SURVEY OF HISTORICAL DATABASES WITH LONGITUDINAL MICRO-DATA

For more information about this questionnaire or questions about entering specific information, please mail George Alter (alter@indiana.edu) and/or Kees Mandemakers (kma@iisg.nl).

1/ Identifying information

	Name of database:	Utah Population Database
	Location:	University of Utah, Salt Lake City, Utah
	Web-address:	http://www.hci.utah.edu/groups/ppr/
	Name contact person:	Geri Mineau or Alison Fraser
	Email-address contact person:	gmineau@hci.utah.edu
	Postal Address:	University of Utah 2000 Circle of Hope Huntsman Cancer Institute Salt Lake City, UT 84112
	Tel. number:	801-581-4567 or 801-581-4468

2/ **Main objective and scope of the database:**

The **Utah Population Database (UPDB)** is a rich source of information for genetic, epidemiological, demographic and public health studies. For over 30 years, researchers have used this resource to identify and study families that have higher than normal incidence of cancer or other diseases, to analyze patterns of genetic inheritance and to identify specific genetic mutations. Demographic studies have observed and analyzed the trends in the fertility transition and changes in mortality patterns for both infants and adults. The UPDB is the only such database of its kind in the US and one of few such resources in the world. The central component of UPDB is an extensive set of Utah family histories, in which family members are linked to demographic and medical information. The UPDB also includes diagnostic records on cancer, cause of death, and medical details associated with births. The UPDB provides access to almost 9 million records and supports over 55 research projects. These data can only be used for biomedical and health related research.

3/ Sources: Please enter Yes or No and the time period for the main sources included in the database

Yes /no	**Start year**	**End Year**	**Type of source**	**Comments**

Yes	1830	1975	Baptisms	The Church of Jesus Christ of Latter-day Saints (LDS) was organized in New York in 1830. Baptism occurs at age 8 or later. Reported on LDS genealogies.
Yes	1790	1975	Marriages from church registers	Reported on LDS genealogies
Yes			Burials	Reported on LDS genealogies
No			Population registers (continuous) maintained by a church	
Yes	1915	2004	Civil birth certificates	Utah 1951-21 and 1947-2004; annual electronic updates
Yes	1978	2004	Civil marriage/divorce certificates	Utah records - annual electronic updates
Yes	1904	2004	Civil death certificates	Utah records - annual electronic updates
No			Civil population registers	
Yes	1880	1880	Census	1880 Utah Territorial Census
Yes	1978	1999	Utah fetal death records	
Yes	2000	2005	Utah Driver License	Annual electronic updates
Yes	1960	2003	Social Security Death Index	For SSNs issued in Utah and Idaho; plus selected others
Yes	1966	2002	Utah Cancer Registry	Annual electronic updates
Yes	1969	2004	Cancer Data Registry of Idaho	Annual electronic updates

4/ How was the sample (or samples) defined?

Genealogies: Over 185,000 Utah families were identified on "Family Group Sheets" from the archives at the Utah Family History Library which is maintained by the LDS Church. These sheets contain demographic and kinship information on three generations. For a family to be initially selected for UPDB from the genealogies, at least one member had to have a vital event (birth, marriage, death) on the Mormon Pioneer Trail or in Utah. Some family members have birth dates as early as 1740; most occur after 1780; the genealogy set ends with birth dates about 1970.

Other records: complete statewide records with no sampling; the fields that are available vary over time and fields that are computerized vary over time.

5/ Units of observation:
Please enter Y or N for each unit that can be followed over time

		Comments:
Yes	Individuals	All records are individuals records
Yes	Married couples	
Yes	Families	Individuals have been linked across generations to create multi-generational families
Yes	Households	Only on the 1880 census
No	Farms	
Y/N	Other:	
Y/N	Other:	
Y/N	Other:	

6/ Describe the geographic area under observation

Utah for the most part. Genealogies include migrants into Utah; early migrants originated in New England, Missouri, Illinois and Northern Europe. The families tended to settle in Utah and southern Idaho.

7/ Is information available about related individuals who are not in the sample?

Examples:

Marriage registers often include occupations of parents.
A population register sample may include everyone in the household of an individual in the sample.

Genealogies and civil death certificates include the names of parents who may never have resided in Utah. For some years the birth place of parents on death certificates is electronic.

8/ What events can be identified? Do events have dates?

Y/N	**Event**	**Are these events dated? Y/N/P(= partial dates, e.g. year only)**

Yes	Birth	Y/P	Genealogies have full birth dates. Utah birth certificates have full birth dates. From 1978 to current the birth certificate have the birth date of the parents, prior to that only the age of the parents was available.
Yes	Marriage	Y/P	Genealogies have full marriage dates. Utah marriage records: From 1978 – 1988 records have minimal computerized information, listing husband and wife name and age, marriage date, and county of marriage. After 1988 records have more complete information for the husband and wife, including birth date and birth place, education, number of marriages and type of marriage (civil or religious).
Y/N	Death	Yes	Genealogies have full death dates Utah death certificates have full death dates
Y/N	Migration	No	Comment: A general range of years could be imputed from the place information associated with events.
Y/N	Other:	Y/N/P	Comment:
Y/N	Other:	Y/N/P	Comment:
Y/N	Other:	Y/N/P	Comment:

9/ These questions describe the way observation is censored.

A) How do individuals enter observation?

Genealogies: These provide information on full sibships regardless of place of birth so there is no censorship.

Vital records: Utah birth records will not include children born in other states. Thus full sibships may not be available on all families. Utah natives, born from period 1921 through 1946 who are not on genealogies, may not be linked to their parents information.

B) How do individuals leave observation?

Genealogies: Families or family members that leave Utah may not be followed; we use the Social Security Index and Idaho cancer records to find some.

Vital records: Individuals born in Utah who leave the state are not followed; we use the Social Security Index and Idaho cancer records to find some.

Some research projects have included additional data, for example, the vital file from Centers for Medicare and Medicaid Services (CMS).

C) Are some entry or exit dates unknown?
Yes.
D) Are some entries or exits interval censored (i.e. the exact date is unknown, but it can be located between two known dates)?
Yes, if an individual is known to have died out of state then the date that they last had an event in Utah and the death date would provide some parameters.

10/ Residence and Household (Y/N/Partly)	
Yes	Can observations be linked to residential locations? Residence is available on death certificates, driver's license data, and for the mother on birth certificates.
No	Are the dates and locations of movements within the observation area recorded?
No	Are all individuals who lived in the households of members of the sample recorded?

11/ Kinship relations
A) How is kinship recorded in the sources?
Genealogies: "Family Group Sheets" from the archives at the Utah Family History Library contain demographic and kinship information on three generations. Vital records: Births and deaths include information on two generations
B) How deep (number of generations) is the available kinship information?
The Utah family histories represent pedigrees that may span as many as ten generations. The majority of families living in Utah are represented in this database with a special emphasis on genealogy records of the pioneers of Utah and their Utah descendants. Merging data from genealogies and birth certificates allows information for new generations to be added and most families can be linked across five generations. For example, looking at all individuals born in Utah in 1950, 79 percent have grandparent information available in UPDB and 67 percent have five or more previous generations documented in the UPDB.

12/ Linkage		
Which sources and units of observation have been linked?		
	Y/N/Partly	**Comments:**
Births/Baptisms	Y	

Marriages/ Divorces	Y	
Deaths/Burials	Y	
Population registers	Y/N/P	
Census	Y	
Utah fetal deaths	Y	
SSDI	Y	
UCR	Y	
CDRI	Y	
Other:	Y/N/P	

How is linkage represented in the database? For example, do all occurrences of an individual include a universal identification number? Are records linked to each other but not to a universal ID?

Yes, there is a universal identification number. Because one person may have many records, it is imperative that we have one demographic control record (person record) to represent an individual. Otherwise, researchers would have to view many records for one individual and decide which information they wanted from each record. Thus we create a "person record" that is composite of information contained on all the records for a single person. As additional records are linked, the validity and quantity of the information on a given person increases. Rather than link different record sets to each other, a new record set is linked to the composite individuals with information from all sources.

13 / What data structures have been added to the information in the sources?

		Comments:
Partial	Date of entry and date of exit by individual	Last living date – provides the last observation of the individual; it could be the death date or some earlier event.
Y/N	Events by individual	
No	Time constant information (date of birth, sex, etc.) by individual	
Y	Husband-Wife pairs	Included in a relationship table
Y	Mother-child and Father-child pairs	Included in a relationship table
Y/N	Other:	
Y/N	Other:	

Y/N	Other:	
Y/N	Other:	
14/ What reference/coding systems have been linked to the data?		
Yes	Occupational titles (like HISCO): Occupational Coding (Occupations and Industry on Births and Death certificates are coded, but incomplete) 1980 Census Occupation and Industry Classification 1990 Census Occupation and Industry Classification These codes apply to occupations for all years, not just 1980 and 1990 An occupation code and industry code were developed for the 1880 census data.	
Yes	Locations (including geo-referenced systems): Address on Utah driver license has been geocoded. The system used was Universal Transverse Mercator (UTM). We have eastings and northings; as long as you stay in the same UTM zone, these map to longitude and latitude. Also have census tract and census block related to 2000 census. Vital records contain place codes for city, county and state.	
Y/N	Other (religion, civil status etc.):	
Yes	Other: Diagnoses Risks, complications, malformations on Birth Certificates coded to ICD 9 for 1978-1988 Cause of Death on Death Certificates coded to ICD6 (1956-1957), ICD7 (1958-67), ICD8 (1968-78), ICD9 (1979-1998), ICD10 (1904-1933, 1999-2004) Cancer Diagnoses coded to International Classification of Diseases for Oncology (ICDO)	
Y/N	Other: Race – Vital records and cancer records have National Center for Health Statistics (NCHS) race codes	
Y/N	Other: Hispanic classification on vital records and cancer records	

15/ Have you developed any software for analysis or data extracting?
Please describe the capabilities and outputs of these programs.

Kinship Analysis Tools:

Kinclass: This program rapidly identifies arbitrary classes of relatives for a set of individuals according to a set of criteria, such as first and second-degree relatives who are still alive.

Dynaped: This program takes the output from a control dataset and a kinship dataset, and performs various types of statistical analysis. The functionality of this program includes

- Calculation of familial disease incidence
- Calculation of familial average phenotypes, e.g. familial excess longevity
- Estimation of founder relative risks via pedigree-structured Poisson regression
- Extension of above methods to alternative inheritance models (e.g. mitochondrial, x-linked, y-linked, imprinting, social)

Descriptive Tools:

Kinship Coefficients: Pairs of individuals are provided and their ancestors are compared to identify a common ancestor. If a common ancestor is identified, a kinship coefficient is calculated to indicate the degree of relatedness.

Pedigree Drawing: Provides all descendants for import into Progeny pedigree drawing program.