades of Demographic Research

Ken R. Smith  
University of Utah

Geraldine P. Mineau  
University of Utah

### ABSTRACT

This paper summarizes the unique characteristics of the Utah Population Database (UPDB) and how it has catalyzed demographic, social and medical research since the mid-1970s. The UPDB is one of the world's richest sources of linked population-based information for demographic, genetic, and epidemiological studies at the Individual-level. UPDB has supported hundreds of demographic and biomedical investigations, with heavy emphasis on families, in large part because of its size, representativeness, inclusion of multi-generational pedigrees, and linkages to numerous data sources. The UPDB contains data on over 11 million individuals from the late 18th century to the present. UPDB data represent Utah's population that appear in administrative records and many of these data are updated due to longstanding efforts to add records as they become available including statewide birth and death certificates, hospitalizations, ambulatory surgeries, and driver licenses. The depth of information within UPDB has been used to support a wide range of family, medical and historical demographic studies which are described here arranged into four broad categories: fertility, mortality, life course analyses and some selected special topics. The paper concludes with a discussion of the future areas of innovation within the UPDB and the types of novel studies that they are likely to facilitate.

**Keywords:** Demography, Multigenerational families, Mortality, Fertility, Databases, Life course

e-ISSN: 2352-6343  
DOI article: <https://doi.org/10.51964/hlcs10916>

© 2021, Smith, Mineau

This open-access work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/), which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit. See <http://creativecommons.org/licenses/>.

# 1 INTRODUCTION

Understanding the sources of variation in human health and well-being for individuals and families throughout history is a central goal of the social, biological and medical sciences. One strategy for contributing to this understanding is to capture and curate extensive data on large, well-defined populations and all of its members over time. What is most desirable for the research community is to obtain wide-ranging data from 'DNA-to-Demography' or 'Proteins-to-Pedigrees'. Gathering and curating big data for social and health-related research have been successful in several instances, many of which have been fundamental contributors to key medical and social science discoveries. The impact of this data collection strategy will be described in more detail in the following pages but it is noteworthy that organizing historic and contemporary demographic and medical information that incorporate quality controls and centralized management benefit many in the research community. Centralized oversight and careful supervision of data that are made accessible to the research community serve to fuel innovative research with wide ranging impact across many disciplines.

Here, we describe the **Utah Population Database (UPDB)** and specifically its scientific impact, primarily to the social sciences and demography, which offers exceptional and unique data and research opportunities for population scientists, demographers, epidemiologists, geneticists, health services researchers, behavioral scientists, among others, all of whom work on population health and medical research. The distinctive quality of the UPDB is that links individual-level administrative and medical records derived from a range of sources spanning decades for some sources and centuries for others (Casey, Schwartz, Stewart, & Adler, 2016; Hurdle, Smith, & Mineau, 2013). The UPDB is a research resource that has been expanded extensively in its 40 years of existence. At this time, UPDB includes information on over 11 million individuals who have basic demographic information and have been a data source for nearly 400 research projects. The time period covers birth cohorts from the 1700s but are more extensive starting in the mid-1800s through the present. The UPDB is in the same genre as other large data projects in the US and UK such as the ongoing longitudinal Framingham Heart Study and the National Survey of Health and Development which are massive in scope, years of coverage, and depth of information.

## 1.1 CONTENT OF THE UPDB

The visionaries who established UPDB over 40 years ago sought to integrate genetics and the social sciences. Its beginnings, during the years 1973–1974, started when several researchers at the University of Utah realized the research opportunities that could be gained by first obtaining extensive genealogy records and constructing a population-based resource that would link these data to high quality medical records in order to investigate the genetic basis of a number of important diseases. Central to this effort was geneticist Mark Skolnick, who led a consortium of key scientists including, cardiologist Roger Williams and demographer Lee L. Bean.

The original set of genealogy records used when the UPDB was being developed comprised approximately 185,000 documents representing, on each form, three generations: a husband-wife pair, their four parents, and the couple's offspring and their respective spouses. These initial documents were selected to represent approximately 1.9 million individuals. Linking these across generations created thousands of multi-generational pedigrees, providing astonishing insights regarding the population (Song & Campbell, 2017).

These early genealogical records comprise the original backbone of UPDB. The founding research team secured access to the Utah Cancer Registry (UCR, a Surveillance, Epidemiology and End Result [SEER] Registry) and Utah death certificates (from the Utah Department of Health) as the basis for medical outcomes to be linked to the genealogies at the individual level.

While the original development of UPDB was derived from genealogy, cancer and death records, the UPDB now includes substantially more records and from diverse sources (see Table 1). In addition to the original genealogy records, UPDB includes (1) All available Utah vital records (births, deaths, marriages, divorces and fetal deaths), (2) Statewide health facilities data including Inpatient Hospital Discharge, Emergency Department, Ambulatory Surgery records from the Utah Department of Health, (3) Health insurance claims data ('All Payer Claims') (4) Utah cancer registry, (4) Social Security death index records, (6) Utah Voter Registration, (7) Utah driver license records, and (8) The 1880 and 1900–1940 Utah individual-level censuses and other data sets from the Department of Health.

Table 1 Overview of key data sources comprising the Utah Population Database

Record Type	Years Available	Records
Original Family History Records	1700's–1975	1,916,649
Birth Certificates	1915–1921, 1926–present	3,067,098
Death Certificates	1904–present	921,081
Marriage Certificates Utah	1978–2010	692,680
Divorce Records Utah	1978–2010	298,928
Fetal Deaths Utah	1978–present	11,434
Ambulatory Surgery Utah	1996–present	7,255,110
Inpatient Hospital Claims Utah	1996–present	6,124,266
Emergency Department	1996–present (in process)	9,716,843
All Payer Claims Data	2013–present	3,561,630
Utah Cancer Registry	1966–present	394,437
Utah Birth Defect Network	1995–2015	21,235
U.S. Census of Utah	1880,1900–1940	2,300,084
Driver License Utah	Updated annually since 1990	3,997,558
Voter Registration Utah	Updated after national elections since 2008	2,068,916

Two distinctive features of UPDB are noteworthy with respect to linkages to other large medical data sets. First, links have been created connecting UPDB with the data warehouses of the two largest health providers in Utah — University of Utah Health and Intermountain Healthcare (DuVall, Fraser, Rowe, Thomas, & Mineau, 2012). These two health care providers represent inpatient and outpatient electronic medical information for approximately 85% of the state's medical encounters starting from the mid-1990s. The medical data per se are not held within the UPDB but are securely maintained by the data warehouses of the providers. Medical data are joined with the demographic and genealogical data in UPDB after the research project receives the necessary approvals from appropriate Institutional Review Boards and the Utah Resource for Genetic and Epidemiologic Research (RGE), which oversees research access to the UPDB.

A second and related medical data linked to the UPDB are those derived from Medicare claims. The Medicare data are available due to funding from National Institutes of Health (NIH) grants in order to facilitate the study of healthy aging and health expectancy among the Medicare-eligible population. These data are available to researchers using the UPDB but they must not only obtain IRB approval for their use but also approval from the Centers for Medicaid and Medicare Services (CMS).

Linking all the records as shown in Table 1 within the UPDB creates unique research opportunities including:

1. Creation of reproductive histories. Using data from the Utah Department of Health that includes Utah birth certificates from 1915 to the present, genealogical holdings of UPDB have been extended considerably. Mothers and fathers on multiple birth certificates are linked. This allows us to see that specific individuals share common parents and are therefore siblings. The children named on these birth certificates are then linked to the birth certificates of their children. Because birth certificates provide gestational age and birth weight as well as other features such as adverse obstetric events and birth complications, this strategy has provided a valuable source for analysis of obstetric health in families and across generations (Hammad et al., 2020; Theilen et al., 2016; Theilen et al., 2018). Many of the genealogies derived from vital records also link into the legacy genealogies that are part of the UPDB.
2. Creation of residential exposures and histories. Location information is derived from several sources in UPDB including Driver License Division (DLD) data, voter registrations, and vital and birth records. One use of DLD is to determine if an individual is currently under observation, while residence information on death records verify an individual was under observation until their death. Voter Registration records are obtained and linked to UPDB which give location information at a particular

point in time. Additionally, DLD data hold information on height and weight from which Body Mass Index (BMI) for each individual can be derived (Chernenko, Meeks, & Smith, 2019; Smith et al., 2008; Smith et al., 2011; Zick et al., 2009). Finally, residential histories within UPDB have been geo-coded which create the opportunity for linking any geo-referenced data set (e.g., census block and air quality monitors) with individual-level data. In general, for contemporary decades where address information was captured, UPDB generates location data down to the Census block-group level which are statistical divisions of US census tracts and have 600 and 3,000 people. Of course, higher level areas of aggregation are also available. For more historic years, Census Enumeration Districts are provided or place names (city, county).

3. **Individual-Level Census Records.** The addition of the micro level census records from 1880 and 1900–1940 to UPDB allows for several types of studies. First, it is now possible to observe mobility, both geographic and socioeconomic, and its causes and consequences. Second, given the manner in which census enumerators were assigned to districts to conduct the full count of the population, the data are arranged into neighborhoods, as noted, represented as Enumeration Districts. Accordingly, individuals identified in the census can be characterized by the quality of their 'neighborhoods' and how these spatial attributes may alter later life outcomes. Finally, these census records provide valuable independent information about family composition, co-residence, and genealogical data that may not be possible from other sources of data in the UPDB.
4. **High Risk Pedigrees and Gene Discovery.** The initial motivation for creating the UPDB stemmed from the concept of joining multigenerational information to medical records in order to facilitate the identification of genes involved in disease risk. With this information, cancer in the early (and present) era of UPDB, and now almost any condition found in electronic medical records, provides scientists with the opportunity to identify families with an excess disease risk (or traits that are positive such as longevity). Family-based studies have an advantage in providing more information from which to identify how genes segregate among related individuals who are or are not affected by the disease. UPDB has provided the family and disease information from which many gene discoveries are based including genes for breast cancer (BRCA1, BRCA2), melanoma (p16), and colon cancer (APC).
5. **Link to Existing Cohorts.** UPDB also has the capacity to link its data to ongoing projects that have arisen independent of UPDB. For example, the Cache County Memory and Health Study was launched in 1995 to study factors related to dementia and Alzheimer's disease risk. These persons were 65 and older at enrollment and were from a single county in Northern Utah. This linkage between an existing cohort and the UPDB provided an opportunity to open up new life course studies of dementia and Alzheimer's disease (Norton et al., 2010, 2011, 2016).

## 1.2 RESOURCE FOR GENETIC AND EPIDEMIOLOGIC (RGE) RESEARCH

Access to UPDB data is regulated by the Utah Resource for Genetic and Epidemiologic Research (RGE). The RGE was created by an Executive Order of the Governor of Utah in 1982. Relying on enabling statutes in state health code, the RGE was established as a "data resource for the collection, storage, study, and dissemination of medical and related information" to operate "for the purpose of reducing morbidity or mortality, or for the purpose of evaluating and improving the quality of hospital and medical care". Originally administered under the direction and supervision of the Utah Department of Health, the RGE was transferred to the University of Utah by a second Executive Order in 1986. RGE is the legal custodian for the data contained within the UPDB and is responsible for developing and maintaining contractual agreements with organizations that contribute data to the UPDB or that links records to the UPDB.

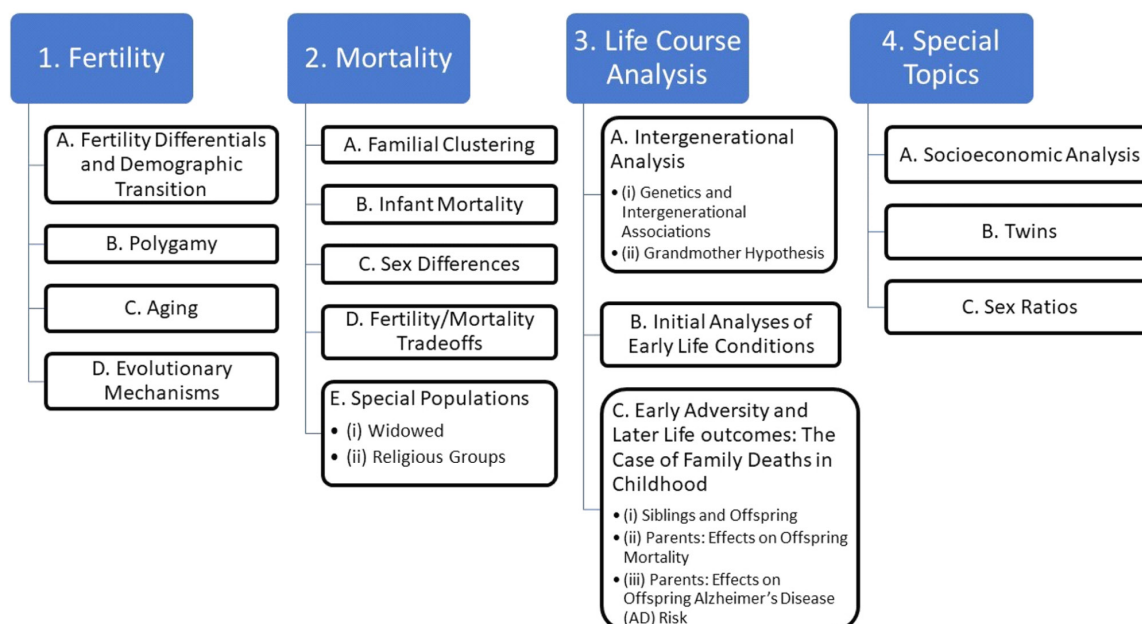
Each project requesting access to data from the UPDB or linked electronic medical records applies to RGE for review. Applications are reviewed by the RGE Committee, which includes University faculty with expertise in several disciplines including demography, genetics, public health and epidemiology, as well as representatives from each of the data contributors. All projects are required to obtain approval by the appropriate Institutional Review Board(s) and Privacy Board(s) before access to data is granted. Ultimately, RGE has the responsibility to protect the sensitive confidential information in UPDB.

## 1.3 IMPACT OF UPDB ON RESEARCH DOMAINS IN DEMOGRAPHY

The birth and maturation of UPDB as a unique data resource was initially developed to promote both demographic and genetics research. The summary of scientific impacts of the UPDB presented here focuses on demographic research and on publications where demographic and genetic elements are combined.

The number of publications in this realm are substantial and we do not attempt to include all publications. Instead, the focus is on publications that exemplify the types of research in four broad demographic domains that represent the value of UPDB to these research endeavors. Understandably, some investigations span more than one of these domains. Accordingly, highlights of a given publication may get attention in more than one section. The organization of the four broad domains and their sub topics are summarized in a roadmap in Figure 1 and this provides the structure of the review that follows.

Figure 1 UPDB Research Domains



## 2 FERTILITY

One of the earliest sets of questions behind the development of the UPDB was born from issues about fertility and how reproduction forms the foundation for the study of genetics and inheritance. Given the settlement history of the Utah territory in the mid-19th century and the migration of adherents to the Church of Jesus Christ of Latter-day Saints, initial research projects focused on the early pattern of natural fertility, and the study of fertility and its change during the demographic transition.

### 2.1 FERTILITY DIFFERENTIALS AND DEMOGRAPHIC TRANSITION

The early years of the UPDB focused on fertility patterns, launched with a two-part analysis of 'Mormon demographic history'. The first examined nuptiality and fertility among once-married couples (Skolnick et al., 1978) and a second on 'the family life cycle and natural fertility' (Mineau, Bean, & Skolnick, 1979). Skolnick and his colleagues motivated their analysis with a recognition that the Church of Jesus Christ of Latter-day Saints had a history of polygamy in the 19th century and endorsed pro-natalist policies all of which led to higher fertility than the nation at large at that time. While these patterns have been known for some time, they argued that the specific characteristics of this fertility behavior in Utah had been undeveloped. In particular, during the early settlement years (in the two decades after 1847, when white settlements in Utah started) marriage ages declined and the proportion of women marrying as teens increased. During the middle of the 19th century, little fertility control was practiced and age at marriage was a major determinant of total and completed fertility. The period 1870–1880 shifted demographic history of Utah when the transcontinental railway opened, exposing Utah residents to greater influence of the secular world. This was also the decade where church members were facing pressure from non-church members in Utah and elsewhere to renounce the practice of polygamy.

The second of these paired papers by Mineau and collaborators (1979) drew on the debate in the 1960's and 1970's about the American family life cycle which was largely based on women born during and after the 1880s and presumably portrayed a 'modern' pattern (i.e., declining and controlled fertility, lower rates

of infant mortality and improved parental life expectancy). They examined a unique American population (Utah) and the opportunity it provided to study the life-cycle under conditions of natural fertility (pre-1870). They extended the study of the family life-cycle to US birth cohorts (1800–1869). In comparisons of the family life cycle at the time, Mineau et al. showed differences in the number of years between the marriage of the last-born child and the age at family dissolution (the 'empty nest' phenomenon). Using micro-data from the UPDB (rather than life table parameters), Mineau et al. concluded that the increasing number of years attributed to the 'empty nest' phenomenon probably has less to do with increased survival rates of parents and more to the interaction between declining fertility levels and decreasing age of mother at birth of the last child.

In a series of papers, Anderton led several studies that provided new insights into fertility limitation and transitions after the early natural fertility settlement years. In 1984, Anderton and colleagues (Anderton, Bean, Willigan, & Mineau, 1984) concluded that during the period of natural fertility, commitment to a pronatalist faith helped to account for differences in fertility levels, but not patterns of fertility decline over time. Factors typically associated with fertility decline in Western Europe such as consistent long-term secularization and urbanization, were found to be more important determinants of cross-sectional fertility levels and longitudinal changes in fertility levels across the birth cohorts studied. This question was extended to consider the role of birth spacing and fertility transitions (Anderton & Bean, 1985). In this analysis they hypothesized that substantial groups of women with long birth intervals could be identified — even during periods when fertility behavior at the aggregate level is consistent with a natural fertility regime. Specifically, they concluded that birth spacing patterns are highly parity-dependent and that the transition is associated with a larger proportion of women shifting to the same spacing schedules associated with smaller families in earlier cohorts. They also showed that changes in birth intervals over time are indirectly associated with age of marriage and that there is evidence of efforts to terminate child-bearing (i.e., shifts in stopping behavior). Overall, they emphasized the importance of distinguishing between spacing and stopping behavior. Later, Hsueh and Anderton (1990) evaluated age, period, and cohort effects on marital fertility during the onset of the Utah fertility transition (1880–1900). They concluded that declining marital fertility in Utah can be explained by both declining fertility levels across historic periods and increasing age-specific limitation across cohorts. They argued that fertility levels were adaptive (through birth spacing across ages) to immediate contexts of childbearing while age-specific fertility truncation increased across cohorts (through diffusion of contraceptive innovations).

These earlier analyses were accompanied by additional foci on central demographic questions regarding fertility. We consider three additional specific domains: the role of polygamy, aging and reproductive senescence, and analyses motivated by evolutionary theories of fitness.

## 2.2 POLYGAMY

In early Utah history, the Church of Jesus Christ of Latter-day Saints supported plural marriages though it was renounced later in the 19th century as a basis for achieving statehood (Ellsworth, 1963; Lyman, 1998). This fact has been the basis for several important historical demographic analyses. Bean and Mineau examined plural marriages in the context of fertility (Bean & Mineau, 1986). They specifically addressed the relationship between polygynous marriages and fertility which the literature tends to show that woman-specific fertility levels are lower in polygynous than in monogamous marriages. Using data for 2,534 polygynists with 7,378 marriages and comparing fertility with once-married monogamous women, they found significant differences in fertility levels by the order of the plural wife. They note that in most polygynous cultures more sister wives marry-in as the husbands increase in status, and therefore husbands will be older at the time the second and later wives are added. With increasing age of the patriarch, widowhood is also more likely so that the risk of pregnancy is affected by wife-order. With UPDB, the average level of fertility of all polygynous wives is lower than that of monogamous wives. This is due to a significantly lower fertility among second and later wives than among monogamous wives, while the fertility of the first wife in polygynous families is higher.

Moorad published a series of papers (Moorad, Promislow, Smith, & Wade, 2011; Moorad, 2013; Moorad & Wade, 2013) using UPDB in which he addresses the role that polygyny plays in sexual selection. Sexual selection is the competition of one sex for reproductive access to the other. Moorad was interested in the hypothesis that sexual selection is stronger in polygamous than in monogamous species. Moorad and his colleagues (2011) found, for example, that over the reproductive lifetimes of Utahns born between 1830–1894, reductions in the rate and degree of polygamy was associated with a 58% reduction in the strength of sexual selection. Polygyny conferred a strong advantage to male fitness (more progeny who survive) as well as a weak disadvantage to female fitness. In contrast, mating with multiple males provided little benefit to females in this population.



## 2.3 AGING

It is axiomatic that the time of marriage and first birth have a profound effect on subsequent fertility. In the US, it has been difficult to show how this pattern plays out in the context of natural fertility. Mineau and Trussell (1982) examined these associations using birth cohorts in UPDB from 1840 to 1879. They showed that older-aged husbands depress marital fertility only at higher marriage durations. They also demonstrated that mother's aging is the most important factor, while father's aging has a moderately negative effect under a natural fertility regime.

The UPDB has been used to compare fertility in relation to other historical databases spanning periods of natural fertility. The most notable example is a large international comparison of reproductive behavior across several populations: The Utah Population Database (UDBP), the *Registre de la population du Québec Ancien* at the Université de Montréal, Canada (RPQA), the LINKS database, hosted by the International Institute of Social History in Amsterdam, the German database based on *Ortssippenbuch* ('book of local kinsmen'), the BALSAC demographic database representing the Saguenay-Lac-St-Jean (SLSJ) region in Québec, and the database collected by the French demographer Louis Henry of the French population between 1670 and 1830 (Eijkemans et al., 2014). The question was whether it was possible to quantify the ages after which women are biologically unable to reproduce? The analysis was motivated by the fact that little is known about the distribution of female age at last birth (ALB) especially now with the widespread availability of modern birth control. The six natural fertility populations comprised nearly 60,000 women. Eijkemans and his colleagues showed that while these populations represent different historical time periods and cultural contexts, the distribution of ages at last birth is quite similar. The curve denoting the end of fertility indicates that <3% of women had their last birth at age 20 years and that about 50% had their age at last birth by age 41, almost 90% by age 45 years and approaching 100% at age 50 years.

## 2.4 EVOLUTIONARY MECHANISMS

The availability of high-quality demographic and life history data spanning decades consistent with natural fertility has attracted the attention of evolutionary biologists and anthropologists. Two analyses are highlighted here that exemplify this line of inquiry.

Jones and Bliege Bird (2013) examined a fundamental hypothesis that argues that fitness, that is reproductive success, should be maximized by an intermediate level of fertility. This prediction has not been widely supported in the human life-history literature and they contend that the difficulty of finding this intermediate reproductive optimum may be a measurement issue. Rather than using lifetime reproductive success as the fitness measure, they proposed a measure that accounts for *variation* in reproductive timing which better reveals preferences about when women are making risky reproductive decision-making. Using UPDB and data from 19th century, they demonstrate that if births are properly timed, a lower-fertility reproductive strategy can have the same fitness as a high-fertility strategy.

Gagnon and collaborators (2009) also considered tradeoffs using frontier populations: UPDB, the *Registre de la population du Québec ancien* (Université de Montréal), and the BALSAC database (Université du Québec à Chicoutimi from the Saguenay-Lac-St-Jean (SLSJ) in Québec. These data provided exceptional opportunities to test the hypothesis of a trade-off between fertility and longevity. Together, these databases allow for comparisons over time and space, and represented one of the largest comparisons of natural fertility cohorts to simultaneously assess reproduction and longevity. They observed a negative influence of parity (more children, worse survival) and a positive influence of age at last birth (ALB; later ALB, better survival) on postreproductive survival in the three populations, as well as a significant interaction between these two variables, patterns that were remarkably similar in the three samples. Strong support was therefore found for a trade-off between parity and longevity and a strong moderating influence of ALB.

## 2.5 FERTILITY SUMMARY

The UPDB has contributed to our understanding of fertility, the forces influencing its variation, and what it tells us about post reproductive consequences. While the volume of studies is considerable, we suggest that UPDB has, in general, provided an opportunity to study in great detail the shifts in fertility for 200 years, at the individual, family, pedigree, and community levels. The historical reach of the data, starting from the early settlement pioneer era, have attracted the attention of anthropologists and biologists, as well as historical demographers, interested in testing evolutionary hypotheses best conducted on humans during years approximating natural fertility conditions. At the other extreme, these historical data connect to present day Utahns such that the potentially enduring effects of past fertility behavior can be assessed.

This feature has led to collaborations with obstetrics and gynecology colleagues who are asking previously underexplored questions that require historical perspectives and data. Some unique components of UPDB have allowed insights into fertility behavior including plural marriage, how demographic change happens when a population moves into a sparsely inhabited geography and how it creates the conditions for a society to grow in the following decades.

## 3 MORTALITY

A powerful advantage of the UPDB is its coverage of key life history and life course traits. The strategy to capture birth, marriage, and death information from a variety of sources including church records, vital registration and medical sources provides an opportunity to study mortality patterns and differentials over time for the Utah population with triangulation across diverse types of data. With death registration beginning in Utah in 1904, UPDB now holds over 100 years of cause of death information all coded to the International Classification of Diseases. Given that mortality information is linked to UPDB's vast genealogical data, many studies have explored familial forces that affect and are affected by the timing and causes of death among family members. Here we summarize selected key contributions of UPDB data applied to questions related to mortality.

### 3.1 FAMILY CLUSTERING

It is well-established that mortality risks are often shared among nuclear and extended family members. The UPDB has been the basis for developing methods which are then applied to increase our understanding of mortality risks beyond individual level measures. One of the earliest and influential methods was introduced by Kerber (1995). The innovation here was to develop a method for estimating excess relative risks of mortality or disease to an individual that could be due to familial factors. Using UPDB, all relatives of a person are identified and measured for the presence and timing of a disease or a cause of death. The count of disease X or cause-of-death Y among the relatives of the person is compared to a count of what would be expected if those relatives had experienced the population risks, given their age and sex. Using exposure years (to generate incidence rates) and adjustments for how closely related a person is to each of their relatives, the method produces a measure similar to a relative risk (which compares observed versus expected rates) called the familial standardized incidence ratio (FSIR) or familial standardized mortality ratio (FSMR). For example, for a study of suicide, a significant FSIR > 1 means that for people who died of suicide, they have relatives who were more likely to die of suicide than would be expected based on the prevailing population suicide rates, an indication of familial clustering.

This technique has been expanded and applied to several mortality questions. Kerber et al. (2001) evaluated the influence of family history of longevity by examining longevity in a cohort of 65+ individuals drawn from UPDB born between 1870–1907. Using the logic of the FSIR applied to longevity, resulting in a measure called Familial Excess Longevity (FEL), they showed that excess longevity aggregates in families, and that the presence of familial aggregation of longevity is a powerful predictor of longevity a given person. O'Brien and colleagues (2007) considered the effects of familial longevity and familial mortality on mortality rates for 10 leading causes of death. FEL and FSMR were estimated for 666,921 individuals over 40 born from 1830 through 1963. They showed that a family history of disease increases the risk of dying from the same cause, whereas a family history of longevity is protective for most age-related diseases including heart disease, stroke, and diabetes, but not cancer (Kerber, O'Brien, Smith, & Mineau, 2008).

The UPDB has been used when applying another measure of familiarity of disease or mortality, the Genealogical Index of Familiarity or GIF (Cannon Albright, 2008). The idea is simply to identify the genetic relationships between all pairs of individuals with the same disease or cause of death and to then estimate the average relatedness of these individuals. This uses the (Malécot) coefficient of kinship to measure the relatedness of each pair of cases between individuals sharing a cause of death, for example. This is repeated for matched controls for the average relatedness one would expect in the general population reflected in UPDB. If the average relatedness of a set of people dying from a given cause is significantly higher than the mean relatedness from a set of matched controls, there is support for excess familiarity for this cause of death though it may represent genetic or environmental forces. GIF have been used to study familial clustering of disease and mortality using UPDB, for example, Alzheimer's disease and coronary heart disease (Horne, Camp, Muhlestein, & Cannon-Albright, 2006; Kauwe, Ridge, Foster, & Cannon-Albright, 2013).

Recently, Van den Berg et al. (2019) have proposed a simple familiarity rule. Using data from two sources, UPDB and LINKS (the Netherlands), they provide strong evidence that longevity is transmitted as a quantitative genetic trait among survivors up to the top 10% of their birth cohort. They also show that if you have first and second-degree relatives who survive to the top 10%, even if your own parents are not longevous, then you enjoy a survival advantage.

A large consortium of universities comprises the Long Life Family Study (LLFS) where they have developed the Family Longevity Selection Score (FLoSS). This score measures familiarity of longevity and was used to select families for LLFS but had not been validated in other populations. The LLFS team computed FLoSS using the lifespan data of 234,155 individuals from UPDB, born between 1779 and 1910 with mortality follow-up through 2012–2013. They reported (Arbeeva et al., 2018) that in UPDB those born after 1900 from 'exceptional' sibships based on the FLoSS had survival curves similar to that of the US participants from comparable LLFS probands. In this way, UPDB served as a basis from which others can test their methods of detecting familiarity of mortality. This study validated the FLoSS as selection criteria in family longevity studies using UPDB.

Clustering of mortality within families also aggregate with other important life history traits including fertility, in particular late female fertility. Women giving birth at advanced reproductive ages in natural fertility conditions have been found to have superior post-menopausal longevity (Smith, Mineau, & Bean, 2002). To determine if survival also improved for relatives of late-fertile women, Smith and colleagues (2009) compared male survival past age 50 for those with and without a late-fertile sister in two populations: UPDB (born 1800–1869) and the Programme de recherche en démographie historique in Québec (born 1670–1750). They reported improved male survival for those with, rather than without, a sister reproducing after age 45, suggesting that late female fertility and slower rates of aging may be promoted by similar genes. This work demonstrates again how UPDB promotes comparative work with other populations and in different eras.

The use of summary measures of familiarity is common using UPDB but UPDB has also advanced the application of statistical models that exploit the genealogies in UPDB but without summary estimates like FSIR or GIF. Garibotti and collaborators (2006) used pedigrees to assess the effects of unobserved environmental and genetic effects on longevity or so-called frailty. With UPDB they used two different frailty models that account for common environments (shared frailty) and genetic effects (correlated frailty). In a model that includes summary measures of familial history of longevity and both types of frailty effects, they found that genetic factors were comparable in their effects to shared environments.

### 3.2 INFANT MORTALITY

The detailed coverage of the population embodied within the UPDB facilitates the study of entire lives, from gravida to grave. It is equally important to investigate specific, primarily susceptible ages where mortality is highest: for infants and for those past mid-life. Lynch et al demonstrated during the early years of the UPDB, the power of genealogies for the study of infant mortality (Lynch, Mineau, & Anderton, 1985). They argue that the demographic study of declines in infant mortality during the last half of the 19th century was hampered by limited data at the individual-level; the US vital record system was not truly national until the 1930s. Accordingly, regional data served as the basis for the study of infant death. Here, they advocated for a form of family reconstitution drawn from the then-named Mormon Historical Project. This idea of using regional data as represented by the now-named UPDB was persuasive since it could be done using individual-level genealogy data. Moreover, the sweep of historical time covered by the UPDB was extensive and covered key and theoretically critical stretches of history encompassing the demographic transition beginning with the migration to Utah, an early natural fertility regime followed by the declining infant mortality and fertility rates. In addressing the value of these data, they report no pattern of sex or geography bias within the genealogies. Their study of early deaths showed that the slower decline in rural infant mortality was due to ongoing problems these areas had with access to health care relative to the more urbanized areas.

Bean, Mineau and Anderton (1992) showed how the timing and spacing of births altered the risks of infant death. Using an early Utah settlement cohort, they demonstrated that children born to older mothers with larger families and shorter birth intervals were more susceptible to 'contagion-and-competition effects'. They showed that during the early settlement of Utah, high levels of fertility were associated with early age at marriage, early childbearing, short birth intervals, and late ages at last birth. These fertility patterns are observed in a range of European and American populations before the quick secular decline in fertility and in many developing nations as outlined in the book *Fertility Change on the American Frontier* (Bean,

Mineau, & Anderton, 1990), populations that also have relatively high rates of infant mortality. Bean and colleagues also used UPDB to show how public health and medical practices served to alter the patterns of infant mortality including public health campaigns, use of herbs, improvement in water quality, and broader sanitation measures (Bean, Smith, Mineau, Fraser, & Lane, 2002).

### 3.3 SEX DIFFERENCES

The pattern of excess male mortality across the life span is well-known. What is also well-known is a paradox — women seem to be sicker while living longer and men seemingly healthier with shorter lifespans. In Utah, the prevalence of unhealthy male risk behaviors is lower than in most other male populations (due to lifestyle such as the prohibition on the use of tobacco and alcohol), whereas women experience greater mortality risks because of elevated fertility rates. UPDB was used by Lindahl-Jacobsen and his colleagues (2013) to assess how Utah's sex differential in mortality differs from Sweden and Denmark, given Utah has lower (in relation to the European nations) risk behaviors for males and higher fertility related risks for females, at least in the early settlement years. This prediction was not supported since Utah male-female differences in mortality were similar to that of Sweden and Denmark, suggesting a central role of biological mechanisms.

As the Lindahl-Jacobsen study suggests, male mortality may be exceeded by female mortality during the reproductive years. Penn and Smith (2007) showed that parents during 19th century Utah incurred fitness costs (i.e., excess mortality) from reproduction with women having higher mortality risks than men. They examined the survivorship and reproductive success of over 20,000 couples married between 1860–1895 and found that parity was negatively associated with parental survivorship, and more so for mothers than fathers. Increasing family size was also associated with lower offspring survival, primarily for later-born children, indicating a tradeoff between offspring quantity versus quality.

Additional analyses of UPDB by Bolund and her collaborators (2016) adopted a life-history perspective which predicts that reduced reproduction should benefit female lifespan when females endure greater costs of reproduction. They show a shift from male-advantaged to female-advantaged adult survival in individuals born before versus during the demographic transition. As fertility decreased over time, female lifespan increased, while male lifespan was stable, supporting the theory that differential costs of reproduction in the two sexes result in the shifting patterns of sex differences in lifespan across human populations.

Harrell and colleagues (2008) examined a novel question about sex differentials: are girls good and boys bad for parental longevity? They find significant but small adverse mortality effects for mothers after age 50 who bore mostly sons. Offspring sex composition did not have a significant effect on paternal mortality. Bearing mostly boys was found to be detrimental to maternal mortality regardless of childhood survival. This study is another example of the intergenerational value of the UPDB and how early circumstances (sex composition of offspring) alters mortality risks of parents.

### 3.4 FERTILITY AND MORTALITY TRADEOFFS

UPDB has been the basis for answering a central question in biology, anthropology, and biodemography: How are mortality risks associated with fertility patterns? One of the oldest questions in this domain relates to the association between consanguinity and survival. Jorde (2001) examined how parental consanguinity was associated with offspring mortality risks. To do this, he estimated inbreeding coefficients for over 300,000 Utahns born between 1847–1945 where approximately 3,500 inbred offspring were identified. For this analysis, elevated relative risks of pre-reproductive mortality were found among the offspring of first-cousin marriages and among the offspring of closer unions. Jorde argues that these mortality risks are larger in populations with low inbreeding and low mortality which allow one to see the effects of consanguinity more readily.

Smith and his colleagues (2002) considered the effects of fertility on longevity among mothers and fathers after age 60. They drew on evolutionary theories of aging and theories predicting social benefits and costs of children to older parents. Using UPDB data on 13,987 couples married between 1860–1899, they found that women with lower parity and those bearing children late in life lived longer post-reproductive lives. Husbands' longevity was less sensitive to reproductive history, although they faced mortality effects similar to their wives for more recent marriage cohorts. The fact that late age fertility during a natural fertility period was associated with better survival, especially for females, is consistent with the idea that slow reproductive senescence (e.g., late age at natural menopause as indicated by later fertility) is associated with overall somatic senescence. They find support for predictions based on evolutionary hypotheses about the tradeoffs

between fertility and mortality. As noted previously, the tradeoffs between fertility and longevity were replicated by Gagnon and Smith and their collaborators (2009) using three frontier populations.

## 3.5 SPECIAL POPULATIONS: WIDOWHOOD AND RELIGION

### 3.5.1 WIDOWHOOD

Mineau, Smith and Bean (2002) examined whether recently widowed individuals, male or female, have higher rates of mortality than comparable married persons over historical time. They employed life course analyses of four marriage cohorts extending from 1860 through 1904 with mortality follow-up to 1990. They found significant differences in the mortality risk for widowed men and women, with widowed men having excess mortality risks in every cohort and nearly every age. A consistent pattern of excess mortality in the comparison of married and widowed women was not observed. A key finding is that the relative mortality risks of widowhood, when they occur, have grown over time as secular trends of mortality decline.

Barclay and his colleagues (2020) addressed a little understood aspect of widowhood — how marital bereavement affects adult mortality in the context of polygamy. They studied over 200,000 men and women born before 1900 and their mortality into the 20th century. They showed that the death of a polygamist husband and the death of a 'sister' wife have similar adverse effects on female mortality. For men, the death of one wife in a polygamous marriage increases mortality to a lesser extent than it does for men in monogamous marriages. For polygamous men, losing additional wives has a dose-response effect. They also demonstrated that the presence of other kin in the household (a second wife, a sister wife, or children) attenuates the adverse effects of bereavement.

### 3.5.2 RELIGIOUS GROUPS

Questions regarding religion are not uncommon when using the UPDB. Mineau, Smith and Bean (2004) sought to examine how religious affiliation affects mortality risk. They examined all-cause mortality for a set of married men and women who survived to age 40 from selected birth cohorts (1850–1919). They found that individuals active in the Church of Jesus Christ of Latter-day Saints have lower mortality risks than those who are inactive or non-LDS in all cohorts and this relationship remains after controlling for socioeconomic status. The protective influences of being an active member of the LDS Church are greatest for the middle-aged and for those born in the more recent birth cohort. These results show that religious affiliation has stronger effects on adult mortality for men rather than women. These observations are consistent with explanations of health practices and social support factors that have been posited to understand the positive relationship between religious involvement and mortality outcome.

## 3.6 MORTALITY SUMMARY

UPDB offers considerable opportunities for research on all-cause and age-specific mortality for all ages and organized into families and pedigrees. The family-oriented nature of UPDB, with its deep genealogical information, has promoted from the very beginning the study of both historical family demographic analysis of mortality risks but also the potential role of genetic factors. We have highlighted how demography and genetics have provided synergistic guidance in the collection and use of mortality data. It is noteworthy that death certification begins in 1904 in Utah and these data, linked to UPDB, provide the very first UPDB 'medical' data elements in terms of cause of death. That this cause-specific information is now available and coded into the International Classification of Disease (ICD) schema has launched numerous studies and is another facilitating component of UPDB leading to collaborations between medical and demographic scholars. It is noteworthy that with linkages to contemporary electronic medical records, the morbidity profiles of individuals and how they may relate to the risk of death are now being analyzed extensively. Finally, the fact that families in Utah tend to be large (by US standards) and Utah is home to a large percentage of residents who are members of the Church of Jesus Christ of Latter-day Saints, these aspects of Utah's population as represented in UPDB attract the attention of researchers. Specifically, UPDB is used to examine the role of religion and family structure on longevity and how shifts in these two forces may be driving changes in mortality risks.

## 4 LIFE COURSE ANALYSIS OF THE EFFECTS OF EARLY LIFE CONDITIONS

The value provided by UPDB and other reconstructed historical databases arises because it permits life course analysis of data spanning full lifetimes and across generations, including genetics and shared traits across generations. UPDB especially allows for extensive and innovative assessment of early and mid-life conditions, including the presence of kin such as grandmothers, and their effects on later-life health and survival. The components of UPDB permit researchers to increase their ability to examine these associations with much greater precision for a large range of important early life factors and key later-life health outcomes. For example, UPDB contains hundreds of thousands of members of birth cohorts from the first half of the 20th century individuals for whom early and midlife conditions are measured and who are linked to their adult medical and mortality records generated decades later.

There are several advantages of UPDB for life course analyses. The complex data links in UPDB provide unparalleled data quality and depth, especially those that focus on families (nuclear, multigenerational, full pedigrees) and health outcomes that span entire life spans of individuals and their relatives. Given the data linkage model of UPDB, one can also study health effects of early life conditions and whether they generate direct effects on subsequent health outcomes or whether they operate through or are moderated by characteristics and circumstances arising during the adult years (e.g., widowhood, proximity to adult children). UPDB often provides ample data to study even moderate statistical interactions with sufficient power which generally require large sample sizes. Relatedly, the great advantage of data linked to the UPDB include key measures that are repeated over time allowing for the construction of trajectories that describe the dynamics of early life circumstances and later outcomes. Given its long-standing focus on genealogies, familial data in UPDB help to address confounding bias through the use of statistical models (fixed and random effects). This means that factors for which we lack direct measures but which are shared by family members (e.g., common family-of-origin environmental exposures, shared genes) can be introduced into the multivariate models.

### 4.1 INTERGENERATIONAL ANALYSIS

#### 4.1.1 GENETICS AND INTERGENERATIONAL ASSOCIATIONS

The influence of the UPDB on advancing our understanding of genetics and inheritance is legendary. Accordingly, it is not feasible to summarize the vast volume of genetic studies and discoveries in this essay. A few important highlights are described here that should be of interest to historical demographers. In 2004, *The New York Times* noted the value of UPDB and how Utah is proving to be an ideal genetic laboratory given the connections between genealogical data and medical records as well as biospecimens. Over a decade later in 2017, *The Atlantic* also illustrated the value of UPDB for genetic discoveries. It is often said that more diseases with genetic origins have been discovered in Utah than at any other university, an achievement attainable to a large extent due to UPDB. An earlier review of family-based genetic studies from a decade ago also provides a history and summary of the value of UPDB in terms of methodologies used to quantify the heritable contribution to traits and to identify genes potentially responsible for these traits (Cannon Albright, 2008).

An example of early genetic work that used the UPDB was by McLellan and colleagues (1984) where they compared gene frequency data for Utah with the gene frequencies from a U.S. population, 13 European populations, and seven populations from three religious isolates. The gene frequencies in Utah were found to be similar to those of their northern European ancestors. This is explained by the large founding size of the pioneer population in Utah and high rates of gene flow. More isolated groups such as the Amish, Hutterites, and Mennonites revealed more divergence from their ancestral populations and each other, due in part to social isolation. In a similar way O'Brien and her co-authors (1994) also examined the genetic structure of the Utah population using UPDB data.

One of the most prominent genetic discoveries that used UPDB was for identifying genes associated with breast and ovarian cancer risk in women (Miki et al., 1994). The identification of BRCA1 mutations has facilitated early diagnosis of breast and ovarian cancer susceptibility in some individuals as well as a better understanding of breast cancer biology. Known mutation carriers of this gene have been used in conjunction with the genealogies within UPDB to impute the genetic status of ancestors who lived decades and centuries before testing was possible (Smith, Hanson, Mineau, & Buys, 2012). This made it possible for a historical demographic study of genetic influences on fertility when modern contraception was not available. There

are other well-known studies of the identification of cancer mutations using UPDB, for example, melanoma and colon cancer (Cannon-Albright, Kamb, & Skolnick, 1996; Neklason et al., 2008).

While not using genetic (sequence) data, UPDB has been used effectively to study historical and intergenerational patterns of demographic phenomena. Anderton and his co-authors (Anderton, Tsuya, Bean, & Mineau, 1987) evaluated the hypothesized relationship between the fertility behavior of mothers and that of their daughters and the role of cohort effects. They further evaluated how both cohort-specific intermediate fertility determinants and mother's relative fertility behavior may explain specific fertility-timing patterns of daughters. Their analyses indicated that both fertility behavior and indirect associations regarding timing of fertility-related life-course events (e.g., marriage) are transmitted intergenerationally, that cohort-specific influences are substantial, and that intergenerational relationships may be more readily elaborated through the examination of fertility relative to cohort levels. Several years later Jennings and her colleagues (Jennings, Sullivan, & Hacker, 2012) used UPDB to show that during the onset of the fertility transition, reproductive behavior was transmitted across generations between women and their mothers, as well as between women and their husbands' family of origin. The findings suggest that the practice of parity-dependent marital fertility control and inter-birth spacing behavior derived, in part, from the previous generation and that the potential for mothers and mothers-in-law to help in the rearing of children encouraged higher marital fertility.

#### 4.1.2 GRANDMOTHER HYPOTHESIS

Given the generational depth of the UPDB, it has attracted attention by scientists interested in the role of social networks and family dynamics. A specific and influential topic has been the Grandmother Hypothesis. This argues that the long human female post-menopausal life span can be explained by the idea that grandmothers provide care to their grandchildren thereby increasing their fitness — facilitating the reproduction of their offspring and survival of grandchildren. A few exemplary papers demonstrate the impact and value of UPDB in addressing this question.

Hawkes and Smith (2009) noted that grandmother effects can be measured in data sets that include births and deaths over several generations, while recognizing unmeasured covariates complicate the task. They examined two complications: cohort shifts in mortality and fertility, and maternal age at death. They show that longevity of grandmothers may actually be associated with fewer grandchildren even when grandmother effects are actually positive (i.e., increased fitness). They further explored this question to address why humans evolved greater longevity while continuing to end female fertility at about the same age as some of our closest relatives, the great apes. With the grandmother hypothesis in mind, they compared age-specific mortality and fertility rates between humans and chimpanzees. They used 19th century women from UPDB to represent non-contracepting humans, and compared their fertility by age with published records for wild chimpanzees. They found wide individual variation in age at last birth in both humans and chimpanzees. This heterogeneity, combined with differences in adult mortality, has large and opposing effects on fertility schedules. There was support for the hypothesis that ages at last birth changed little while greater longevity evolved in humans.

To remain on the topic of the Grandmother Hypothesis but with focus on fertility, Dillon and collaborators (2020) assessed the role of grandmothers in fertility outcomes in a comparative historical demographic study based on four populations from Scandinavia (Sweden) and North America (two in Québec and Utah). The individual-level data, including UPDB, are all longitudinal and multigenerational, allowing them to address the impact of maternal and paternal grandmothers on the fertility of their daughters and daughters-in-law, while attending to heterogeneous effects across space and time as well as within-family differences via the use of fixed effects models. They found associations of paternal grandmother presence with higher fertility across the regions, as well as a general fertility advantage associated with the post-reproductive availability of the maternal grandmother. Overall, grandmothers were generally associated with high-fertility outcomes, but that the mechanism for this effect was co-determined by family configurations, resource allocation and the advent of fertility control.

## 4.2 INITIAL ANALYSIS OF EARLY LIFE CONDITIONS

For life course analyses, one of the first comprehensive assessments of childhood and young adulthood life and its effects on later life outcomes with UPDB was conducted by Smith and colleagues (2009). They considered how key early family circumstances affect mortality risks decades later. Early-life conditions were measured by parental mortality, parental fertility, religious upbringing, and parental SES. They noted an important issue: prior to these early-life conditions are familial and genetic factors that affect life span. Accordingly, they

examined the role of parental and familial patterns of longevity on mortality risks demonstrating the power of familial data by using frailty models to control for unobserved heterogeneity within families, all based on sib-pair data for 12,000 sib-pairs. They reported modest but significant effects of key childhood conditions (birth order, sibship size, parental religious affiliation, parental SES, and parental death in childhood). The effects of familial patterns of longevity were large and suggest that family history of key demographic measures may be an important but overlooked early life condition.

### 4.3 EARLY ADVERSITY AND LATER LIFE OUTCOMES: THE CASE OF FAMILY DEATHS IN CHILDHOOD

The range of early life conditions are manifold. For contemporary settings where individuals are asked to recall circumstances of their childhood and youth, batteries of questions exist such as the Adverse Childhood Experiences (ACE) instrument (Felitti et al., 1998; Greeson et al., 2014). Unfortunately, for historical demographers, the types of experiences and trauma encompassed in such measures are not generally available in the historic record. But this circumstance does not preclude demographers from exploring indicators of such events if they can be measured. The UPDB has identified a class of traumatic and adverse events that can be detected: deaths of family members when a person is young. This class of variables is arguably an unambiguous indicator of stress in childhood, is visible in most historical databases and when family relationships are well measured, and one can analyze how the type of death (sibling, parent, child) and cause of death may have differing effects.

#### 4.3.1 SIBLINGS AND OFFSPRING

Van Dijk, Janssens and Smith (2019) observed that the literature has yielded mixed evidence about the influence of infant and child mortality in *birth cohorts* on adult mortality. These studies generally do not examine the specific role of mortality *within a family context* when micro data are available. They examined how exposure to mortality as a child is related to their adult mortality risk between ages 18 and 85 in UPDB (1874–2015) and the LINKS data from Zeeland (the Netherlands) 1812–1957. They found that childhood exposure to community mortality and sibling deaths increases adult mortality rates. Effects of sibling mortality on adult all-cause mortality risk were stronger in Utah, where sibling deaths were less common in relation to Zeeland. Exposure to sibling deaths from infection was related to the siblings' risk of adult mortality from cardiovascular disease and diabetes mellitus, a result consistent with an inflammatory immune response mechanism.

The direct measure of biomarkers for inflammation were available in a separate study (based on the Cache County Memory and Health Study linked to UPDB) which showed that sibling deaths also elevated inflammation, a key biomarker for mortality, as measured by high-sensitivity C-reactive protein (CRP) (Norton, Hatch, Munger, & Smith, 2017). This study demonstrated a link between significant psychosocial stress in early life and immune-inflammatory functioning in late life, and reinforces a mechanism explaining the link between early-life adversity and late-life health.

#### 4.3.2 PARENTS: EFFECTS ON OFFSPRING MORTALITY

A study by Smith and colleagues (2014) asked whether a parental death is associated with enduring mortality risks after age 65? The years following parental death may initiate circumstances in which the adverse effects of paternal death operate. Accordingly, they examined the offspring's marital status, adult SES, fertility, and later-life health status, where the latter relies on the comprehensive Charlson Comorbidity Index (Charlson, Szatrowski, Peterson, & Gold, 1994) using Medicare data. They show that offspring whose parents died when they were children, but especially when they were adolescents/young adults, have modest but significant mortality risks after age 65. Strikingly, there were weak mediating influences of later-life comorbidities, marital status, fertility and adult socioeconomic status.

One of the hypothesized effects of parental death is suicide risk of the offspring. Hollingshaus and Smith (2015) examined this question while also considering whether the surviving parent remarried. Using UPDB for birth cohorts between 1886 and 1960 (N = 663,729, including 4,533 suicides), they demonstrated that parental death was associated with an excess risk of adult offspring suicide before age 50, and with increased risk of cardiovascular disease deaths (CVD) for adults of all ages. Daughters whose surviving parents remarried had (in relation to those who did not remarry) a smaller risk of suicide before age 50 (though not statistically significant), but significantly higher risk after age 50. Parental remarriage had no effect on male suicide risk. This analysis illustrates the value of using death certificate and detailed health information as a method to explicate possible mechanisms linking early events to later health outcomes.



### 4.3.3 PARENTS: EFFECTS ON OFFSPRING ALZHEIMER'S DISEASE (AD) RISK

With UPDB, all-cause mortality and suicide specifically have been identified as risks for those experiencing the death of a parent. This is possible with UPDB and other databases linked to vital records. UPDB has also had impact by virtue of links to cohorts whose clinical assessments and questionnaire response have been linked to UPDB. This has led to a series of papers on parental death in childhood and AD risk.

In a pair of papers, Norton and colleagues (2009, 2011) examined early parental death and late-life dementia risk in offspring based on links between UPDB and the Cache County Memory and Health Study. They showed parental death during one's childhood is associated with higher prevalence of AD, with different effects based on the ages of an individual when they experienced father's versus mother's death. The strength of these associations was attenuated by remarriage of the widowed parent.

The same team of investigators examined family deaths more broadly using UPDB genealogical and mortality data and their effects on AD risk (Greene et al., 2014; Norton et al., 2016). Norton and her team examined whether parental death during one's childhood, and offspring and spouse deaths during adulthood are associated with faster cognitive decline and higher Alzheimer's disease (AD) risk in late life. Using 4,545 non-demented participants from the Cache County Memory and Health Study linked to UPDB found that age moderated the relationship between family deaths and AD. For persons aged 65–69 years at baseline those exposed to more deaths during adulthood faced a two-fold AD risk whereas those over 80 years had a lower AD risk. Their findings again emphasized the value of linking family history from UPDB to outcomes from an epidemiological cohort where they were able to demonstrate an effect about the link between family member deaths during adulthood and AD risk later in life. In a related study, Greene and his collaborators (Greene et al., 2014) tested the hypothesis that experiencing an offspring death was associated with an increased rate of cognitive decline in late life. They also used UPDB linked to the Cache County Memory and Health Study based on 3,174 non-demented residents aged 65–105. They reported that subjects who experienced offspring death before age 30 experienced a significantly faster rate of cognitive decline in late life, but only if they had an *APOE e4* allele (a major genetic risk factor for Alzheimer's Disease). Also, an offspring death was only related to faster cognitive decline when there were no subsequent births. Other stressful life events were also shown to affect the rate of cognitive decline in this cohort (Tschanz et al., 2013).

## 4.4 LIFE COURSE SUMMARY

The depth of familial data spanning over 200 years contained with UPDB creates an opportunity to link the circumstances of an individual early in their lives to events over all years of their existence until death or at least for many decades. With the genealogical structure of UPDB, the data can then link the circumstances of the parents and more distant ancestors to the life course of the target individual for fertility, health, socioeconomic, and mortality outcomes. UPDB data provide the kind of data that supports and is consistent with the perspective advanced by the Developmental Origins of Health and Disease (DOHaD) concept (Gage, Munafò, & Davey Smith, 2016; Hanson, Poston, & Gluckman, 2019; Penkler, Hanson, Biesma, & Muller, 2019). Overall, deep longitudinal data with nearly complete family ascertainment over many generations provide the necessary ingredients for life-course research. Indeed, in the contemporary period, it is now possible to consider the role of 'shocks' long ago and how they may affect the expression of genetic predispositions (i.e., epigenetics) today. It is also worth noting that the challenge for life course research is the vexing problem of loss to follow-up that is unavoidable when so many years of follow-up for millions of families are involved. UPDB takes steps to monitor and time-stamp, when possible, the arrival to and departure from Utah.

## 5 SPECIAL TOPICS

### 5.1 SOCIO-ECONOMIC STATUS (SES)

It is well known that socioeconomic status (SES) is a central facet affecting and being affected by demographic structure and change. Nonetheless, users of historic records seeking to identify measures of SES face challenges given the vagaries of the source records and their completeness and consistency over time. UPDB data are not immune to these challenges. Despite this, UPDB data contain relatively consistent measures of SES over many decades that rely most heavily on occupation and industry. These data are derived primarily from vital

records (birth and death certificates; birth from 1915 and death from 1904) and from the US Census of Utah for the available decennial censuses spanning 1880–1940. These data pertain to individuals (adults and by extension, to their offspring). Given the depth of genealogical and spatial information in UPDB, it is possible to construct SES measures of kin to determine a familial composite SES indicator of a person's kin and their geographic proximity. 'Neighborhood' or community indicators of SES can also be derived within UPDB data as well through the more conventional methods of linking location data in UPDB data to Census (or any other geo-referenced) records at the county, census tract or census block group level.

A persistent question exists in the study of SES and health outcomes, particularly mortality risk: has the inverse relationship between SES and mortality risk always existed or is it a recent phenomenon? Bengtsson, Dribe and Helgertz (2020) have acknowledged that today there is a consistent mortality pattern by SES at all ages but not all confirm this association. They note that if a gradient did not exist in the past, then when did the association begin. They conclude that adult mortality risks for men and women in southern Sweden over a 200-year period was associated with social class risks emerging for middle-aged persons only after 1950 for women and after 1970 for men, and later for ages 60–89. These findings occurred when Sweden became a modern welfare state with universal health care system, suggesting the importance of psychosocial factors. In contrast, Smith and his team (2009) examined with UPDB data how key early family circumstances affect mortality risks decades later including parental mortality, parental fertility religious upbringing (Mineau et al., 2004), and *parental SES* for individuals born in the 19th century. Using frailty models for 12,000 sib-pairs, they found significant adult mortality effects associated with parental SES in childhood. This suggests that during a period of limited formal social support mechanisms, an inverse SES-mortality association was detected in Utah.

The conduct of innovative studies using SES in UPDB is represented in several ways. Temby and Smith (2014) elaborated on the SES and mortality association by considering the interaction between SES and a positive family history of longevity. They considered, for example, whether individuals with lower levels of SES may experience an attenuated longevity penalty if they have long-lived relatives. Analysis of survival past age 40 for men born between 1840 to 1909 showed that mortality risks for men with the highest SES was reduced more as familial longevity increased than it does for the lowest SES men. Mortality risks for farmers also declines more as familial longevity increases in relation to non-farmers. They suggest that a type of gene-environment interaction occurs whereby the benefits of a family history of longevity are more available to those who have higher status occupations.

A novel examination between the interplay of the SES of offspring and the mortality risks of their parents was considered by Zimmer and his colleagues (2016b). They tested the hypothesis that SES effects may 'flow up' from offspring to parents: higher offspring SES associates with lower parental mortality after age 40 after controlling for parental SES. They used 30,000 individuals born between 1864–1883 whose offspring were born between 1886–1920 where SES was based on the Nam-Powers occupational status scores (Nam & Boyd, 2004; Nam & Powers, 1983) divided into quartiles and a category for farmers. They showed a longevity penalty for parents whose offspring have low SES and a longevity dividend for those with high-SES offspring. They expanded on this hypothesis (Zimmer, Hanson, & Smith, 2016a) by considering morbidity as measured by the Charlson Comorbidity Index (Charlson et al., 1994). They used sex-specific group-based trajectory patterns of morbidity and survival where group morbidity trajectories were ranked from least to most healthy. They showed that increasing (one's own) SES in childhood is associated with membership in groups that have more favorable morbidity trajectories as well as survival probabilities. SES in adulthood has additive impact, especially for females. These two studies illustrate the insights gained from both longitudinal and intergenerational analytic strategies: the influence of offspring SES on well-being of parents and the role of both childhood and adult SES in independently influencing old-age morbidity risks.

UPDB data have been effective in helping to expose SES differentials in fertility as well. Dribe and his cross-national comparative team (2017) used longitudinal individual-level data from five populations in Europe and North America to examine linked SES and fertility during the fertility transition. They specifically studied the dynamism of SES differences in marital fertility and related these to fertility behavior during the demographic transition. They found no support for the hypothesis of universally high fertility among the upper classes in pre-transitional society, but did provide findings consistent with the hypothesis that the upper classes preceded other groups in reducing their fertility. Farmers and unskilled workers were the latest to start limiting their fertility. Within Utah, Maloney, Hanson and Smith (2014) used UPDB to examine differences across occupational classes in fertility levels and in the timing and pace of change in fertility in Utah in the late 19th and early 20th centuries. They showed that families of white-collar workers led changes in many fertility-related behaviors including age at first marriage and first birth interval while farm families continued

to have high fertility levels and bore children into later ages. They identified patterns of fertility change tied to variation in important economic circumstances such as the length of education and training required for particular occupations, or the need for family-based labor on the farm.

## 5.2 TWINS

Given the richness of the genealogical, and therefore fertility information of UPDB, a specific topic has attracted the attention of historical demographers: twins. We note that UPDB does not comprehensively contain zygosity information so it is not possible to assess identical and fraternal twins in the database. The exception relates to opposite sex twins who are by definition dizygotic (non-identical). One of the earliest examination of twins using UPDB was work by Carmelli, Hasstedt and Andersen (1981). They investigated demographic and genetic aspects of human twinning. They found that Utah has an elevated incidence of twinning in relation to the US white population. They concluded that most of the general decrease in twinning during the 19th and 20th centuries was due to a maternal age effect. Couples bearing children in later US birth cohorts (after 1900) began to practice family limitation, limiting fertility at older ages. Utah couples did not significantly alter their behavior during the early years of childbearing so their fertility remained high with an average number of children above four thereby enhancing the number of twin births.

Decades later, Robson and Smith (2011, 2012) tested two hypotheses regarding twinning in human populations that have alternative predictions about the effects of bearing twins on maternal life time reproduction and survival. The maternal depletion hypothesis argues that mothers of twins will suffer negative outcomes while a 'robustness' hypothesis argues that while twinning is costly, it may reveal mothers with a greater capacity to bear that cost. Using UPDB, they examine mothers who lived at least to the age of 50 and found evidence consistent with the robustness hypothesis: mothers of twins had lower postmenopausal mortality, shorter average inter-birth intervals, later ages at last birth and higher lifetime fertility than their singleton-only bearing counterparts. They concluded that bearing twins is more likely for those with a robust phenotype and may be a useful indicator of maternal heterogeneity.

While the robustness hypothesis was supported with respect to the mother, a follow-up study (Chernenko, Hollingshaus, Robson, Hanson, & Smith, 2018) examined mortality patterns for the singleton offspring of mothers who had twins compared to the single offspring of mothers who had not had twins to determine whether they share the hypothesized robust phenotype of their mothers. They showed that singleton offspring of twinning mothers experience a survival disadvantage prior to age 5, no survival benefit or penalty between ages 5 and 49, and — for males only — a significant survival advantage after age 50. They also found a survival disadvantage in early life for singleton offspring of twinning mothers born immediately after the twinset for both sexes. They conclude that while bearing twins may reflect a robust maternal phenotype, the toll of bearing twins may disadvantage subsequent offspring, especially during infancy.

## 5.3 SEX RATIOS

Investigators have used UPDB to examine several factors associated with a key structural demographic measure, sex ratios, and their consequences. Analysis of the fertility histories of women born between 1850–1900 by Bohnert and colleagues (2012) considered whether there was evidence of sex preference in these early decades of Utah's history. They found more male children, as expressed by birth stopping behavior after the birth of a male child and shorter birth intervals in higher-parity births when most previous children were female. Evidence was presented focusing on two sub-populations, farmers and individuals with stronger ties to the Church of Jesus Christ of Latter-day Saints. They showed that the former, while having relatively high fertility rates, had similar preferences for male children as the other Utahns.

Farmers, who presumably had a need for family labor, were more interested in the quantity than in the sex mix of their children. Schacht and Smith (2017) examined the historical patterns of sex ratios and their importance in understating shifts in other demographic phenomena. They further evaluated whether the sex ratio at birth (SRB) may be patterned by maternal condition and/or environmental stressors (Schacht, Tharp, & Smith, 2019). Using UPDB data for the population during the interwar period (1918–1939), inclusive of three distinct eras (Spanish Flu, Roaring '20 s, and the Great Depression), they assessed two theoretical frameworks used to study patterning in SRB — (1) 'frail males' and (2) adaptive sex-biased investment theory (Trivers & Willard, 1973). The first approach centers on greater male susceptibility to exogenous stressors and argues that offspring survival should be expected to differ between 'good' and 'bad' times. The second approach predicts that mothers themselves play a direct role in manipulating offspring SRB, and that those in better condition should invest more in sons. Consistent with the 'frail male' predictions, they found that boys are less likely to be born during the environmentally challenging times of the Spanish Flu and Great

Depression. However, they found no evidence that maternal condition is associated with sex ratios at birth, a result inconsistent with the Trivers-Willard hypothesis.

## 5.4 SPECIAL TOPICS

Topics highlighted here are, quite simply, key selected examples of demographic interest but there are certainly many others. We note that other cross-cutting topics include studies that are spatially oriented (Smith et al., 2011; Stroup et al., 2017; Zick et al., 2009) as well as less-studied family formation topics such as those investigating step-children (Schacht, Meeks, Fraser, & Smith, 2021). The range of possible topics is considerable and in cases where the event is rare (e.g., extreme longevity, very young fertility) or is likely to vary by context (e.g., different centuries or nations), the size and breadth of UPDB lends itself to comparative analyses (e.g., Dillon et al., 2020; Dribe et al., 2017; Gagnon et al., 2009; Smith et al., 2009).

## 6 CONCLUSIONS AND FUTURE DIRECTIONS

Many historical databases represent considerable investments of time, talent, and institutional commitment. This characterization certainly fits the UPDB. While the ideas and actions of hundreds of investigators and their academic departments and universities have already generated considerable value to scientific advances broadly and demography specifically, the future provides us with additional opportunities and challenges.

### 6.1 GENETICS AND DEMOGRAPHY

We have described numerous instances where geneticists and demographers have benefitted from each other's perspectives and methodologies (Adams, Lam, Hermalin, & Smouse, 1990; Bean, 1990). While they inform each other, the UPDB and its deep genealogical data are based on high quality and desirably redundant data (i.e., multiple reports of the same family connections) which are not always confirmed genetically. At its core, between genealogies from the Genealogical Society of Utah and vital records as well as other sources, there is the possibility that some of the genetic relationships may need verification. Mistaken connections are rare but happen nonetheless due to the human nature of record keeping, regulations that protect identities (e.g., birth versus adoptive parents on birth certificates), and explicit actions to insulate the parties from adverse consequences (e.g., naming an individual as the father incorrectly). The rise of 'Big Genetics' (Smith, Hanson, & Mineau, 2016) is increasingly generating sequence data that is quantifying the likelihood that two persons are related genetically and in what way. Validating the genealogical connections this way, to the extent that these DNA measures are suitably available, may improve the quality of the UPDB and its genealogies. Similarly, genealogies may also provide corrections to the processing of genetic information, which also has imperfections, and serve to help correct erroneously linked persons identified through sequence data.

The burgeoning field of biodemography over recent decades, emerging against the backdrop of the Human Genome Project and the HapMap Project, reflects the union of ideas and data derived from demography, evolutionary biology, genetics, and medicine. Recognition of these connections was central to the concept and creation of the UPDB — and outlined in *Convergent Issues in Genetics and Demography* (Adams et al., 1990) in a chapter by Dr. Lee L. Bean, one of the original co-developers of the UPDB titled 'Utah Population Database: Demographic and Genetic Convergence and Divergence'. This early insight has led to several key genetic discoveries and insights (Adams et al., 1990; Cawthon et al., 2020; Cawthon, Smith, O'Brien, Sivatchenko, & Kerber, 2003; Hanson et al., 2020; Miki, et al., 1994; Neklason et al., 2008; Norton et al., 2010, 2016; Smith et al., 2012). The contribution of the UPDB to genetics is considerable and per se is beyond the scope of this review but that specific impact of the UPDB is acknowledged below.

We can imagine demographic analyses that can effectively control for genetic variants that may contribute to our understanding of central outcomes such as fertility and mortality. This appears to be more common in epidemiology (e.g., Carreras-Torres et al., 2014) where polygenic risk scores (summary values of overall risk as measured across a range of many genes and alleles) but it is easy to see how this may attract the interest of demographers.

One example noted previously may be applied more broadly which is to use known carriers of certain genetic variants in the population and then exploit information about family lineages and modes of inheritance to

identify (obligate) carriers in the past based on where a person is in the family pedigree. For example, if we know that two children have a specific genetic variant, and the disease follows an autosomal dominant mode of inheritance, we can assume that one of the parents (the common ancestor) is also an obligate carrier. This can be expanded to more distant relatives. This method was applied using BRCA1 mutations (which increase the risk of breast and ovarian cancer) in modern samples and studying mortality and fertility differentials of common ancestors for whom we can ascertain their mutation status (Smith et al., 2012).

For demographers who generally work with large and sometimes complete population data of individuals, they would prefer to have genetic information on all persons under investigation along with other commonly analyzed covariates. This is an ambition and is slowly becoming available in certain data sets such as the US Health and Retirement Survey but it is still rare, though the lower costs for genetic sequencing is making this approach more affordable. As genetic data become more available for large samples, the prospects for conducting gene x environmental interactions, obligate carriers analyses and social genomics (Das, 2021; Sanz-de-Galdeano, Terskaya, & Upegui, 2020) will become more common. For UPDB, it is the norm that selected families and pedigrees displaying traits of interest are recruited and their DNA samples collected, rather than wholesale genetic data collection of large sections of the population.

With UPDB, it is worth pointing out that in cases where genomic sequence data to characterize possible genetic predispositions are lacking on a population basis, it is quite feasible to use family history of a trait as an empirical and practical alternative. Use of such family history information has been described in this review and is used as a broader indicator of shared genes and environments. While admittedly a different type of 'genetic' risk indicator, it offers several advantages such as being feasible, inexpensive, adaptable to the type of inheritance pattern (e.g., autosomal dominant, maternal (mitochondrial) inheritance), and having the flexibility for investigators to create multiple family history measures for different traits or behaviors. Others have advocated the use of family history as an attractive tool for assessing possible genetic susceptibility (Rich et al., 2004; Scheuner, Wang, Raffel, Larabell, & Rotter, 1997).

## 6.2 GEOGRAPHY AND DEMOGRAPHY

These two disciplines have been close cousins for centuries and UPDB reflects this close connection and accordingly it contains vast data elements important to both. UPDB has devoted resources to enhance the ability of investigators to test spatially-oriented hypotheses linked to health and demographic outcomes. Geo-referencing spatial data in order to locate individuals and their relatives and neighbors over decades has been a focus of UPDB (Leiser et al., 2020; Stroup et al., 2017). The challenge is to link persons to an area with comparable precision subject to data constraints. In the US, ZIP codes may be available but not precise addresses. In the more distant past, place names are generally available and locating them with consistency with systematic coding is on-going and generally achievable, especially when the need is to study complete lives and all their locations. Again, inclusion of some persons for whom there is geographic information may come at the price of precision. Certainly, working with other organizations, notably the National Historical Geographic Information System (NHGIS) with IPUMS is providing invaluable benefits in this respect. The increasing availability of environmental exposure data, such as from the US Environmental Protection Agency's Air Quality System Data Mart, can be layered on top of geo-referenced data (precise spatial coordinates, US census tracts or block groups) to develop environmental exposures to support studies of their influence on demographic outcomes.

## 6.3 CENSUS RECORDS

UPDB contains linked individual data from the US Censuses of Utah from 1880–1940. This achievement has taken place over the past 40 years motivated and funded by specific research projects but the resulting Census data are now a feature of UPDB. Given the GIS data and geo-referenced persons at various points in time during their lives, it is also common to link areal data from the Census (and other GIS data sources) to UPDB. This historic geospatial information may have broad area indicators (Census Enumeration Districts identifiers) which may be used to control for shared environmental features without necessarily knowing what those precise features are.

The University of Utah now houses one of the Federal Statistical Research Data Centers (FSRDC), a facility that provides access to otherwise secure data collected by the federal government and made available to approved investigators. The FSRDC infrastructure allows other data, such as the UPDB, to link to the US Census and data it collects, if approval is secured from the Census and the institutional owners of the data contained within UPDB. This linkage provides an outstanding opportunity to link securely and with privacy

safeguards UPDB to other federal data such as the decennial Census and the American Community Survey. Such linked data can then be analyzed within the FSRDC (The Wasatch Front Research Data Center at the University of Utah).

## 6.4 ANCESTORS THEN, DESCENDANTS NOW

The UPDB grows every year as new data accrue to the agencies who collect them, who in turn provide approved data to UPDB for linkage. The newly installed data represent both additional information about a given individual as well as new descendants being added to a Utah family tree either through birth, marriage or migration. This allows new opportunities to examine how prior events or genetic signals continue to impart their influence on current residents of Utah.

As discussed in the section on Life Course Analysis, this in-depth and ever-increasing volume of allows demographers to think broadly about how the configuration of one's family pedigree and their spatial orientation may have previously unknown effects on key outcomes. Additional questions can be addressed related to the presence of a grandmother and daughter fertility, density of kin in the neighborhood and survival, adoption of reproductive behavior as a function of that behavior in close relatives, and the volume and nature of specific causes of death in relatives and how that alters an individual's risk of death from those causes.

## 6.5 BRIDGING DATABASES AND COMPARATIVE ANALYSES

The growth of historical data bases throughout the globe give rise to the potential of comparative research but also direct collaboration by identifying family lineages that touch more than database. In the case of UPDB, the founding European population was based in northern and western Europe. For example, Scandinavian and British founders served as the basis of the early pioneers who settled Utah's early frontier expansion. Accordingly, many present data Utahns can claim their heritage from those parts of Europe and indeed common ancestors appear in UPDB and several of the European historical demographic databases. The opportunity exists to examine the demographic impact of staying versus leaving one's home country — some left Europe to the US and others remained. What has become of these two broad segments of a family and their respective descendants? Indeed, what can we learn from those who left Europe for the US only to return later?

## 6.6 FINAL THOUGHTS

The UPDB has contributed to the growth and development of demography through the sheer number of people and families represented in its data holdings and the wide-ranging data available about each individual. But we recognize that UPDB has succeeded for other reasons. One factor for UPDB's success relates to the relatively small size of Utah's population and institutions. Utah's small population at the outset in the mid-1970s likely contributed to its launch. The volume of data was more manageable and the ability of key institutions to interact was conducive to creating a collaborative atmosphere between the initial participating institutions: The Genealogical Society of Utah, The University of Utah, and the Utah Department of Health. On the latter point, the geographic proximity of these institutions contributed to interactions, negotiations and agreements that would likely have been more problematic in much larger states. The consistent and robust support of the University of Utah and the Huntsman Cancer Institute to maintain and fund the UPDB is without question a key ingredient in explaining the success of the database.

In the end, the growth and evolution of investigators and topics reliant on UPDB can in part be attributed to the catalyzing effects of big data on team science (Sellers, Caporaso, Lapidus, Petersen, & Trent, 2006; Shah, Pico, & Freedman, 2016; Stokols, Misra, Moser, Hall, & Taylor, 2008). The diversity and quality of UPDB data that is curated and made safely available has served to induce many ambitious projects that involve investigators from multiple disciplines that would not have been possible otherwise. This has created teams that often combine medical, population and social sciences. These multidisciplinary efforts serve to strengthen the science under investigation.

## ACKNOWLEDGEMENTS

Support for this work was partially provided by R01 AG022095 (Early Life Conditions, Survival, and Health: A Pedigree-Based Population Study). We also thank the Pedigree and Population Resource of Huntsman Cancer Institute, University of Utah (funded in part by the Huntsman Cancer Foundation) for its role in the ongoing collection, maintenance and support of the Utah Population Database (UPDB). We also acknowledge partial support for the UPDB through grant P30 CA2014 from the National Cancer Institute, University of Utah and from the University of Utah's program in Personalized Health and Center for Clinical and Translational Science.

## REFERENCES

- Adams, J., Lam, D. A., Hermalin, A. I., & Smouse, P. E. (Eds.) (1990). *Convergent issues in genetics and demography*. New York: Oxford University Press.
- Anderton, D. L., & Bean, L. L. (1985). Birth spacing and fertility limitation: A behavioral analysis of a nineteenth century frontier population. *Demography*, 22(2), 169–183. doi: [10.2307/2061176](https://doi.org/10.2307/2061176)
- Anderton, D. L., Bean, L. L., Willigan, J. D., & Mineau, G. P. (1984). Adoption of fertility limitation in an American frontier population: An analysis and simulation of socio-religious subgroups. *Social Biology*, 31(1–2), 140–159. doi: [10.1080/19485565.1984.9988569](https://doi.org/10.1080/19485565.1984.9988569)
- Anderton, D. L., Tsuya, N. O., Bean, L. L., & Mineau, G. P. (1987). Intergenerational transmission of relative fertility and life course patterns. *Demography*, 24(4), 467–480. doi: [10.2307/2061386](https://doi.org/10.2307/2061386)
- Arbeeva, L. S., Hanson, H. A., Arbeev, K. G., Kulminski, A. M., Stallard, E., Ukraintseva, S. V., . . . Yashin, A. I. (2018). How well does the family longevity selection score work: A validation test using the Utah Population Database. *Front Public Health*, 6, 277. doi: [10.3389/fpubh.2018.00277](https://doi.org/10.3389/fpubh.2018.00277)
- Barclay, K. J., Thorén, R. D., Hanson, H. A., & Smith, K. R. (2020). The effects of marital status, fertility, and bereavement on adult mortality in polygamous and monogamous households: Evidence from the Utah Population Database. *Demography*, 57(6), 2169–2198. doi: [10.1007/s13524-020-00918-z](https://doi.org/10.1007/s13524-020-00918-z)
- Bean, L. L. (1990). The Utah Population Database: Demographic and genetic convergence and divergence. In J. Adams, D. A. Lam, A. I. Hermalin & P. E. Smouse (Eds.), *Convergent issues in genetics demography* (pp. 231–244). New York: Oxford University Press.
- Bean, L. L., & Mineau, G. P. (1986). The polygyny-fertility hypothesis: A re-evaluation. *Population Studies*, 40(1), 67–81. doi: [10.1080/0032472031000141846](https://doi.org/10.1080/0032472031000141846)
- Bean, L. L., Mineau, G. P., & Anderton, D. L. (1990). *Fertility change on the American frontier: Adaptation and innovation*. Berkeley: University of California Press.
- Bean, L. L., Mineau, G. P., & Anderton, D. L. (1992). High-risk childbearing: Fertility and infant mortality on the American frontier. *Social Science History*, 16(3), 337–363. doi: [10.1017/S0145553200016539](https://doi.org/10.1017/S0145553200016539)
- Bean, L. L., Smith, K. R., Mineau, G. P., Fraser, A., & Lane, D. (2002). Infant deaths in Utah, 1850–1939. *Utah Historical Quarterly*, 70(2), 158–173.
- Bengtsson, T., Dribe, M., & Helgertz, J. (2020). When did the health gradient emerge? Social class and adult mortality in southern Sweden, 1813–2015. *Demography*, 57(3), 953–977. doi: [10.1007/s13524-020-00877-5](https://doi.org/10.1007/s13524-020-00877-5)
- Bohnert, N., Jåstad, H. L., Vechbanyongratana, J., & Walhout, E. (2012). Offspring sex preference in frontier America. *The Journal of Interdisciplinary History*, 42(4), 519–541. doi: [10.1162/JINH\\_a\\_00303](https://doi.org/10.1162/JINH_a_00303)
- Bolund, E., Lummaa, V., Smith, K. R., Hanson, H. A., & Maklakov, A. A. (2016). Reduced costs of reproduction in females mediate a shift from a male-biased to a female-biased lifespan in humans. *Scientific reports*, 6, 24672. doi: [10.1038/SREP24672](https://doi.org/10.1038/SREP24672)
- Cannon Albright, L. A. (2008). Utah family-based analysis: Past, present and future. *Human Heredity*, 65(4), 209–220. doi: [10.1159/000112368](https://doi.org/10.1159/000112368)
- Cannon-Albright, L. A., Kamb, A., & Skolnick, M. (1996). A review of inherited predisposition to melanoma. *Seminars in Oncology*, 23(6), 667–672.
- Carmelli, D., Hasstedt, S., & Andersen, S. (1981). Demography and genetics of human twinning in the Utah Mormon Genealogy. *Progress in Clinical and Biological Research*, 69A, 81–93.
- Carreras-Torres, R., Kundu, S., Zanetti, D., Esteban, E., Via, M., & Moral, P. (2014). Genetic risk score of NOS gene variants associated with myocardial infarction correlates with coronary incidence across Europe. *PLoS ONE*, 9(5), e96504. doi: [10.1371/journal.pone.0096504](https://doi.org/10.1371/journal.pone.0096504)

- Casey, J. A., Schwartz, B. S., Stewart, W. F., & Adler, N. E. (2016). Using electronic health records for population health research: A review of methods and applications. *Annual Review of Public Health, 37*, 61–81. doi: [10.1146/annurev-publhealth-032315-021353](https://doi.org/10.1146/annurev-publhealth-032315-021353)
- Cawthon, R. M., Meeks, H. D., Sasani, T. A., Smith, K. R., Kerber, R. A., O'Brien, E., . . . Jorde, L. B. (2020). Germline mutation rates in young adults predict longevity and reproductive lifespan. *Scientific reports, 10*(1), 10001. doi: [10.1038/s41598-020-66867-0](https://doi.org/10.1038/s41598-020-66867-0)
- Cawthon, R. M., Smith, K. R., O'Brien, E., Sivatchenko, A., & Kerber, R. A. (2003). Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet, 361*(9355), 393–395. doi: [10.1016/S0140-6736\(03\)12384-7](https://doi.org/10.1016/S0140-6736(03)12384-7)
- Charlson, M., Szatrowski, T. P., Peterson, J., & Gold, J. (1994). Validation of a combined comorbidity index. *Journal of Clinical Epidemiology, 47*(11), 1245–1251. doi: [10.1016/0895-4356\(94\)90129-5](https://doi.org/10.1016/0895-4356(94)90129-5)
- Chernenko, A., Hollingshaus, M., Robson, S., Hanson, H. A., & Smith, K. R. (2018). Tykes, toddlers, teens, and twins of robust mothers: Do the offspring of twinning mothers share in their mother's robust phenotype? *Biodemography and Social Biology, 64*(2), 102–113. doi: [10.1080/19485565.2018.1486697](https://doi.org/10.1080/19485565.2018.1486697)
- Chernenko, A., Meeks, H., & Smith, K. R. (2019). Examining validity of body mass index calculated using height and weight data from the US driver license. *BMC Public Health, 19*(1), 100. doi: [10.1186/s12889-019-6391-3](https://doi.org/10.1186/s12889-019-6391-3)
- Das, A. (2021). Transpersonal genetic effects among older U.S. couples: A longitudinal study. *The Journals of Gerontology: Series B, 76*(1), 184–194. doi: [10.1093/geronb/gbz151](https://doi.org/10.1093/geronb/gbz151)
- Dillon, L., Chernenko, A., Dribe, M., Engelhardt, S., Gagnon, A., Hanson, H. A., . . . Vézina, H. (2020). Did grandmothers enhance reproductive success in historic populations?: Testing evolutionary theories on historical demographic data in Scandinavia and North America. In O. Burger, R. Lee & R. Sear (Eds.), *Human Evolutionary Demography*. Retrieved from <https://osf.io/qgjp8/>
- Dribe, M., Breschi, M., Gagnon, A., Gauvreau, D., Hanson, H. A., Maloney, T. N., . . . Vézina, H. (2017). Socio-economic status and fertility decline: Insights from historical transitions in Europe and North America. *Population Studies, 71*(1), 3–21. doi: [10.1080/00324728.2016.1253857](https://doi.org/10.1080/00324728.2016.1253857)
- DuVall, S. L., Fraser, A. M., Rowe, K., Thomas, A., & Mineau, G. P. (2012). Evaluation of record linkage between a large healthcare provider and the Utah Population Database. *Journal of the American Medical Informatics Association, 19*(e1), e54–e59. doi: [10.1136/amiajnl-2011-000335](https://doi.org/10.1136/amiajnl-2011-000335)
- Eijkemans, M. J., van Poppel, F., Habbema, D. F., Smith, K. R., Leridon, H., & te Velde, E. R. (2014). Too old to have children? Lessons from natural fertility populations. *Human Reproduction, 29*(6), 1304–1312. doi: [10.1093/humrep/deu056](https://doi.org/10.1093/humrep/deu056)
- Ellsworth, S. G. (1963). Utah's struggle for statehood. *Utah Historical Quarterly, 31*(1), 60–69.
- Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., . . . Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. *American Journal of Preventive Medicine, 14*(4), 245–258. doi: [10.1016/s0749-3797\(98\)00017-8](https://doi.org/10.1016/s0749-3797(98)00017-8)
- Gage, S. H., Munafo, M. R., & Davey Smith, G. (2016). Causal inference in Developmental Origins of Health and Disease (DOHaD) research. *Annual Review of Psychology, 67*, 567–585. doi: [10.1146/annurev-psych-122414-033352](https://doi.org/10.1146/annurev-psych-122414-033352)
- Gagnon, A., Smith, K. R., Tremblay, M., Vézina, H., Paré, P.-P., & Desjardins, B. (2009). Is there a trade-off between fertility and longevity? A comparative study of women from three large historical databases accounting for mortality selection. *American Journal of Human Biology, 21*(4), 533–540. doi: [10.1002/ajhb.20893](https://doi.org/10.1002/ajhb.20893)
- Garibotti, G., Smith, K. R., Kerber, R. A., & Boucher, K. M. (2006). Longevity and correlated frailty in multigenerational families. *The Journals of Gerontology: Series A, 61*(12), 1253–1261. doi: [10.1093/gerona/61.12.1253](https://doi.org/10.1093/gerona/61.12.1253)
- Greene, D., Tschanz, J. T., Smith, K. R., Østbye, T., Corcoran, C., Welsh-Bohmer, K. A., & Norton, M. C. (2014). Impact of offspring death on cognitive health in late life: The Cache County study. *The American Journal of Geriatric Psychiatry, 22*(11), 1307–1315. doi: [10.1016/j.jagp.2013.05.002](https://doi.org/10.1016/j.jagp.2013.05.002)
- Greeson, J. K., Briggs, E. C., Layne, C. M., Belcher, H. M. E., Ostrowski, S. A., Kim, S., . . . Fairbank, J. A. (2014). Traumatic childhood experiences in the 21st century: Broadening and building on the ACE studies with data from the National Child Traumatic Stress Network. *Journal of Interpersonal Violence, 29*(3), 536–556. doi: [10.1177/0886260513505217](https://doi.org/10.1177/0886260513505217)
- Hammad, I. A., Meeks, H., Fraser, A., Theilen, L. H., Esplin, M. S., Smith, K. R., & Varner, M. W. (2020). Risks of cause-specific mortality in offspring of pregnancies complicated by hypertensive disease of pregnancy. *American Journal of Obstetrics and Gynecology, 222*(1), 75.e1–75.e9. doi: [10.1016/j.ajog.2019.07.024](https://doi.org/10.1016/j.ajog.2019.07.024)



- Hanson, H. A., Leiser, C. L., Madsen, M. J., Gardner, J., Knight, S., Cessna, M., . . . Camp, N. J. (2020). Family study designs informed by tumor heterogeneity and multi-cancer pleiotropies: The power of the Utah Population Database. *Cancer Epidemiology, Biomarkers & Prevention*, 29(4), 807–815. doi: [10.1158/1055-9965.EPI-19-0912](https://doi.org/10.1158/1055-9965.EPI-19-0912)
- Hanson, M. A., Poston, L., & Gluckman, P. D. (2019). DOHaD: The challenge of translating the science to policy. *Journal of Developmental Origins of Health and Disease*, 10(3), 263–267. doi: [10.1017/S2040174419000205](https://doi.org/10.1017/S2040174419000205)
- Harrell, C. J., Smith, K. R., & Mineau, G. P. (2008). Are girls good and boys bad for parental longevity? The effects of sex composition of offspring on parental mortality past age 50. *Human Nature*, 19, 56–69. doi: [10.1007/s12110-008-9028-2](https://doi.org/10.1007/s12110-008-9028-2)
- Hawkes, K., & Smith, K. R. (2009). Brief communication: Evaluating grandmother effects. *American Journal of Physical Anthropology*, 140(1), 173–176. doi: [10.1002/ajpa.21061](https://doi.org/10.1002/ajpa.21061)
- Hawkes, K., & Smith, K. R. (2010). Do women stop early? Similarities in fertility decline in humans and chimpanzees. *Annals of the New York Academy of Sciences*, 1204(1), 43–53. doi: [10.1111/j.1749-6632.2010.05527.x](https://doi.org/10.1111/j.1749-6632.2010.05527.x)
- Hollingshaus, M. S., & Smith, K. R. (2015). Life and death in the family: Early parental death, parental remarriage, and offspring suicide risk in adulthood. *Social Science & Medicine*, 131, 181–189. doi: [10.1016/j.socscimed.2015.02.008](https://doi.org/10.1016/j.socscimed.2015.02.008)
- Horne, B. D., Camp, N. J., Muhlestein, J. B., & Cannon-Albright, L. A. (2006). Identification of excess clustering of coronary heart diseases among extended pedigrees in a genealogical population database. *American heart journal*, 152(2), 305–311. doi: [10.1016/j.ahj.2005.12.028](https://doi.org/10.1016/j.ahj.2005.12.028)
- Hsueh, Y., & Anderton, D. L. (1990). Temporal dimensions of the fertility transition: An age-period-cohort analysis of frontier fertility. *Sociological Perspectives*, 33(4), 447–464. doi: [10.2307/1389166](https://doi.org/10.2307/1389166)
- Hurdle, J. F., Smith, K. R., & Mineau, G. P. (2013). Mining electronic health records: An additional perspective. *Nature Reviews Genetics*, 14, 75–75. doi: [10.1038/nrg3208-c1](https://doi.org/10.1038/nrg3208-c1)
- Jennings, J. A., Sullivan, A. R., & Hacker, J. D. (2012). Intergenerational transmission of reproductive behavior during the demographic transition. *The Journal of Interdisciplinary History*, 42(4), 543–569. doi: [10.1162/jinh\\_a\\_00304](https://doi.org/10.1162/jinh_a_00304)
- Jones, J. H., & Bliege Bird, R. (2013). The marginal valuation of fertility. *Evolution and Human Behavior*, 35(1). doi: [10.1016/j.evolhumbehav.2013.10.002](https://doi.org/10.1016/j.evolhumbehav.2013.10.002)
- Jorde, L. B. (2001). Consanguinity and prereproductive mortality in the Utah Mormon population. *Human Heredity*, 52(2), 61–65. doi: [10.1159/000053356](https://doi.org/10.1159/000053356)
- Kauwe, J. S. K., Ridge, P. G., Foster, N. L., & Cannon-Albright, L. A. (2013). Strong evidence for a genetic contribution to late-onset Alzheimer's disease mortality: A population-based study. *PLoS ONE*, 8(10), e77087. doi: [10.1371/journal.pone.0077087](https://doi.org/10.1371/journal.pone.0077087)
- Kerber, R. A. (1995). Method for calculating risk associated with family history of a disease. *Genetic Epidemiology*, 12(3), 291–301. doi: [10.1002/gepi.1370120306](https://doi.org/10.1002/gepi.1370120306)
- Kerber, R. A., O'Brien, E., Smith, K. R., & Cawthon, R. M. (2001). Familial excess longevity in Utah genealogies. *The Journals of Gerontology: Series A*, 56(3), B130–B139. doi: [10.1093/gerona/56.3.b130](https://doi.org/10.1093/gerona/56.3.b130)
- Kerber, R., O'Brien, E., Smith, K. R., & Mineau, G. P. (2008). Familial aggregation of elderly cause-specific mortality: Analysis of extended pedigrees in Utah, 1904–2002. In T. Bengtsson & G. P. Mineau (Eds.), *International Studies in Population: Volume 7. Kinship and demographic behavior in the past* (pp. 243–258). Dordrecht: Springer. Available from [10.1007/978-1-4020-6733-4\\_11](https://doi.org/10.1007/978-1-4020-6733-4_11)
- Leiser, C. L., Taddie, M., Hemmert, R., Richards Steed, R., VanDerslice, J. A., Henry, K., . . . Hanson, H. A. (2020). Spatial clusters of cancer incidence: Analyzing 1940 census data linked to 1966–2017 cancer records. *Cancer Causes & Control*, 31, 609–615. doi: [10.1007/s10552-020-01302-3](https://doi.org/10.1007/s10552-020-01302-3)
- Lindahl-Jacobsen, R., Hanson, H. A., Oksuzyan, A., Mineau, G. P., Christensen, K., & Smith, K. R. (2013). The male-female health-survival paradox and sex differences in cohort life expectancy in Utah, Denmark, and Sweden 1850–1910. *Annals of Epidemiology*, 23(4), 161–166. doi: [10.1016/j.annepidem.2013.02.001](https://doi.org/10.1016/j.annepidem.2013.02.001)
- Lyman, E. L. (1998). Mormon leaders in politics: The transition to statehood in 1896. *Journal of Mormon History*, 24(2), 30–54. Available from <https://digitalcommons.usu.edu/mormonhistory/vol24/iss2/1>
- Lynch, K. A., Mineau, G. P., & Anderton, D. L. (1985). Estimates of infant mortality on the western frontier: The use of genealogical data. *Historical Methods*, 18(4), 155–164. doi: [10.1080/01615440.1985.10594160](https://doi.org/10.1080/01615440.1985.10594160)
- Maloney, T. N., Hanson, H., & Smith, K. R. (2014). Occupation and fertility on the frontier: Evidence from the state of Utah. *Demographic Research*, 30, 853–886. doi: [10.4054/DemRes.2014.30.29](https://doi.org/10.4054/DemRes.2014.30.29)
- McLellan, T., Jorde, L. B., & Skolnick, M. H. (1984). Genetic distances between the Utah Mormons and related populations. *American Journal of Human Genetics*, 36(4), 836–857.

- Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., . . . Skolnick, M. H. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, 266(5182), 66–71. doi: [10.1126/science.7545954](https://doi.org/10.1126/science.7545954)
- Mineau, G. P., Bean, L. L., & Skolnick, M. (1979). Mormon demographic history II: The family life cycle and natural fertility. *Population Studies*, 33(3), 429–446. doi: [10.2307/2173890](https://doi.org/10.2307/2173890)
- Mineau, G. P., Smith, K. R., & Bean, L. L. (2002). Historical trends of survival among widows and widowers. *Social Science & Medicine*, 54(2), 245–254. doi: [10.1016/S0277-9536\(01\)00024-7](https://doi.org/10.1016/S0277-9536(01)00024-7)
- Mineau, G., Smith, K., & Bean, L. (2004). Adult mortality risks and religious affiliation: The role of social milieu in biodemographic studies. *Annales de Démographie Historique*, 108(2), 85–104. doi: [10.3917/adh.108.0085](https://doi.org/10.3917/adh.108.0085)
- Mineau, G. P., & Trussell, J. (1982). A specification of marital fertility by parents' age, age at marriage and marital duration. *Demography*, 19(3), 335–349. doi: [10.2307/2060975](https://doi.org/10.2307/2060975)
- Moorad, J. A. (2013). Multi-level sexual selection: Individual and family-level selection for mating success in a historical human population. *Evolution*, 67(6), 1635–1648. doi: [10.1111/evo.12050](https://doi.org/10.1111/evo.12050)
- Moorad, J. A., Promislow, D. E. L., Smith, K. R., & Wade, M. J. (2011). Mating system change reduces the strength of sexual selection in an American frontier population of the 19th century. *Evolution and Human Behavior*, 32(2), 147–155. doi: [10.1016/j.evolhumbehav.2010.10.004](https://doi.org/10.1016/j.evolhumbehav.2010.10.004)
- Moorad, J. A., & Wade, M. J. (2013). Selection gradients, the opportunity for selection, and the coefficient of determination. *The American Naturalist*, 181(3), 291–300. doi: [10.1086/669158](https://doi.org/10.1086/669158)
- Nam, C. B., & Boyd, M. (2004). Occupational status in 2000: Over a century of census-based measurement. *Population Research and Policy Review*, 23, 327–358. Available from [10.1023/B:POPU.0000040045.51228.34](https://doi.org/10.1023/B:POPU.0000040045.51228.34)
- Nam, C. B., & Powers, M. G. (1983). *The socioeconomic approach to status measurement*. Houston, TX: Cap and Gown Press.
- Neklason, D. W., Stevens, J., Boucher, K. M., Kerber, R. A., Matsunami, N., Barlow, J., . . . Burt, R. W. (2008). American founder mutation for attenuated familial adenomatous polyposis. *Clinical Gastroenterology Hepatology*, 6(1), 46–52. doi: [10.1016/j.cgh.2007.09.017](https://doi.org/10.1016/j.cgh.2007.09.017)
- Norton, M. C., Fauth, E., Clark, C. J., Hatch, D., Greene, D., Pfister, R., . . . Smith, K. R. (2016). Family member deaths across adulthood predict Alzheimer's disease risk: The Cache County Study. *International Journal of Geriatric Psychiatry*, 31(3), 256–263. doi: [10.1002/gps.4319](https://doi.org/10.1002/gps.4319)
- Norton, M. C., Hatch, D. J., Munger, R. G., & Smith, K. R. (2017). Family member deaths in childhood predict systemic inflammation in late life. *Biodemography and Social Biology*, 63(2), 104–115. doi: [10.1080/19485565.2017.1281099](https://doi.org/10.1080/19485565.2017.1281099)
- Norton, M. C., Østbye, T., Smith, K. R., Munger, R. G., & Tschanz, J. T. (2009). Early parental death and late-life dementia risk: Findings from the Cache County Study. *Age and Ageing*, 38(3), 340–343. doi: [10.1093/ageing/afp023](https://doi.org/10.1093/ageing/afp023)
- Norton, M. C., Smith, K. R., Østbye, T., Tschanz, J. T., Corcoran, C., Schwartz, S., . . . Welsh-Bohmer, K. A. (2010). Increased risk of dementia when spouse has dementia? The Cache County Study. *Journal of the American Geriatrics Society*, 58(5), 895–900. doi: [10.1111/j.1532-5415.2010.02806.x](https://doi.org/10.1111/j.1532-5415.2010.02806.x)
- Norton, M. C., Smith, K. R., Østbye, T., Tschanz, J. T., Schwartz, S., Corcoran, C., . . . Welsh-Bohmer, K. A. (2011). Early parental death and remarriage of widowed parents as risk factors for Alzheimer disease: The Cache County study. *The American Journal of Geriatric Psychiatry*, 19(9), 814–824. doi: [10.1097/JGP.0b013e3182011b38](https://doi.org/10.1097/JGP.0b013e3182011b38)
- O'Brien, E., Kerber, R., Smith, K., Mineau, G., Boucher, K., & Reed, D. L. (2007). Familial mortality in the Utah Population Database: Characterizing a human aging phenotype. *The Journals of Gerontology: Series A*, 62(8), 803–812. doi: [10.1093/gerona/62.8.803](https://doi.org/10.1093/gerona/62.8.803)
- O'Brien, E., Rogers, A. R., Beesley, J., & Jorde, L. B. (1994). Genetic structure of the Utah Mormons: Comparison of results based on RFLPs, blood groups, migration matrices, isonymy, and pedigrees. *Human Biology*, 66(5), 743–759.
- Penkler, M., Hanson, M., Biesma, R., & Muller, R. (2019). DOHaD in science and society: Emergent opportunities and novel responsibilities. *Journal of Developmental Origins of Health and Disease*, 10(3), 268–273. doi: [10.1017/S2040174418000892](https://doi.org/10.1017/S2040174418000892)
- Penn, D. J., & Smith, K. R. (2007). Differential fitness costs of reproduction between the sexes. *Proceedings of the National Academy of Sciences of the United States of America*, 104(2), 553–558. doi: [10.1073/pnas.0609301103](https://doi.org/10.1073/pnas.0609301103)
- Rich, E. C., Burke, W., Heaton, C. J., Haga, S., Pinsky, L., Short, M. P., & Acheson, L. (2004). Reconsidering the family history in primary care. *Journal of the General Internal Medicine*, 19(3), 273–280. doi: [10.1111/j.1525-1497.2004.30401.x](https://doi.org/10.1111/j.1525-1497.2004.30401.x)

- Robson, S. L., & Smith, K. R. (2011). Twinning in humans: Maternal heterogeneity in reproduction and survival. *Proceedings of the Royal Society B*, 278(1725), 3755–3761. doi: [10.1098/rspb.2011.0573](https://doi.org/10.1098/rspb.2011.0573)
- Robson, S. L., & Smith, K. R. (2012). Parity progression ratios confirm higher lifetime fertility in women who bear twins. *Proceedings of the Royal Society B*, 279(1738). doi: [10.1098/rspb.2012.0436](https://doi.org/10.1098/rspb.2012.0436)
- Sanz-de-Galdeano, A., Terskaya, A., & Upegui, A. (2020). Association of a genetic risk score with BMI along the life-cycle: Evidence from several US cohorts. *PLoS ONE*, 15(9), e0239067. doi: [10.1371/journal.pone.0239067](https://doi.org/10.1371/journal.pone.0239067)
- Schacht, R., Meeks, H., Fraser, A., & Smith, K. R. (2021). Was Cinderella just a fairy tale? Survival differences between stepchildren and their half-siblings. *Philosophical Transactions of the Royal Society B*, 376(1827), 20200032. Available from [10.1098/rstb.2020.0032](https://doi.org/10.1098/rstb.2020.0032)
- Schacht, R., & Smith, K. R. (2017). Causes and consequences of adult sex ratio imbalance in a historical U.S. population. *Philosophical Transactions of the Royal Society B*, 372(1729). doi: [10.1098/rstb.2016.0314](https://doi.org/10.1098/rstb.2016.0314)
- Schacht, R., Tharp, D., & Smith, K. R. (2019). Sex ratios at birth vary with environmental harshness but not maternal condition. *Scientific Reports*, 9, 9066. doi: [10.1038/s41598-019-45316-7](https://doi.org/10.1038/s41598-019-45316-7)
- Scheuner, M. T., Wang, S.-J., Raffel, L. J., Larabell, S. K., & Rotter, J. I. (1997). Family history: A comprehensive genetic risk assessment method for the chronic conditions of adulthood. *American Journal of Medical Genetics*, 71(3), 315–324. doi: [10.1002/\(sici\)1096-8628\(19970822\)71:3<315::aid-ajmg12>3.0.co;2-n](https://doi.org/10.1002/(sici)1096-8628(19970822)71:3<315::aid-ajmg12>3.0.co;2-n)
- Sellers, T. A., Caporaso, N., Lapidus, S., Petersen, G. M., & Trent, J. (2006). Opportunities and barriers in the age of team science: Strategies for success. *Cancer Causes & Control*, 17(3), 229–237. doi: [10.1007/s10552-005-0546-5](https://doi.org/10.1007/s10552-005-0546-5)
- Shah, R., Pico, A. R., & Freedman, J. E. (2016). Translational epidemiology: Entering a brave new world of team science. *Circulation Research*, 119(10), 1060–1062. doi: [10.1161/CIRCRESAHA.116.309881](https://doi.org/10.1161/CIRCRESAHA.116.309881)
- Skolnick, M., Bean, L., May, D., Arbon, V., De Nevers, K., & Cartwright, P. (1978). Mormon demographic history I. Nuptiality and fertility of once-married couples. *Population Studies*, 32(1), 5–19. doi: [10.1080/00324728.1978.10412788](https://doi.org/10.1080/00324728.1978.10412788)
- Smith, K. R., Brown, B. B., Yamada, I., Kowaleski-Jones, L., Zick, C. D., & Fan, J. X. (2008). Walkability and body mass index density, design, and new diversity measures. *The American Journal of Preventive Medicine*, 35(3), 237–244. doi: [10.1016/j.amepre.2008.05.028](https://doi.org/10.1016/j.amepre.2008.05.028)
- Smith, K. R., Gagnon, A., Cawthon, R. M., Mineau, G. P., Mazan, R., & Desjardins, B. (2009). Familial aggregation of survival and late female reproduction. *The Journals of Gerontology: Series A*, 64A(7), 740–744. doi: [10.1093/gerona/glp055](https://doi.org/10.1093/gerona/glp055)
- Smith, K. R., Hanson, H., & Mineau, G. (2016). A six-pack of big ideas for historical demography. In K. Matthijs, S. Hin, J. Kok, & H. Matsuo (Eds.), *The future of historical demography* (pp. 211–214). Leuven/Den Haag: Acco.
- Smith, K. R., Hanson, H. A., Mineau, G. P., & Buys, S. S. (2012). Effects of BRCA1 and BRCA2 mutations on female fertility. *Proceedings of the Royal Society B*, 279(1732), 1389–1395. doi: [10.1098/rspb.2011.1697](https://doi.org/10.1098/rspb.2011.1697)
- Smith, K. R., Hanson, H. A., Norton, M. C., Hollingshaus, M. S., & Mineau, G. P. (2014). Survival of offspring who experience early parental death: Early life conditions and later-life mortality. *Social Science & Medicine*, 119, 180–190. doi: [10.1016/j.socscimed.2013.11.054](https://doi.org/10.1016/j.socscimed.2013.11.054)
- Smith, K. R., Mineau, G. P., & Bean, L. L. (2002). Fertility and post-reproductive longevity. *Social Biology*, 49(3–4), 185–205.
- Smith, K. R., Mineau, G. P., Garibotti, G., & Kerber, R. (2009). Effects of childhood and middle-adulthood family conditions on later-life mortality: Evidence from the Utah Population Database, 1850–2002. *Social Science & Medicine*, 68(9), 1649–1658. doi: [10.1016/j.socscimed.2009.02.010](https://doi.org/10.1016/j.socscimed.2009.02.010)
- Smith, K. R., Zick, C. D., Kowaleski-Jones, L., Brown, B. B., Fan, J. X., & Yamada, I. (2011). Effects of neighborhood walkability on healthy weight: Assessing selection and causal influences. *Social Science Research*, 40(5), 1445–1455. doi: [10.1016/j.ssresearch.2011.04.009](https://doi.org/10.1016/j.ssresearch.2011.04.009)
- Song, X., & Campbell, C. D. (2017). Genealogical microdata and their significance for social science. *Annual Review of Sociology*, 43, 75–99. doi: [10.1146/annurev-soc-073014-112157](https://doi.org/10.1146/annurev-soc-073014-112157)
- Stokols, D., Misra, S., Moser, R. P., Hall, K. L., & Taylor, B. K. (2008). The ecology of team science: Understanding contextual influences on transdisciplinary collaboration. *American Journal of Preventive Medicine*, 35(2 Suppl), S96–115. doi: [10.1016/j.amepre.2008.05.003](https://doi.org/10.1016/j.amepre.2008.05.003)
- Stroup, A. M., Herget, K. A., Hanson, H. A., Reed, D. L., Butler, J. T., Henry, K. A., . . . Smith, K. R. (2017). Baby Boomers and birth certificates: Early-life socioeconomic status and cancer risk in adulthood. *Cancer Epidemiology, Biomarkers & Prevention*, 26(1), 75–84. doi: [10.1158/1055-9965.EPI-16-0371](https://doi.org/10.1158/1055-9965.EPI-16-0371)
- Temby, O. F., & Smith, K. R. (2014). The association between adult mortality risk and family history of longevity: The moderating effects of socioeconomic status. *Journal of Biosocial Science*, 46(6), 703–716. doi: [10.1017/s0021932013000515](https://doi.org/10.1017/s0021932013000515)

- Theilen, L. H., Fraser, A., Hollingshaus, M. S., Schliep, K. C., Varner, M. W., Smith, K. R., & Esplin, M. S. (2016). All-cause and cause-specific mortality after hypertensive disease of pregnancy. *Obstetrics & Gynecology*, 128(2), 238–244. doi: [10.1097/AOG.0000000000001534](https://doi.org/10.1097/AOG.0000000000001534)
- Theilen, L. H., Meeks, H., Fraser, A., Esplin, M. S., Smith, K. R., & Varner, M. W. (2018). Long-term mortality risk and life expectancy following recurrent hypertensive disease of pregnancy. *American Journal of Obstetrics & Gynecology*, 219(1), P107.E101-107.E6. doi: [10.1016/j.ajog.2018.04.002](https://doi.org/10.1016/j.ajog.2018.04.002)
- Trivers, R. L., & Willard, D. E. (1973). Natural selection of parental ability to vary the sex ratio of offspring. *Science*, 179(4068), 90–92. doi: [10.1126/science.179.4068.90](https://doi.org/10.1126/science.179.4068.90)
- Tschanz, J. T., Pfister, R., Wanzek, J., Corcoran, C., Smith, K., Tschanz, B. T., . . . Norton, M. C. (2013). Stressful life events and cognitive decline in late life: Moderation by education and age. The Cache County Study. *International Journal of Geriatric Psychiatry*, 28(8), 821–830. doi: [10.1002/gps.3888](https://doi.org/10.1002/gps.3888)
- van den Berg, N., Beekman, M., Smith, K. R., Janssens, A., & Slagboom, P. E. (2017). Historical demography and longevity genetics: Back to the future. *Ageing Research Reviews*, 38, 28–39. doi: [10.1016/j.arr.2017.06.005](https://doi.org/10.1016/j.arr.2017.06.005)
- van den Berg, N., Rodríguez-Girondo, M., van Dijk, I. K., Mourits, R. J., Mandemakers, K., Janssens, A. A. P. O., . . . Slagboom, P. E. (2019). Longevity defined as top 10% survivors and beyond is transmitted as a quantitative genetic trait. *Nature Communications*, 10(1), 35. doi: [10.1038/s41467-018-07925-0](https://doi.org/10.1038/s41467-018-07925-0)
- van Dijk, I. K., Janssens, A., & Smith, K. R. (2019). The long harm of childhood: Childhood exposure to mortality and subsequent risk of adult mortality in Utah and the Netherlands. *European Journal of Population*, 35(5), 851–871. doi: [10.1007/s10680-018-9505-1](https://doi.org/10.1007/s10680-018-9505-1)
- Zick, C. D., Smith, K. R., Fan, J. X., Brown, B. B., Yamada, I., & Kowaleski-Jones, L. (2009). Running to the store? The relationship between neighborhood environments and the risk of obesity. *Social Science & Medicine*, 69(10), 1493–1500. doi: [10.1016/j.socscimed.2009.08.032](https://doi.org/10.1016/j.socscimed.2009.08.032)
- Zimmer, Z., Hanson, H., & Smith, K. (2016a). Childhood socioeconomic status, adult socioeconomic status, and old-age health trajectories: Connecting early, middle, and late life. *Demographic Research*, 34, 285–320. doi: [10.4054/DemRes.2016.34.10](https://doi.org/10.4054/DemRes.2016.34.10)
- Zimmer, Z., Hanson, H. A., & Smith, K. R. (2016b). Offspring socioeconomic status and parent mortality within a historical population. *Demography*, 53(5), 1583–1603. doi: [10.1007/s13524-016-0502-x](https://doi.org/10.1007/s13524-016-0502-x)