Lecture notes
from the summer school in historical demography and statistics, Cluj-Napoca, Romania, 12–19 June 2016

By Tor-Ivar Krogsæter

What these notes are and are not

These notes were written during the summer school in Cluj-Napoca hosted by Universitatea Babeș-Bolyai in co-operation with European Science Foundation and EHPS Network. The topic of the summer school was historical demography and statistics, aided by several computer workshops in Access and R.

All lectures have been duly marked with date and name of lecturer, as well as the title of the lecture. The notes however are *my own*, and only represent how *I* interpreted the lectures I had the pleasure of attending. Naturally, I tried my best to write as good and accurate notes as possible, but errors could have occurred. Where I was uncertain of something, I generally showed this with double or triple question marks (yes, I know, I was inconsistent).

Should anyone have questions about these notes, feel free to contact me either at tor-ivar.krogsaeter@uit.no or the1cannedman@gmail.com; I will do my best to reply to any questions anyone might have. I ask that anyone using these notes in any kind of publication (be it in private spheres, such as personal blogs, or public spheres, such as magazines or blogs belonging to organisations of any sort) will please contact me to request permission. Should anyone desire to quote my notes, here is the citation information:

- "Lecture notes from the summer school in historical demography and statistics, Cluj-Napoca, Romania, 12–19 June 2016";
- the author is Krogsæter, Tor-Ivar;
- the date and year of publication is 2 september 2016;
- its fixed URL is http://munin.uit.no/handle/???...;
- and it was published by
  - English: UiT Norway's Arctic University
  - Norwegian: UiT Norges arktiske universitet

On typography: Modern day's availability of computers, has facilitated easy access to powerful tools for creating wonderful documents; unfortunately, these tools are too often used without any time spent educating oneself in the typographic skills necessary to fully make use of said tools. I have spent quite some time trying to master these things during the writing of my master's thesis, and hopefully this will show in the present document. True small caps are used (hopefully correctly and consistently), but ligatures I have unfortunately not been able to enable, as OpenOffice and LibreOffice currently only support this by way of graphite fonts. Microsoft Office, however, supports these OpenType features quite well, but is horrible at reading ODF Documents. I hope the reader can forgive the lack of ligatures in this document (except the final pages, manually inserted, ɔ: by cheating). Hopefully OpenOffice and LibreOffice will catch up in the future, particularly considering how many years they spent being far superior to Microsoft's solution in all the important ways. It apparently is not a problem for users running Linux or MacOS.

I hope you will find interesting information in this document, and wish you, dear reader: *cūra ut valeās!*

13 June 2016: Siegfried Gruber

Sources for historical demography

Censuses

Microdata vs. Macrodata

Provides all data on each individual

Allows investigation on small groups.

Mass data is no longer a problem to handle. This allows broader comparisons.

Problem with how the data is handled: To save money, it is sent to cheaper countries for transcription.

Sources

Anything can be a source

Private sources, such as diaries, can describe entire lives.

Images as sources are now possible to store practically limitless. This allows people access to this kind of data without having to visit archives.

Two major differences in what one can deduct: composition or change

Cross-sectional data

Describes the population characteristics at one point in time.

Change can be hard to find: A child born after a census but died before the next, will be invisible.

Longitudinal data

Covers many points in time, such as date of birth, marriage, baptism.

This data generally don't have income information, information on population size etc.; this needs to be covered by a second source.

Numerous comparative sources, such as church books, available.

Census list

"A census of population may be defined as the total process of collecting, compiling, evaluating, analyzing and publishing or otherwise disseminating demographic, economic and social data pertaining at a specified time, to all persons in a country or delimited part of a country" (UN 1998, p. 3)

For a long time there were separate censuses for civilians and military personell.

Characteristics

National legal authority; defined enumeration are, completely covered; simultaneous, individual and periodic enumeration; publication and dissemination of results.

Changes in time in census lists

General practice in Europe since 19th century; USA required by constitution.

Early lists had very general information; later would add more information, such as income, disease and other sickness, etc.

Now most countries' censuses are done registry-based. This has several problems

People hiding, houses that are hard to find, etc.

Historically

Early civilizations had census, particularly to find out who and from where they could collect taxes. This of course, led to the above mentioned problem.

Earliest modern censuses: France/Canada 1665, Iceland 1703.

Lithuania 1764: Jewish census, due to special taxations on Jews. Only blank, handwritten paper.

Iceland 1801: Same as in Denmark and Norway. Census forms printed in Danish; Danes establishing a working democracy.

Dép. de la Roer, Borcette/Burtscheid 1812: Everything in French, except the names (German): Part of Napoleonic France. Up until that time, in France it was common to only enumerate children under the age of 14; only children 14 and up were listed nominatively.

Mecklenburg-Schwerin, 1819: Done over a long period of time, simply because it took a long time to finish.

Wallachia, 1838: Again, not printed paper. Households enumerated, further information on details.

Austria-Hungary, 1869: Long descriptions on every column provided. Bilingual. Religion. Area one belonged to (because that was where one would get social support). Absence or presence.

Serbia 1884: Every five years. Printed. Administrative division entered on top. Religion; citizenship. Disabilities. Several columns left empty: 100 % of standard, or lazy census taker.

Russia 1897: Only real census of Russia. Done for the entire empire. Social class included, as well as disabilities and education, military service.

Mecklenburg-Schwerin 1900: 1 page for every person. Very detailed information given on several areas: date and place of birth; employment; actual place living; religion; citizen of German empire, and if not, which state; for military personnel where they have served.

Albania 1918: During the end of WW1. Bilingual, based on Austrian-Hungarian model. Ethnicity; reading and writing skills; degrees of farming capability; column for occupation; column of whether or not in agriculture. No information on people missing.

Registration systems

Starting point: All information gathered. Thereafter, system is updated.

Some started in 18th century, but incomplete. In Belgium (1846) and Netherlands (1850) base inventory of population plus all characteristics and changes.

Forms included not just people leaving, but also where they were leaving to.

The Ottoman register 1874.

Status animarum

These sources are mostly in the parishes or the bishop's archive.

Büchenau 1749. No printed forms, just notes written on blank paper, separated by a line. Also: It is written in Latin.

A later form can be seen from Münster in 1750. Specifically asking for servants; if not included, you miss a lot of population. In some censuses, servants could be counted both at their family and at their household of work. In protestant areas, one could also find notes on how well schooled they were in their catechism. In Latin

Ebensee 1779.

Moldavia 1782: For Catholics. Still in Latin.

Register-type family book, Diocese of Rottenburg: Husband, wife, birth dates, their children. No information on people leaving.

Registration of vital events

In Europe, Christian churches developed registration systems for births, marriages and deaths: The areas of interest for them.

Civil registration occurred later (France).

In France, the king ordered books on baptism since 1539.

Rituale Romanum, 1614: Liber matrimonium, liber status animarum, liber defunctorum.
The Liber animarum was important for good registers amongst others on who were alive to fight.

Protestant churches:

More regional → no general instructions, though similar to Catholic. Only England, France and Scandinavian countries had common, standardised books by the 18th century.

Baptism, Tragöß 1737: Latin, no printed paper. Ordered by date.
Baptism, Tragöß 1823: Printed form, German (?). Year, date, place, house number, child's name, religion, gender, literacy, father and mother, godparents' name and standing/occupation/relation.

Marriage, Tragöß 1711: Not printed. Ordered by date.
Marriage, Tragöß 1842–43: Printed, German. Information mostly like 1823 baptism.

Deaths, Tragöß 1793: Now in German, though not printed. Date of death, date of burial, name, page, who performed the burial, cause of death.
Deaths, Tragöß 1836: Now printed. Religion, sex, age added.

Same can be seen in Ottoman areas, though not as complete.

Other sources

Tax lists of various kinds. Various "soul revisions".

Digitisation projects

Much has been made available online now. France has made their data available, searchable by village. Austria has done similar work with church books.

East and Southeast Europe generally worse off. Census taking started later. As in Western, Central and Northern Europe, they later started having the people themselves fill them out → problems with understanding undermines the quality of the data.

Many sources destroyed by e.g. wars, but also the administration themselves.

Questions to sources

Who counts?

Who would avoid being counted? Criminals, illegal immigrants, etc.

Mobile populations are difficult to enumerate.

Information about ethnic and/or religious minorities.

Challenges

Incomplete or erroneous data, e.g.: child older than parents.

Ambiguous, fuzzy or missing data: Remarriages could make it hard to know whether or not age of children is correct.

Transcription is time consuming.

Research approaches

Life course approaches: What changes happen during life?

Dyadic relationships: Who lives with whom?

Analysis of household structures.

Challenges in analysis of household structures

Analysis of persons and not families or households.

Pay equal attention to enumerator as well as denominator.

Examples

IPUMS USA

Minnesota Pop. Center.

Samples censuses since mid 19th century.

All done in the same database structure.

IPUMS International

Collects data from all 79 countries.

Requests samples of data from 1960 to present.

560 million persons recorded.

All done in the same database structure.

NAPP

Minnesota Pop. Center.

Samples mainly immigration data, i.e. data from GB, Scandinavia, Ireland and Canada.

Common features of the projects

Data is free, but registration required. Proper citation required.

Standardised variables.

Occupations are coded into HISCO (OccHISCO).

Allows cross-country comparisons.

13 June 2016: Peter Teibenbacher

Nuptiality

Definition

Nuptiality refers to marriage as a population phenomenon, including the rate at which it occurs, the characteristics of people united in marriage, and the dissolution of such unions (through divorce, separation, widowhood and annulment).

nupitae, nuptialis

Indicators

Number married and never married

Marriage rates (crude and age specific).

First marriage mean and median age, SMAM (singulate mean age at marriage)

Gender gaps: Portions, age at marriage, widowhood

Years in marriage, i.e. how many years people spend in marriage.

Related issues

Family is normally defined via marriage in the past. Nowadays, people live in various forms of unions.

Polygamy–monogamy

Divorce rates (crude, age specific, marriage period specific). Elites got divorces more easily. It was more expected amongst normal people to stay within marriage.
The protestant church allowed divorces more easily, since it wasn't a sacrament? At least the fact of their refusal of the role of the Pope mattered.

Widowhood, separation and annulment

Remarriage. Marriage restrictions.

Illegitimacy: Informal marriage → legal bastards.

European Marriage Pattern: EMP.

Main trends in nuptiality

Arranged → love-marriage

Early marriage → late marriage. In EMP areas: a U-curve (premodernity later, modernity earlier, postmodernity later again). The tendency in premod.: Requirement of capability to support, in particular in rural areas. Marriage age decreased during the 20th century, particularly in the 50s and 60s; for women, the marriage age dropped from 27 to 22.
In other parts of the world, the development is more linear from low to high age.

From societal rules and economic constraints to individual, deliberate decision making.

Holy → secular: In the past, also outside Christian Europe, marriage was also a holy issue.

Formal → cohabitation (PACS): Pacte civil de solidarité

Marriage as a religious issue

Across the world, marriage traditionally has in most places been a holy issue. This poses problems in many areas when one wants to remarry.

Concilium Tridentinum (1563) decreed that each marriage was to be registered in registration books by the local priest. In modern times, states began using these books to control registration and to use them for administration and juridical issues.

Marriage as a state issue: civil marriage

The enlightenment: Civil marriage became required in most countries; the holy part of it was an addendum. In China (1950), similar laws were enacted, overruling old tradition.

Civil marriage allows easier divorce. Church remarriage, however, often remained as difficult.

Marriage as socio-sexual unification

Historically considered the sexual unification of a man and woman meant to create legal children. This is the reason it in most countries has been connected to the age of sexual marriage.

South-Africa: Already had a historical institution of woman–woman marriage, a purely socio-economical institution.

Monogamy and polygamy

Christian societies, as well as Hinduism, mostly claim monogamy to be the only acceptable form of marriage; Mormons, though, have commonly practised polygamy till the mid 50s.

China: Big mother (biological mother, but not necessarily), also called queen wife; the others were "aunts" or similar.

Hammarubi laws, 18th century B.C.E.: As long as the first wife didn't give children, he could marry a second wife, though she would be inferior to the first.

Polyandry: Causes? Male surplus?

We can assume that in earlier times, polygamy was a good solution for ensuring reproduction, as men tended to die early due to hunting and war. Later, it is likely that this social structure led to the male dominance of women.

Marriage as a societal issue

Particularly in early 18th century, the poor, the non-landed people or the non-housed people were restricted from marriage, maybe for risk of over-population. These restrictions were not commonplace in earlier times. When industrialisation set in, over-population was no longer an issue, due to both increased need of labour, as well as increased food-production capability.

In some black African tribes, there still today are restrictions on men marrying without providing at least sufficient cattle or land.

Today, marriage ages increase in developing countries, due to desire of securing a good education and a good job before trying to support a family.

Regulations and restrictions have always been employed by the rich against the poor. This, however, was against the Christian doctrine, in which it is told that marriage is a Christian sacrament and duty.

Marriage restriction → opportunity to marry → opportunity not to marry → in Japan: self-declaration of opportunity to not be in any relationships at all. This is especially important for women: They are now empowered to provide for themselves.

Universal marriage (percent of people ever married) is high (80 %+) outside Europe; early marriage as well. Tendency leads towards the same marriage ages and universal marriage rate as Central/Western Europe (60 % universal marriage rate in Europe). Anti-sexual/anti-nuptial society the next step?

In pre-modern times, marriage was dominated by parents.

Fire-horse syndrome (Japan): Unlucky years and superstitions causes drop in birth rates.

Who marries whom?

In civil law, first cousins were allowed to marry; in church law, however, this was prohibited.

It was common for brothers-in-law married the widowed, to strengthen the family-tribe. This same practice can be seen in between cousins (consanguinity), which is especially prevalent in Muslim countries.

The family

In early society, the bride would leave her family to live with her husband.

Later, especially in Western and Northern regions, the new couple created a new, small family: the nuclear family. In Africa, Asia and the old Americas, however, joint families were the normal case and the bride entered the groom's family as an association; these joint families were also common in Eastern and South-Eastern Europe. In earlier times, naturally, where the young couple often would stay with his parents for some time, this often soon deteriorated into a nuclear family, due to short life expectancy.

Remarriage

Women not too old (usually > 50) were expected to remarry when widowed, though if the farm had male heirs, this would have to wait until the son inherited the farm (so as to not cause confusion of power hierarchies).

Marriage as an economic issue

Geographical differences: Women were more empowered in Scandinavia and some parts of Europe, than in Africa, Asia and the rest of Europe. They often could not own anything, and therefore were depending on their husband for their livelihood.

Marriage is expensive. Italians spend € 11 200 just for their wedding; in Austria, it's less than half this. In former times, the bride's parents had to take care of the wedding. In other words: Marrying a daughter was more expensive than a son. In modern China, *naked marriages* (no ring, ceremony, honeymoon, home, car) has become a way of protesting against modern consumption pressure and status symbols. The man is not supposed to provide a home for the bride; the bride is not expected to be in need of the grooms financial support.

Marriage as a seasonal issue

Time of year could also matter: Back in the days, when the harvest was (ability to provide a feast); nowadays, when it's a good time to go on a honeymoon.

Nijmegen: Marrying increased during the festive May month.
China: Marrying increased around the beginning of the new year. Marriage was both a festive and a serious matter, so providing as much luck as possible was considered important.

Fertility

Numbers for modern families are incompatible with older numbers, as such a large part of the population live in cohabitation. Married couples have low fertility rates: 1.2–1.3 or so.

(Western) European Marriage Pattern (W)EMP

Different ways of dividing Europe in areas by demographic patterns.

Western Europe vs. Eastern Europe

Higher age at first marriage (> 25 vs. < 25)

Lower portions of universal marriage (< 80 % vs. > 80 %)

Higher portions of never married women (20–40 % vs. 10–20 %)

More non-marital births (> 15% vs. < 10 % in Eastern Europe, usually < 5 %)

Africa, Asia, (Pre-Colombian Americas)

Overall marriage, low age at first marriage, almost no non-marital births.

13 June 2016: Kees Mandemakers

Database design

Historical sources

Of varying kind. In general, all sources have some structure, otherwise, we couldn't read it. Of course, this structure isn't explicit. Structured data, on the other hand, already provides the meaning to the pieces of information.

Studying historical sources can give insight into things such as why some die at a younger age than those born just a year or two earlier or later; why some grow taller than others.

Databases summed up

DB = Storage system of data suitable for processing data in all kind of ways.

DBs allow numerous ways of sorting, searching, selecting and reworking the data. It allows working with complex, related data; controlled data entry (e.g. controlling high/low frequency entries to locate errors); viewing data in several ways; simple data analysis (grouping, counting).

What a DB is: Table, entity, key, record, attributes, relations. Knowing these things → understanding the basic principles of DBs.

Table

Creating separate tables for each category of information, one can more easily organise the data, by linking the tables. A combination of rows for some certain entities.

Entities

Defines what's in the table, i.e. the characteristics of that which is described. E.g.: You would not put a column to tell it's grey (if they're all grey), but length of trunk, sex, power, etc. E.g. 2: In a table about a person, you would not put in info that he has a heart, lungs, etc., as that isn't relevant to the *person*; rather, it is relevant to a separate table about *species.*

Key

The primary key provides each record it's unique identification

Record

= Row.

Attributes

The different entries in a record (= row).

Designing a DB

Entity Relationship Diagram (ERD). Distinction between entities and relations between entities. Tables are records with entities and keys.

First three rules of harmonisation:

1. No repeating fields.
2. All attributes belong to the same entity (functional dependency)
3. No subdependencies in one table (except secondary keys)

Example of this: For a course with a set number of students, creating keys for student_1, student_2, etc., would violate the first rule; creating a sub-table for student, with name, birth-date and address, as well as id_course, would violate the second rule, because there are two IDs connecting the name, birth-date and address. A good simple DB would have five tables, one for course, one for student, one for registration, one for teacher and one that connects them, maybe called teaching. In an ERD diagram, this would be something like this:
student → registration (n:m) ← course → teaching (n:m) ← teacher

A secondary key does not have to be unique; it is unique in the table in which it belongs. Eksempel: I befolkningstabellen vår, står PID-en unikt til hver person, mens fødestedet har en unik fødselskode (som er en sekundærnøkkel). I tabellen over fødestedene, derimot, er dette den unike nøkkelen, primærnøkkelen.

Spreadsheets

Do not use a spreadsheet as a database; rather, export data to the spreadsheet to further manage it or perhaps to create graphs.

Exercise: Creating a database

New York census of 1880

Tables:

- [person]
  - id_person {no.}
  - address_street {char}
  - address_house_no {char}
  - dwelling_no {char}
  - family_no {int}
  - first_name
  - last_name
  - race {single letter}
  - sex
  - age
    - age _int {int}
      age_frac {char}

        - age_month {int [if age_int < 1]}
- [house]
- [geography]
    - id_geo {no.}
    - state ↓
        - [states]
            - id_state {no.}
            - name_state
    - county ↓
        - [counties]
            - id_county {no.}
            - name_county
    - city ↓
        - [cities]
            - id_city {no.}
            - name_city
- enumeration_info

14 June 2016: Péter Őri

Life tables

Mortality

Demographic phenomenon characterised by a special vital event, i.e. deaths.

How many people die between age x and $x + n$? What proportion of life is alive at age x?

Sources

Parish registers, written by local clergy, can now be used for modern demographic analysis. During the 18th and 19th century its content became increasingly complete. With women, several pieces of information were included:

- The name of their husband.
- Their estimated age, as guessed by the priest or those nearby
- The cause of death, estimated by the local priest, though this is often hard to compare to today's data, due to lack of expertise, amongst many things.

The possibility of measuring mortality

Often times, the only thing that can be told clearly, is simple matters such as sex and age.

$$RoI = \frac{n \text{ events by } t}{m \text{ members exposed by } t}$$

Illustration 1: Where RoI *is Rate of Incidence,* t *is time, and "members exposed" is to be understood as "members exposed to risk of incident".*

Persons at risk of dying: The problem of the denominator.

- Anyone present at any time during the interval
- Persons present in the middle of the interval
- Persons present at the beginning of the interval.

Crude rates: calculated for the whole population:

- Numerator: number of events (deaths)
- Denominator: mid-year population (pop. size at the beginning of the year + pop. size at the end of the year]

$$\text{crude rate} = \frac{p_1 + p_2}{2}$$

Illustration 2: p*: population size at the (1) beginning and (2) end of the year*

The problem of crude death rates

All population is at risk of death at all times. Crude rates refer to the whole population, not those at realistic risk; in other words, they don't account for the composition effect (age, sex, etc.). Mortality varies by many aspects. The solution to the problem is to add age-specific rates:

$$\text{RoI}(\text{age}) = \frac{n_t(\text{age})}{p(\text{age}) \leftarrow \text{risk}_t}$$

Age-specific **R**ate **o**f **I**ncidence =
(**n**umber of events that occur in a given **t**ime interval to persons of a certain **age**) ÷
(number of members of the **p**opulation of that **age** exposed to the **risk** of event during the same **t**ime interval).

By dividing the population into age groups, one can get a clearer picture of this:

- The population for each age group is enumerated.
- The death toll for the same age groups is enumerated.
- By dividing the death toll by the population size of the various age groups, one gets a picture of how high the death risks were for each age group; multiplying these two columns' information (population by age group and age-specific death rates), one gets the number of people expected to die in each age group)
- The differences between the sexes: Breaking down into different age groups by sex? Question not understood by lecturer.

When only having rough death numbers by age category, or when working with small population sizes, problems arise. By using age-specific death rates and comparing these to actual death, one can compare the deaths and normalise the numbers. Example:

- In Zsámbék, age group 0–5 have 565 people; in Perbál, it is 293; in Tök, it is 144.
- The age-specific death rate for Pest county, is 0.093.
- The expected deaths numbers, therefore are
  - Zsámbék: 565 × 0.093 = 53
  - Perbál: 293 × 0.093 = 27
  - Tök: 144 × 0.093 = 13

Summing the *expected* number of deaths and comparing this to the *actual* number of deaths, yields a comparative mortality rate, for the mentioned towns the number yielded 1.61, 1.83 and 1.36. This number could again be compared to the crude death rate, to get an expression of the actual crude death rate for smaller places.

Life tables

Age-specific mortality rates: m_x

Useful functions:

- ${}_nq_x$ Probability that a person alive at age x will die before age $x + n$.
- l_x Number of people alive at age x.
- ${}_nd_x$ Number of deaths by age x.
- ${}_nL_x$ Number of person years lived between age x and age $x + n$.
  $L_x = l_{x+1} + .5(.2 \text{ or } .3)d_x$
- ${}_nm_x$ Death rate for persons between age x and age $x + n$.
- T_x Total person years lived after age x (the sum of L_x values)

The root of the death table: Population number.

Characteristics

A life table may be longitudinal. People may be followed over time; one can register the death cases, and in the end create the death tables. The problem is, of course, that it will take around a hundred years to create such a table; as such, we usually work with older information.

It may be cross-sectional, based on the events observed in a calendar year, or as mentioned above – longitudinal – basing it on the events observed in a given cohort or population (e.g. in a birth cohort), giving the proportion of those surviving in longitudinal data.

Probability of surviving between age x and age $x + 1$: ???

Concluding remarks

Life expectancy at birth is *not* identical with the mean age at death, except in the rare case of closed populations with zero growth, where the age structure, fertility and mortality are stable. Example of a normal case: Pest county: $e_0 = 36.6$ vs. mean age at death 23.3

Separate life tables for males and females are important, and its significance grew in the 20th century.

Family reconstitution

A base approach to connect fertility, marriage, death, etc. Goal: reconstruction of family trees or ancestor tables across generations and space, by genealogists, historians, ethnologists, etc.; reconstruction of all family units across generations living in a given place, by the same plus local historians; or, in historical demography (since about 1956), the identification of family units based on marriages, all family members and demographic events belonging to these family units.

The latter method was developed by French demographers, particularly connected to the *annales* school. Most of these studies refer to villages or local populations; samples of larger towns is not so common. In Germany, the local genealogies are known as Ortsfamilienbücher. In datasets, such as the one provided by Germany (www.online-ofb.de), one can reconstruct for examle the reproductive history of women by comparing demographic events (birth-date, marriage, child-birth, death-date). Parish records are our main source. Migration leaves holes in these records, which can be identified by incomplete dates; complete dates often mean that this was *after* the person immigrated to the parish.

	Parish registers	Vital statistics	Censuses	Genealogies	Household registers
Longitudinal	✓	✓		✓	✓
Individual level	✓		✓	✓	✓
Detail on households			✓		✓
Geographic specificity	✓	✓	✓		✓
Complete community	✓	✓	✓		✓
Population at risk		✓	✓	✓	✓
Timing of vital events	✓	✓		✓	✓

Household registers contain all the demographic information at the same time, as well as information on the people living in that specific location.

Source information

Censuses or census-type sources are the most ancient, but are not regularly repeated before the 19th century. Census-type information should be recorded regularly, and this can be found in some few parish records, allowing a fairly complete reconstruction of these villages or parishes.

Research questions

In the 17th–18th century, mercantilism was the leading idea. Society saw stagnation or slowly growing populations, and the goal was marriage control, migration, pro-natalism and more. By the 20th century, in particular the latter part, a rapidly growing population (especially in developing countries) has been prevalent. Developed countries, on the other hand, face an ageing population matched by reduced fertility. These factors contributed to the demographers' focus on fertility, trying to understand which factors affected these changes.

The demographic transition theory was developed by Davis and Notestein in 1945, leading to strong theoretical, empirical and methodological development. Computers greatly changed our opportunities for studying populations. Social history became increasingly interested in demography; again, led by the *annales* school. The new paradigm favoured parish records as sources and family reconstitution as method.

These new demographers were interested in finding the transitional points, and there was a focus on the differences between premodern and modern societies. Social historians focused on long term changes of mortality. They showed how long-term changes could show development in connection to food prices, growth-limits in premodern times and the development of new demographic regimes. Demographers, on the other hand, were more interested in particularly fertility studies. The historians and demographers agreed on the sources, though.

Louis Henry was leading in suggesting how the sources could be used in a subtle, indirect way. Fleury and Henry's book of 1956 as well as Gautier and Henry's of 1958 and Pierre Goubert's publication of 1960 (*Beauvais et le Beauvaisis de 1600 à 1730*) were of great importance and impact.

Parish records as sources

The contents of these data sets can be very different from each other, and therefore hard to use. The different demographic events recorded, might not contain the same information, as different information was considered more or less important in different contexts. Civil death register of the last centuries, often contain detailed information on several subjects relevant to the individual and the incident of death itself.

The problems of quality

Under-registration poses a strong problem: An example could be times of epidemics; another example is infant deaths, as the time between birth and death could be crucial for whether or not the death was registered; of course, if the clergyman himself died or wasn't present in the village, much time could pass before he was replaced (in case of death), leaving holes in the records.

Length of time between birth and baptism to estimate the accuracy of infant death registration. Deceased infants and children in the birth register could help estimate the accuracy of birth registers (problem of migration). Infant mortality rate by age cannot be very low; if it is, it is a clear sign of incompleteness. In the same matter, the percentage of infants and children amongst dead must reach a certain level; if it doesn't it can show the inaccuracy of registration of infant child deaths. Also, unusually long intervals between births of numerous repetitions of first names, could indicate the problems of registration.

The process of family reconstruction

The first thing in need of doing is the registration of all events (earlier on cards, today digital). After this, the standardisation of names is necessary, and this often has to be done manually. This manual editing, also often includes deciding family relations, such as to whom the child belonged. Automatic reconstruction may then commence.

Depending on the kind of data available, more or less accurate studies and/or conclusions may be drawn. There are four main categories of family cards used:

- MF: Date of marriage and date of dissolution is known
- MO: DoM known, DoD unknown
- EF: DoM unknown, DoD known
- EO: DoM unknown, DoD unknown

Nominative analysis

First interest: Calculate fertility rates. If one can reconstruct the number of years spent in marriage by age groups, one can calculate her "at risk"-rate, i.e. her chance of getting pregnant. It is crucial to know the date of marriage and the date of dissolution. Without this, age-specific rates cannot be calculated. On the non-married women, we have no information at all.

Natural fertility – birth control

There is great variety in the fertility of historical populations. No "natural fertility" in actuality; rather, there were very different levels of fertility, influenced by time spent in marriage and length of birth intervals. There is, however, evidence of early birth control in som regions, towns and/or social groups well before the industrial revolution. The differences between fertility regimes of traditional and modern (transitional) societies are hard to map.

14 June 2016: Kees Mandemakers

Access and Query Language

What is a query and types of queries

- Selection query: makes a selection from the table.
- Table query: what you have selected is written as a table.
- Append: to add records to an existing table.
- Update: changing information within an existing table.
- Delete: deletes records in a table.

Operators

Delimits information

- Relational operators (numerical): `>`, `<`, `=`, `<>` (not equal), `>=`, `<=`
- Relational operators (text): `=`, `<>`, `like *`
  - `like`: To identify things containing a phrase; function similar to `dir` in DOS.
- Logical operators: `and`, `or`, `not`, `xor`?

The `Σ`-button works on `select` and `make table`.

Joining tables

Three types:

- Inner join: only records that are fulfilled in both tables, effectively an `and` operation.
- Outer join: only records that are not fulfilled in both tables, effectively `not` operation.

Nothing in DBs

Numerically: `0` or `null`. Text: space, tab and null.

Functions

When writing functions, remember to surround variables by brackets. Example: `mid([Occ_father;1;5])`, which starts in the first position and shows the first five characters.

Social stratification/mobility

Marriage certificate

Usually contains information of *two* generations in *two* directions: both the bride's and the groom's parents. Witnesses might also be included. Social background of the couple can be further investigated by looking into the certificates of their children again, or the marriage certificates of the mother and father. Doing this, we have created a *social model* for three generations.

This kind of modelling cannot be done in programs such as Access; instead the data is prepared in a DB software, and (modelled) analysed in a statistics program.

Social stratification

Approaches, principles

To map the social background, two questions:

1. Which approach?
2. Which data to use?

1.a.: Social positions can be mapped on a scale of say 1–100, i.e. "occupational status". **S**ocial **E**conomic **S**tatus Attributes, like average schooling and income, is not used for this.

1.b.: Another way to do this, would be by basing the position in production process (Marxist tradition), or by position in the labour market (blue/white collar), income, education, responsibility, directing, ownership, technical level. For this, sc. Weber, Goldthorpe. The Marxist tradition basically works with the three classes: owners employing others, owners employing themselves, and labourers (without capital). Goldthorpe focuses on market–work relations:

1. Higher grade professionals
2. Lower grade professionals
3. Routine non-manual workers
4. Small self employed ("petty bourgeoisie")
    1. Small self employed without employees
    2. Small self employed agriculture
    3. Supervisors manual work, technicians
5. Supervisors
6. Skilled manual work
7. Unskilled manual work
    1. Unskilled workers
    2. Farm labourers

The service class in Western societies has grown very large, more than every second person.

2. Income, wealth, hereditary titles, position, occupational titles, data about social exclusion (caste, religion), social prestige.

The practice of coding

Original title →
standardisation (spelling) →
coding (HISCO) (→ coding HISCAM) →
coding (HISCLASS)

Historical International Standard Classification of Occupation:

Specialised for historical data, but designed for modern history (20th century and onwards). There is a small inclination to "dialectisation", and new titles do not have a standards committee for approval. NAPP HISCO seems to alleviate this problem.

HISCOs main structure is based in the following:

- 0/1: Professional, technical and related
- 2: Administrative and managerial
- 3: Clerical and related
- 4: Sales
- 5: Service
- 6: Agriculture, fishery
- 7/8/9: Production (industry), transport and labourers

There further are 76 minor groups, 296 unit groups and 1675 micro groups. Also, status, product and relations is included.

HISCLASS

Developed for the period before the 20th century. 5 different criteria to go from HISCOM to HISCLASS: By passing through a tree of categories, through manual/non-manual → skill → supervision → sector → class labels → number. By this system, one gets 12 distinct classes of workers.

15 June 2016: Peter Teibenbacher

Transition in migration in a long term perspective

Migration, spacial movement, may be measured by subtracting the emigration rate from the immigration rate, usually measured in per mille.

Issues

Assimilation–acculturation–integration–inclusion, segregation and discrimination, macro–micro.

Subdivision

- permanent vs. temporary: example emigrants vs. commuters
- single vs. group migration: most go in groups, but individual migrators occus
- circular vs. linear: commuters vs. emigrants
- forced vs. voluntary: refugees vs. labour migrants
- types of migration:
  - ritual,
  - labour,
  - environmental migration: problematic issue, as this migration could be caused by other factors. Modern families might not be satisfied with that which ancestors considered adequate.

Research questions

How would you count them and who is a migrant? In the 18th and 19th century, people were separated in present and non-present or native and stranger – the criteria being to be present and to be born locally. Real emigrants, permanently leaving states, were registered in separate lists.

If if one knows them, one doesn't know the reason for the migration. Migration is rarely done in one step, and one rarely knows the previous situations. To cover this, aggregated statistics through qualitative surveys is needed. Where did you go? Why did you go? Whom did you meet? In addition to these sources – which tell us something about migration directly – there are sources telling about migration indirectly, such as censuses, marriage registration forms, applications for citizenship; these would tell where the person was born as well as where persons of affiliation was born.

Defining a migrant

To answer the preliminary questions: Anyone moving is a migrant. In official statistics migrants in general and immigrants are often mixed up; they cover all migration, e.g. as in the UN Worldbank, the CIA Factbook. An example was the negative net migration rate of Ireland in 2007 and 2013–2014; many of those leaving were not originally Irish, but people who had been labour immigrants or the likes; and of the native Irish, we don't know whether they left the country permanently. In other words: A table titled net migration rate is not sufficient, and it usually doesn't "distinguish between lawful migrants and undocumented migrants" (CIA FB). The BBC on the other hand, define "migrant" as someone not yet granted asylum or other kinds of permission to stay in the country. UNESCO differentiates between immigrants and migration

within the boundary of the country or area in question, but nothing more. The UN defines international migration as those who on mid-year (1 July) lived in a country or are in which they were not born. Worldbank merely calculates the difference between immigrants and emigrants, including both citizens and non-citizens.

The human factor is often neglected. The formal definition is delimited by the interests of the governing body, and therefore serves this unit's purposes.

The first generation begins the change, the second generation is in the middle ground, and the third generation is fully adapted to the new culture.

Zelinsky's model of migration transition (1997)

Based on the main thesis "Migration (variants and frequency) is dependent on (economic) development.

He begun by defining the history in five different periods (i pre-modern traditional, ii early transitional, iii late transitional, iv advanced, and v superadvanced society). He further subdivided migration in four different types: emigration, rural–urban migration, urban–urban migration, and any other circular migration.

His findings

There is almost no migration in the first period. During the second phase one sees a strong increase in rural–urban migration as well as emigration in general; urban–urban emigration had a slightly weak increase. In the third phase urban–urban migration is still increasing, any other circular migration is continually increasing, whereas emigration decreases; rural–urban migration peaks during this periods end and emigration peaks at the beginning. In the advanced society, rural–urban migration begins to decline and emigration continues its decline; circular migration increases sharply, to a great extent due to better "tools" for transportation; and urban–urban migration has a slow, yet steady increase. In the superadvanced society, the rural–urban migration practically comes to an end.

It is worth noting that the analysis does not include exceptional migration, such as due to war (forced). Developed countries show a stronger tendency for circular migration. There are no other recent models.

Other kinds of migration

There has always been waves of large scale migration: immigration waves to the US during the sixteenth to the twenty-first century; expelling of the Jews, the Barbarian invasions. Also, in previous times, there has always been a lot of people on the road: marriage migration, commuters, beggars and so forth. The problem of (voluntary) migration started only during the seventeenth century or so.

Influencing factors for voluntary migration

Estimated situation at place of origin, calculating the way, expected situation at end point. A push-and-pull model could explain these things: strong disadvantages at origin are expected to be lost; the risks of leaving are evaluated; and strong advantages are expected gained at targeted place. Summed: The higher surplus and the lower cost, the higher probability of migrating.

The negative factors aren't necessarily material. An example could be a person with a good job living in an area with a bad school; this person could decide to move to a worse job for the benefit of his children. Elderlies, particularly in rural areas, are more stable; for them the migration calculation yields a poor result, as they have roots in the area and a short expected lifespan in the new area.

Laws of migration, macro and micro models

Macro models favour general factors like labour and general "laws" of migration, similar to natural science's laws, almost claiming causality.

Micro models ...??

Ravenstein

Ravenstein created a model for explaining migration, his "Ten Commandments", i.e. his Law of Migration (1880s):

1. Most migrants travel short distances.
2. Migration proceeds step by step.
   This, however, is not so much true any more: Africa, for instance, sees *lots* of migration, though these is often not appreciated in Europe. In older times, one would move to the first big place looking for a job; then, when their resources were adequate, they'd move to the next place.
3. Longer distance migrants prefer to go to great centres of commerce or industry.
   Today migrants often choose to go to the rural areas outside of cities, as the living there is cheaper, there is more farmland or countryside to appreciate. For shops, schools, jobs, however, they go to the urban areas.
4. Each stream of migration produces a counter-stream.
   This is still true today. By leaving, they are changing the factors at their place of origin.
5. Urban dwellers are less migratory than people in rural areas.
   In these days, however, commuters counter this.
6. Females are more migratory than males in internal migration, but males are more common in international migration.
   Whether this is true today, is hard to say.
7. Large towns owe more of their growth to migration than to natural increase.
   It changed in the last two decades before WW1. Migration reduced somehow, maybe the market was filled? Economic growth stagnated.
8. The volume of migration increases with the development of industry and commerce and as transport improves.
   This, of course, is still true.
9. Most migration is from the agricultural areas to centres of commerce and industry.
   Now, though, most industrial countries don't really have agricultural areas anymore.
10. The main causes of migration are economic.
    Defining economy as a case of scarcity allows everything to be considered economy.

Neoclassical theory

"The neoclassical theory understands migration to be driven by differences in returns to labour across markets. ... According to this theory, migration is driven by geographic differences in labour supply and demand and the ...?"

NEM: The New Economics of Migration

"... shifted the focus of migration research from individual independence to mutual interdependence (Stark 1991). The key argument is that migration decisions are not made by isolated individual actors but typically by families or ...?"

World System Theory

This links determinants of migration to structural changes in world markets and views migration as a function of globalisation, the increased interdependence of economies and the emergence of new forms of production.

Migration as a network

Instead of looking at the determinants, it examines what perpetuates migration in time and space (Massey et al. 1993). "My brother is already there, so I will follow him."

Migration systems theory

Migration alters the social, cultural, economic and institutional conditions both at origin and destination, and therefore it forms an entire developmental space within which migration processes operate (de Haas 2009b). While migration systems theory has its roots in geography, migration network theory is of sociological and anthropological origin (Castles and Miller 2009). Whereas network theory mainly focuses on the vital role of personal relations between migrants and non-migrants, migration system theory goes further and stresses that migration restructures the entire societal – or "developmental" – context of the concrete spaces in which it takes place, both at the receiving and at the sending end (de Haas 2008).

Other models

The life-cycle migration model: The migrator is young, doesn't lose much, and therefore moves. This excludes children and the elderly.

Intervening opportunities model: Moving as (in particular) job opportunities show up, or other large improvement opportunities in life.

Gravity model: The shorter the distance and the larger the mass, the stronger the gravitational pull.

Rates of migration

Europe–Russia: Northern regions have had a positive net migration since the 1960s, and it has increased. The east and south have lost people the past years. The west has had continuously increasing net migration since the 1950s, excluding the decade from the late 70s to the early 80s.

After 2008 things changed in the EU/EEC area. Ireland, for instance, changed from positive to negative rate. In Austria people move both east and west.

Europe – both historically and presently – has mostly internal migration. From the sixteenth century, there was a large increase of emigration. Catastrophes, such as the Irish potato famine (1845–1852) caused a hundred-year population decrease. The Jews have migrated within Europe for centuries

The Americas: internal migration and lots of immigration.

Africa: internal migration, immigration of Arabs since seventh century, immigration of Europeans since the nineteenth century; emigration of Africans since the 1980s. The export of African slaves, of course, account for around 12 million in total to America, and also centuries of export to Asia.

Asia: mostly internal migration, emigration since the late nineteenth century. Both Indians and Chinese have migrated to the whole world; the Japanese have mainly populated the Americas.

Australia: ??

Surprisingly the south–south migration (migration between developing countries) is higher than the south–north migration (from developing countries to developed countries).

15 June 2016: Richard Zijdeman

Introduction to R

Quantitative research method

The method is chose to answer descriptive and explanatory questions on populations, as opposed to discussing single persons. Inferences need to be drawn by generalising. Sample hypotheses could be e.g. "Why do children get different jobs than their parents?" or "Why do women get different jobs than men?" This is done by using the PTE scheme (problem, theory, explanation).

Example case: Migrational problem

How many, to what extent? → Explanatory question, such as finding the mechanism behind it.

We know that the work of men is valued higher than the work of women, i.e. higher salary. The difference between the wages could be analysed by finding the mechanism.

The method

To do the work, one usually has to choose a sample. For the analysis workflow:

import → ↱ review → descriptives
regression* ↲
↳communicate

The tools

Using spreadsheets is a dangerous process, particularly due to two problems:

1. tempting to input and clean data
2. ???
3. defaults messing up entered date: 01200 → 1200

Using syntax (scripting)

Syntax/scripting offers efficiency, quality (error checking and reviewability), replicatability, communication (many journals now require both your data *and* your code)

* explanatory analysis

15 June 2016: Siegfried Gruber

Hands-on computer workshop

- censusmosaic.org
  - Requires login.
- nappdata.org/napp:
  - Download section allows you to log which downloads you have made and comment them, so you'll know what you've done.
  - Also supports online data analysis.
  - Requires login.

Household classification systems

How to measure and classify households

- Size is simply done by numbers. For very large sizes, ranges could be used.
- Generational classification: 1–4 generations, i.e. only couple; couple and children; grandparents, couple and children; or grandparents down. Also a dummy-class can be used for the unknown.
- Structure:
  - 1-person households
  - Family households
    - Special case: Courts should be doubly classified.
  - Institution/group quarters
    - Prisons, hospitals etc.
    - Monastery
    - Courts (v. s.)
  - Different structualisation:
    - Households with only relatives
    - Households without relatives
    - ...??
    - Le Play 1855, 1871:
      - *famille instable* (nuclear family)
      - *souche* (stem family)
      - *partiarcal* (join family)

- Compound family units (Hemmel–Laslett system):
    1. solitaries (0 CFU)
        a. widowed with children
        b. single/unknown
    2. no family (0 CFU)
        a. co-resident siblings, no other relatives
        b. co-resident relations of other kinds
        c. persons not evidently related
    3. simple family households (1 CFU)
        a. married couple, no children
        b. married couple, children
        c. widower, with children
        d. widow, with children
    4. extended family households (1 CFU)
        a. extended UPWARDS
        b. extended DOWNWARDS
        c. extended LATERALLY
        d. combinations of 4a–4c
    5. multiple family households (≥2 CFUS)
        a. secondary units UP
        b. secondary units DOWN
        c. secondary units LATERAL
        d. *frérèches*
        e. other multiple family households
    6. Incomplete classifiable households

Note: for this kind of terminology, lodgers, servants, serfs etc. do not count. Also, extensions are always done from the CFU; if the CFU has a household head, it is the unit of account.

Review: Each system stresses some aspect. If they are based on structure, they are more complex. Statistical agencies use their own system, making comparison more difficult. In historical demography Hammel–Laslett system is the most common.

Mapping the family units, may be done with the standard shapes: △ for men, ○ for women, ◇ for unspecified. The head of the household is blacked; marriage is shown by a line connecting the couple below; children are shown by connecting lines above. A dead person is stricken out by a slash. Circles around the units can be used to show how families are connected, e.g. to show that a polygamous household is considered two units, both headed by the husband. Example:

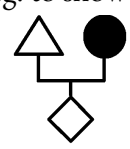


Illustration 3: Shows a household headed by the wife, the parents having one child of unknown sex.

16 June 2016: Tim Riswick

Fertility

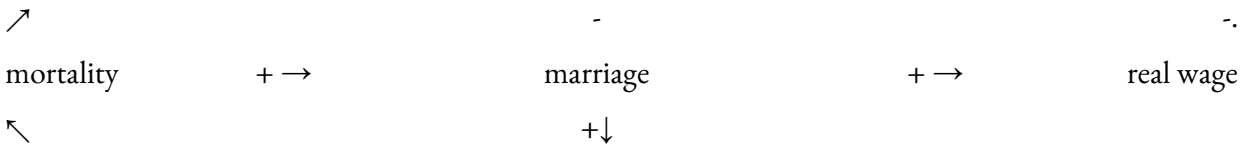
Nijmegen's project *The Power of the Family. Family Influences on Long-Term Fertility Decline in Europe, 1850–2010* (2011–2016).

Fertility from an individual level

It was assumed that in a normal life course, you'd get children; in particular for women, children was the predominant factor connecting the stages of life.

Thomas Malthus

Saw fertility from the bigger perspective. Paraphrased: "The world can be divided in two parts: In Western Europe we manage population by marrying late, which limits the population; in the rest of the world, people just die."



Fertility today has fallen to such a low degree, that some have argued humanity will die as a result; others argue that we are facing a "population bomb".

Terminology and theories

- Fertility: realised live births
- Fecundity: physiological ability to bear children
- Fecundability: probability of conceiving per menstrual cycle
- Natural fertility: realised live births without contraception. Naturally, women would give birth to 11–16 children.

Theories of fertility decline

In the first phase mortality is really high. When it begins declining, population growth commences. A while after the death rate has dropped to a minimum, birth rate declines, leading in turn to a stagnation in population growth (Frank W. ???)

A good way of getting to know the theories and the status of the data, is by studying www.gapminder.org.

- Demographic transition and modernisation
- Princeton European Fertility Project:
  - Can we explain it by looking at the smaller units?
  - Decline is not due to adaptation to new economic circumstances; rather, it is due to diffusion of behaviour. New ideas? New dynamics?
  - Coale: "ready, willing, able"

    - Responses:
        - Charles Tilly: Criticised them for not taking history seriously.
          They wanted to link it to the economic theory, which didn't really explain anything.
        - Robert Woods: Cannot connect it to one thing.
        - Ray Hall: Social-economic variables – things that can be measured – can't explain it all; they ignore culture and changed ideas.
- Easterlin–Crimmins model: cost-benefit; supply, demand and cost of control
    - Differentiates economy from other factors, such as social factors ("I do not want to marry early.")
- David Reher: infant mortality (supply)
    - When people experience that children don't die so often, they are conscious of this, and therefore plan thereby.
- John Caldwell: intergenerational transfers (demand) – wealth flow theory
    - He observed that in Africa, parents with children going to school, had fewer children: children getting education costs money for the parents; children working provided for themselves.
    - He didn't provide data for this; rather he built his theory on his observations.
    - Wealth transfer:
        - Wealth transfer for parents with many children: parents ← children
        - Wealth transfer for parents with few children: parents → children
- Simon Szreter: social context
    - Religion, information, in other words whom one spoke with mattered more. Did people *believe* having children would cost more money.
        - This was observed in Netherlands: Areas speaking the same dialect experienced the same changes in fertility.

Fertility decision making: female agency

When having their own income, they have a stronger agency – more ability to make decisions on their own behalf. The focus in other words shifted from "couples" as a unit to women and their bargaining power, and extended kin. Another factor in this power struggle, would be the age difference between the man and woman.

Unified growth theory

Couples fertility to industrialisation: income–fertility connection is key. Becker argues that there is a quality–quantity trade-off.

Arguments against UGT

They do not realise the importance of infant mortality. Human capital increase does not explain the first stage of the transition: is there a source for selection, or should we allow more time?

Etc.

State control and influence

One-child policy of China. Propaganda during war time to "produce soldiers". However, during peace times, this isn't necessarily true; these campaigns could just as well be to increase the power of various social classes or racial groups.

Not many theories uses this factor, but it is important to be aware of.

Factors influencing fertility decline

Economic development, demography (infant mortality), culture, position of women, technology, government. Interestingly, the class chose economic development as the most important factor.

When asked about the question of growth in the child population, the class believed a growth from 2 bn. children to 3 bn. children during the next century. Surprisingly: high mortality rate seems to be the most important reason people rear many children. The most important factors can be summed as such:

1. Increased age at first marriage,
2. increased educational level,
3. increased integration of women into the labour force, and
4. increased access to family planning.

Religion has very little to do with the number of children per woman! (Hans Rosling's Ted Talk). Ansley Coale agrees with his view: 2 bn. children per woman is the limit, which will lead to a peak population of 10 bn., provided the poor countries get access to the same four criteria.

Measuring fertility

Several factors could be determinants of fertility. Indirect factors need to be accounted for. For example, in a society in which you have to be married to legally have children, if only 20 % of women get married, this would strongly impact how many children were born. Also there are biological factors, such as breast-feeding or induced abortion.

Comparing fertility: Are you going to look at period fertility (PF) or cohort fertility (CF)?

- Period fertility looks at a short period of time, whilst in cohort fertility, you'd study a group over a long period of time.
- In PF studies, data is easily accessed. In CF on the other hand, you would need longitudinal data.

For more, see the table in the pdf.

Measures:

- $\text{crude birth rate} = \frac{\text{births in year "x"}}{\text{midyear population in year "x"}}$
- Other measures include general fertility rate, age-specific fertility rate. With the latter, good-quality data is required; for comparison, this is the best method; also, it is usually expressed in number of children per mille.

- Reproductive histories: Cohort measure:
  Starting, spacing and stopping: First birth, interval between, last birth. Comparing these numbers gives a good indication of how fertility changes over time and can give an idea of which factors contribute.
- McDonald's formula: Requires very good data. For this reason, it is difficult to use it for older societies or societies in which data collection isn't on par. It specifically addresses legitimacy.

Eurasian comparisons

Demographic comparisons to hard times played a key role in balancing population and resources (according to Malthus). The Malthusian model stresses the differences past–present, West–the rest, collective–individual, and passivity–agency; the latter.

Today, one can see a connection between high rate of female infanticide, male late marriage, [more].

Debate Wolf–Lee. Wolf argues that "malnutrition is causing lower fertility"; Lee argues that "people are deliberately lowering fertility". Wolf uses interviews and Taiwanese household registers as main sources; Lee uses local population records in China as main source to generalise. Wolf links poverty and nutritional status; Lee studies active decision making, i.e. infanticide. The problem of the debate – even in the journals and books in which they write – they do not come to a joint conclusion (i.e. they completely disagree).

Study: Fertility four ways

Urban–rural studies, Netherlands compared to Taiwan. In the Netherlands the age-specific general fertility was almost similar. In Taiwan, however, the general fertility was much higher in rural areas. Both followed the same pattern of getting children, though. The same strong difference (rural–urban) can be seen in marriage age; in the Netherlands, the difference isn't so pronounced.

In Taiwan rural areas, women were more reliant on their kin, whereas in urban areas they had a higher degree of self-empowering. The marriage age can be explained by the different in the European and Asian marriage pattern.

The numbers are hard to compare with their respective continents for Taiwan for several reasons. In the Netherlands, though, it is (most likely) comparable to Europe in general. In Taiwan, having few brothers yields a high chance of not marrying early when living in urban areas.

Ethnographies of reproduction

Pronatalist ideology. How does the transition compare between southern Netherlands and Taiwan?

Several images were produced in China showing nineteenth century couples killing infants; in the Netherlands this was illegal. Other contributing factors, naturally, were a abortion and contraception, and courtship; with regards to courtship, unlike in China, Dutch couples to some extent could choose each other. In the Netherlands the catholic church clearly instructed women that their job was to rear children. In Taiwan you could hear questions like "Why would I get married if I did not want to have children?" and stories of women being forced by their family-in-law to get more children if they only had daughters. In Taiwan it was on the other hand of interest for women to have their own business, and this appears to have been what society expected; did this mean they were more empowered than their European sisters?

Some recent examples of research on fertility

Short economic stress had an impact on reproduction, though not in Japan. People were able to make active decisions, though not everywhere. Fertility is clearly influenced by physiology, psychology and behaviour, but cultural influence needs to be examined more closely than has been the case.

See the papers provided for more information.

A contemporary example: Japan

Japan and Taiwan have the lowest birth rate in the world at this moment. BBC documentary *No Sex Please, We're Japanese*. Population is expected to fall by a third. In a school district (Ubari), 21 primary schools has been reduced to 1. In the 50s, most families had eight children (according to one of the interviewees. Despite great efforts, young people simply wouldn't move back to the towns, exemplified by the maternity unit in the Ubari hospital having been completely shut down; the doctor interviewed told that the amount of children born per year in the city now was zero. Most of the world's countries that discuss population, discuss population increase; in Japan, the problem is catastrophic population decline Even in Tokyo they are noticing a change: fewer children born and higher age of mothers; being 31 years old with first child might now be young.

The decline began during the economic boom in 1974, and the speed of decline has increased. To maintain population, every woman must give birth to two children (2.07). The rate now is at 1.4. The grandmothers (that generation) gave birth to fewer babies; the current mothers are giving birth to even fewer, and it is expected that the children-generation of present time will have even fewer. With a high debt and fewer workers, even though it is the third largest economy it will not make enough money to pay off its debt.

Far less people get married in Japan than in other countries, and that is a problem, considering how traditional the culture still is. Of those married, only 27 % reported having sex weekly. Some men now rather engage in virtual friendships, as in not with an avatar, but with an actual electronic person (R PG: Otaku-culture, *Love Plus*). Although they aren't that many, they are a part of the problem. Why don't Japanese men get involved with women? Why would they challenge themselves with all the problems of a relationship, when they could instead involve themselves in a virtual world – e.g. through manga – that might be seen as superior? It is an idea of purity, of wholesomeness. Most male students are very pessimistic about the future, and are sure they will not do better than they're parents. Relatively speaking Japanese women find the men wanton. There is a generation of women expecting their own independence, but also for the *men* to be independent; the parent generation who went to university were guaranteed a job, and it was life-long, so they didn't have to be particularly autonomous.

The Japanese were completely humiliated by the war, loosing their idea of their great history as well as an immediate secularisation due to the God Emperor declaring their capitulation by radio. The generations that rebuilt the new Japan, creating a new national identity by their own diligence. The current generation of men seems to have an idea of themselves being unable to match their fathers' and grandfathers' efforts, and therefore recede into a state of non-action; the women on the other hand, have a different cultural perspective, inheriting their mothers' and grandmothers' point of view, having had hard-working husbands who at the same time were more like robots, something undesirable by them. The women therefore have an ideal of autonomous men, whereas the men compare themselves to their forefathers, seeing themselves as incapable of measuring up to them.

16 June 2016: Richard Zijdeman with Mihaela Haragus and Oana Sorescu-Iudean

R computer workshop

Empirical testing workflow:

Import your dataset → Tidy it →
[Datawrangling: Transform → Visualisation → Model] →
Communicate

Resources

- Stackoverflow (from Stackexchange): Reputation based Q&A.
- Another great source is *R for Data Science*: http://r4ds.had.co.nz/.
- On GGPlot2: docs.ggplot2.org. Study the vignettes for more information.
- Dplyr (d[ata]plyer): Package for picking data apart.

Etiquette for asking questions online

- be polite
- be concise
- short background
- replicable example
- debrief your efforts so far.

Regression analysis

Easily explained, harder to master.

Subject not covered yesterday

Environment field shows what is loaded into memory. It contains script and more.

On graphs

Graph theory is the science of how to communicate graphs properly.

17 June 2016: Christa Matthys

Life course analysis – introduction

Principles

What defines a life course? What defines a demographic life course? The idea that a life course is a succession of stages in your life, and all these stages are marked by a beginning and an end event. A life course in itself (not demographic life course) would include more events than the ones studied through registered events. One analyses what might be the common denominators, and then tries to identify the odd ones out. By using the information present, one attempts to explain human behaviour.

Literature

Methods of Life Course Research (J. Z. Glete & Glen H. Elder Jr.) and "Principles and prospects of the life course paradigm" (Jan Kok, *Annales de démographie historique*, 2007, 113, pp. 203–230)

What constitutes life course analysis?

Life course is not a methodology; it is a theoretical framework. This matters! The events mark the transition from one stage to another, e.g.: first job (from unemployed to employed), new job (from secretary to boss), marriage (from unmarried to married), death of spouse (from married to widowed). Another example: If doing a life course study on students, an event could be the first time one raised his/her hand to speak.

Events in one persons life could effect the life course of someone else, such as dying and widowing a spouse. The data are individually dynamic (though they could be static), and is connected by family and history.

Different trajectories of the LIFE COURSE ← the intersection of age, period and effect TIMING ←

1. history and culture LOCATION IN PLACE AND TIME,
2. development of the individual HUMAN AGENCY,
3. and social relations LINKED LIVES

Examples:

1) Romania eighteenth century? Moldova twentieth century?
2) Not only looking at the characteristics of a person at the moment it happens (e.g. having a child now, being married now, having an occupation now), but as to how this relates to previous events (already having had children, previous marriage
3) What happens when a spouse dies? Which effect does the choices of one's family members have on an individual, e.g. does having siblings that have urbanised themselves make it more likely for the last one to do the same?

An example: observation on survival of 50 individuals over time, 1975–77 to 2006

Getting censured data into the analysis could be problematic, for example losing someone from observation; this is called a censored data (unit) – in other words this is someone for whom one cannot know the result.

If the study was interested in say *birth*, one would – unlike in a mortality study – get many events of interest; these events are called multiple failures of object.

What are the benefits of studying life courses?

The historian is interested in studying temporal change. By using the theoretical framework of life course analysis, which prefers dynamic data, the historian has access to a large set of skills with which to study research object. In other words, the life course approach is more suitable to the historians field of study. Qualitative research is of course at the heart of the historian's field, but by using LCA and EHA (*q. v. i.*), the historian can get a better understanding of what actually are relevant factors.

Practice

Some terminology:

- We are looking at patterns in human behaviour. Therefore we study groups of people, called COHORTS. The most common cohorts in historical demography are birth cohorts and marriage cohorts. If for instance one was interested in the difference in divorce, one could study particular eras, years, or even days.
- How different attributes – COVARIATES – such as sex and age, or income, class or education, impact the societal, familial and individual level on the occurrence and TIMING of events is analysed. Why do couples stop having children, for instance? Is it dictated by generational gaps? Is it the amount of children they already have? The covariates are on three levels: societal (born in Slovakia or France), familial (How many siblings? Brothers? Sisters? Which number?) and individual (such as name).
- How far have you come in your life by now? How long does it take for this event to happen? Mainly quantitative life course research has been done until now. This is called SURVIVAL ANALYSIS, which is the same as EVENT HISTORY ANALYSIS (EHA).

Event history analysis

The biggest challenge is to get data of high enough quality. Data requirements include well-defined time frames and a reliable sample of population. The events must be accurately recorded (minimum date and type). Linking records must be possible, and sufficient constants and time-varying covariates must be present.

- Multiple beginning dates might be used (such as birth cohorts), or one could choose a set date in which the study begins. A specific end time would be necessary as well.
- One way of taking a reliable sample, would be to find which combinations of the population yields the most reliable population sample.
- The linkage of sources is often of importance, e.g. when working with parish records. In almost all the cases when doing life course analysis, one links multiple sources to study the objects. To do that, one naturally needs to be certain that it is the same person in each record.
- At every time during the study, the ideal is to be able to say who is in the study and not; good criteria are key to accomplishing this.

Population registers

They are generally the best source known for event history analysis. Netherlands, France, Italy and Sweden are of the few that have something like this; also, there is something similar in Japan. They are limited in time and space, have a well-defined population and time (though there is some under-registration due to emigration), and they define accurately life events and covariates and makes linking possible.

Family reconstitution

Databases based on FR is the next based things. These are usually used to calculate "classic measures", as these have few covariates. They represent a sample population (only married couples), and this makes it harder to track as they are geographically limited; only the non-migrating may be studied through these (unless being able to link different records). Accuracy varies, and one cannot deal with missing data.

Other sources

These may be troublesome: genealogies, frequent censuses, combination of various sources.

Population registers by example of the Dutch

Beginning in 1846, though some sources are from earlier. Organised by legal, secular community. For each village or town, there was a set book for keeping the population register alphabetised. Each double page would give the information for one household. On the top of the pages were the addresses. The updates were done by way of the decennial censuses.

Their uniqueness lies in them being updated continuously after the censuses. Marital status, and similar, were of course added, but more importantly, birth and immigration to a set household was also included. This means that persons that were only present *in between* two censuses were recorded as well, such as infants born or died, or temporary immigrants. Another important piece of information in these records, is information on whom members of the households married. Instruction on people emigrating was also maintained, and with information of which page they were being updated in (if within the same village); since both location to where they moved and date was recorded, it would be easy to consult the new village's records in order to create an entire life course for that person.

Seasonal work-migrants: Problematic, as there is great variance from village to village in how well these were recorded.

Case: Louis Mestdagh between 1847 and 1866

Problem 1: The spelling of his name. During the time, French was the *lingua franca* in Belgium. As such, the Dutch name Lodewijk might be spelled either Louis (French) or Ludovicus (Latin).

Linking the three moments in his life – sex, birth date, birth place – that are common denominators, one can easily trace him in the sources.

Problem 2: When no date is given for an event, how is it registered? Use the date of the recording. In his case, he is recorded as a beginning teacher 1 January 1847, and changed occupation 1 January 1857. A beginning and ending date *must* be given for each event, otherwise the analysis would simply not work.

End events: Select the ones important for the study. The central event for a study is called the FAILURE EVENT: failures are "when the event of interest occurs". Other events could also be included, such as getting married, getting children, dying, and so on; all of these could be failure events.

Available datasets

An overview is provided on http://www.ehps-net.eu/databases

Interpreting event history models

A COX MODEL could be shown either with coefficients or with exponentiated coefficients ($\exp(b)$ or e^{coeff}) or hazard ratios. Coefficients that are higher indicate higher odds to experience the event. The $\exp(b)$ variant and hazard ratios are interpreted similarly: A value of < 1 indicates decreased probability of experiencing the event.

You always have reference category for categorical variable. In a publication you need to find the following:

- The descriptive statistics, such as which groups are included, how many censured cases there are (people for whom we do not know the failure of.
- Number of observations – i.e. episodes, not persons – objects and events (i.e. events of interest). In other words: observations includes everything; events include only the observations that are of specific interest for the study.
- Quality check of a model: How the assured the quality of the data.
- Statistical significance for each model.

Case: Jan van Bavel: "Family control, bridal pregnancy ..." (see slide)

Studies the chance for women to give birth outside of marriage.

But what if you don't have "perfect" data for EHA?

A life course framework might be useless in this case, but even though the data might be ...

- scattered or fragmentary,
- inaccurate, thus complicating record linkage, or
- the analysis of timing which is necessary to do EHA, ...

... the perspective might still yield interesting results. The impact of earlier life events is crucial to understanding human behaviour. The fact that you can study deviant groups, those not following the conventional pattern, allows us to get a more detailed understanding of varying the human experience is.

The method

When it is hard to know the entire population, select a well-defined and homogeneous sub-population. An example could be never-married women, deaf people or prostitutes. Avoid limiting oneself to just the classical sources. In the first case (never-married women), it is a fairly classical study, using census data combined with parish registers. In the latter cases, cohorts were selected from military and school registers (deaf people) or legal sources (prostitutes); regarding the last of these, it was not possible to do this everywhere, as prostitution wasn't illegal everywhere.

A control group is necessary, which has to be as well-defined. Creating this control group, it is important not to make preliminary conclusions by way of selection.

One also should combine sources. Quantitative and qualitative sources could complement each other: With the case of the prostitutes, the court cases often contain brief descriptions of the defendant's life

course, which could be connected to the public records. For the deaf and unmarried, they combined documents using censuses, PR-S, inventories, declarations of inheritance, poor listings, lists of orphans, individual files from institutions; all of which highlighted the standard of living, social networks, etc.

More basic statistical models are useful, such as a Kaplan Meier curve. Too advanced statistics would often not be possible, due to too many censored data records. The KM-curve isn't multi-varied, and has "no explanatory power".

Finally, much is achieved by digitising and as much cooperation as possible. If your data isn't so good, working with others on other datasets could make it easier both to finish the work and get published. "Be confident!"

Assignment: Applying the life course approach to domestic servants

They are all over the world in all societies. They constitute a large number of people, are usually migrators. In general they were of working class background with their employer being of a higher class. These factors all contributed to putting them at the centre of various social crossroads.

Being a servant or not could affect one's chance of marriage and opportunities later in life. Many were able to save money, and thereby marrying up in society. With regards to migration, having siblings that had left home, made it more likely that someone moved.

During the nineteenth century, service was done more often in urban areas, and became increasingly more of a feminine occupation; for many, it provided a way of facilitating social mobility. Servants were more likely to marry up into the middle class, both by way of their savings and their well-developed manners. The people of the servant class were in fact the one's with the most bank savings. (Teresa McBride, 1970s). Higgs, however, opposes her. Due to the fact that many servants weren't registered as such when they married, a life course approach is necessary to actually understand their various roles in society.

Cross-sectional analyses are useful to highlight the role of servants in local or regional demographic systems, while life course analysis examines the effect of ...???

17 June 2016: Richard L. Zijdeman

Introduction to R, part 3

Suggested case:
How do different causes for not being able to sign correlate with the age and sex for the bride and groom?

Session management

Setting up a session

`rm(list = ls())` # give me everything in memory (ls), put that into list, then remove that.

`sessionInfo()` # check for "other attached packages"

`detach("package:[packagename]", unload=TRUE)` #removes package from memory

`setwd("[foldername]")` #sets the working directory. Note: forward slash, not backslash.

`getwd()` #to see whether you're in the right directory

`dir()` #shows what's in your `wd`.

Loading data

`read.table()` # generic function

`read.csv()` # does the same as `read.table()` but for CSV-s (comma separated v), expecting commas.

`library(foreign)` # allows reading files in foreign languages, e.g. SPSS and Stata.

`library(readxl)` # fast excel-package. Not the package with the most options, but quickest and easiest usage. Alternatives:

- xlsx(Java required)
- gdata(perl-based)
- openxlsx ()?????

Common features of CSV

`file`, `header`, `sep` (read.csv default: ","; read.csv2 default: ";"), `skip`, `nrows`, `stringsAsFactors`, `encoding` (e.g. `latin1` or `UTF-8`).

- `nrows` is useful for reading a sample number of *n* first rows, to avoid reading in large amounts of bad data.
- `stringsAsFactors` automatically converts strings to factors (such as the `am` or `vs` in the vehicle test data set.

Common features of read_excel

- `path`: your file, including directory
- `sheet`: name or number of sheet. Remember to delete unused sheets; it's a waste of space.
- `col_names`: col names in first row

Basic statistics

Box and whisker plot:

Used to show distribution of data. The horizontal bar in the middle shows the mid point, but not the mean (!). In other words, it shows the median, with 50 % above and below. The box is surrounding the midline, showing 25 % of the observation objects above and below; the box in other words shows the first and third quartile range. A narrow box indicates most of the samples are within a narrow range. Outliers are indicated by dots above and below the whiskers.

To make a box like this in R, use the following code:

```
p ← ggplot(hmar, aes(sign_groom, age_groom))
p + geom_boxplot()
```

with `hmar` being the dataset.

T-test

Three forms of a T-test (`t.test`). For our purpose, the test for independent samples is the most interesting. Before this we do an F-test, to compare two variances (`var.test`), assuming that these are not different from zero. If it is smaller than 0.05, it is significant. Using the plots we had, we can tell that grooms being able to sign themselves (have education) statistically significantly differ from those that can't.

More information

View the Read Me-file in his folder.

18 June 2016: Christa Matthys

Overview of theories of mortality

Sources and methodologies

Vital registrations are required, which may be obtained from parish registers (burial records), civil registration, (actual death certificates) etc. One also has to have an overview of the population size, as the most basic enumeration is the CRUDE DEATH RATE (given per thousand). Other elements are LIFE TABLE ELEMENTS, which gives us information on AGE SPECIFIC MORTALITY (an observed measure) as well as LIFE EXPECTANCY, which at a certain point reaches zero.

National data may be viewed at mortality.org.

What portion of the people die and when do they die? CDR tells us how likely it is that people die, whereas LTE provides a measure of when they are most likely to die.

Causes of death ≠ health determinants

The first time someone tried to make a more general classification was in the eighteenth century, and in England and Wales attempts were made to standardise this, which began spreading to other countries.

In 1893 the ICD-0 was created, an international CLASSIFICATION OF DISEASES, which is used to this day amongst others by WHO (www.who.int and www.who.int/classifications/icd/en/); we are now at 12 or 13; it is revised every other year or so.

Infant mortality

The reason to focus specifically on infant mortality, is that mortality in less developed countries (for us: historically) was extremely high. From the eighteenth century on, it began to decline in Europe, though slowly, and a proper decline could only be seen from the mid-nineteenth century onwards. Somewhere around 1825 and till the late nineteenth century, there was however, a slow *increase*, which might be explained by the increased industrialisation again causing increased urbanisation.

Key terminology:

- PERINATAL: Deaths during first week after birth + N stillbirths per 1000 live births *after 24 weeks of pregnancy*
- NEONATAL: Deaths during the first 28 days after birth per 1000 live births.
- POST-NEONATAL: Deaths after 28 days (but during the first year) after birth per 1000 live births.

Is this caused by nature or nurture?

- NATURE: Genetic factors, age and health of mother
- NURTURE: care (breastfeeding, poverty, assistance, position of women with the household, medical knowledge and knowledge about hygiene). There is a link with fertility; there is a very complicated interaction between infant mortality and ???.
- FERTILITY and its mediating role.

Death clustering

Are there HIGH RISK FAMILIES? Some families seem to have a disproportionately high child and infant mortality. Why?

In a research project on Flanders where there was almost no breastfeeding, the interval between children was close to one year, and almost all children died, with the mother usually dying after number six. The man remarried, getting another group of 5–10 children, but would usually have just 4–5 children out of twenty. 63.3 % of families had not infant death, which means that the infant deaths were clustered with just ⅓ of families.

Breastfeeding

It enormously effects survival rate. In earlier times, access to clean water was difficult to obtain, so substitutes were not a reasonable option; also, the replacements were not good for children, such as raw cow milk (the richer) or bread mixed with water (the poorer). En elite societies doctors had to tell people not to give infants liquor, coffee or chocolate. Another aspect is the numerous important nutrients provided by breast milk.

Sexuality and superstition concerning mother's milk might interfere, for example the idea that sperm could spoil the milk causing the infant to die. Another aspect was the idea of the female beauty that would have to be preserved. Wet-nursing was common, but often not a good option, as she might have to feed more than one child. For many women, labour changed the societal structure, allowing women access to paid jobs; they would mostly not be allowed to bring their infants to the factory. Social status played a role, as in most European countries women wouldn't breastfeed in public, amongst others for religious reasons (it was banned by the Catholic church).

In the south-western part of the Netherlands, infant mortality was higher. Studies have shown that religion played an important role; protestant areas had far lesser rates. In Belgium, however, the infant mortality rates were instead effected by labour conditions; there were many large farms in the north-western part; and in the areas of slightly lower (yet still high) rates, the families were poor, but they did their industry at home (spinning, weaving and the likes), allowing the to breastfeed. Another important factor was geography: In coastal areas – for instance in Belgium and the Netherlands where malaria was present (particularly bad for pregnant women) – one also notices an increased infant mortality rate. Maternal mortality is also the highest in areas of high infant mortality (confirmed for Belgium).

Familial circumstances

Resource competition: *n* children, sex parity... < socio-economic role of children.

Biological evolution theory: "grandmother hypothesis". It is very rare in nature for females (exception: elephants and killer whales) to survive their reproductive age. For the exceptions, the survival of the females allows the transmission of knowledge. See youtube.com/watch?v=6kpzZOm8gFg.

The social-economic role of the children would also be an aspect; to the provide for their family?

In a study on the presence of kin and mortality, one can see that mothers are very important, whereas fathers are not so important; paternal grandfathers even seem to have a *negative* effect on the survival of infants. Grandmothers have a positive effect, whereas for maternal grandfather seems to not have any effect. Older siblings have a strong, positive effect. See Rebecca Sear and David Coall: "How Much Does Family Matter".

Social differences

Studies before ± 2000

A lot of studies in the past, but the older studies are usually limited in temporal and local scope. They mostly focus on England and UK. The reason is that there is not a lot of material present. To study social variation in mortality, you do need the numbers for each social group, and we hardly have any information on that; this means a researcher needs to be very creative with his sources.

One person who's tried to tackle this, is Livi-Bacci who tried to connect the studies that have been done. He found that until the beginning of the nineteenth century, there was not much of a difference between classes; life expectancy was 35–40 years. More divergence didn't ensue until towards the end of the nineteenth century.

Later studies

Introducing life course analysis → special issue on long-term social differential in mortality. The role of industrialisation was examined, to see whether it improved or deteriorated the living standards, or was the no change? Are the social classes approaching with regards to mortality? The elites aren't necessarily better off. Their health situation wasn't more constant than the working class. The elite has always been slightly better off due to better access to proper food and health care → less variation over time. With the working class there has been more variation. However, there is no increase in social mortality differences. The difference existed before the decline, especially in the urban areas.

Trends

Small boom of anthropometric studies. One isn't looking at just mortality any more. Another trend is to not just look at occupation or social class, for instance by comparing siblings.

Mortality transition

We currently do not have a better framework for the following. It is being challenged, and this will be addressed later. The idea holds these concepts:

- Transition from high to low mortality
- Changes in causes of death
- Different stages → various theories
    - Western Europe's first demographic transition:
        1. Pre-decline mortality
        2. Stage 1 (starting around 1750)
        3. Pause (industrialisation)
        4. Stage 2 (starting around 1850
    - Second demographic transition:
        - Began around 1850

Economic or agricultural conditions would have strong impact (peaks). During the beginning industrialisation in Belgium, the CDR (crude death rate) would flatten. After WW2 the variations are to a strong degree

declining. Life expectancy has for countries such as Ukraine and Russia decreased from around the mid 1960s; apparently men are more prone to die of cardio-vascular diseases.

The first demographic transition

The stages represent the idea that predecline mortality was high and had strong fluctuations. The annual fluctuations were more profound, due to "mortality crises" (wars, famine) and epidemics. Today, however, the idea of predecline mortality is to a strong degree rejected, due to us not having good enough records (or no records) for the earlier times (though not entirely wrong). The case of the Domesday Book exemplifies the problem, in fact: It has been generalised to apply for all of Europe, which is problematic in many ways.

With regards to industrialisation, it is generally agreed that it *had* an impact (changes flatten), though not a consensus of why.

The second demographic transition

Great increase in life expectancy and decrease in mortality rate.

Omran's theory on epidemiological transition (1971)

Very influential book arguing that decline is explained by changing causes of death.

- Epidemics: Age of pestilence and famine, with very high mortality and fluctuating life expectancy being less than 30 years.
- Pandemics receding: Steady increase in life expectancy to about 60 years.
- Degenerative diseases: Man-made diseases become more common.

We see the development between eighteenth and nineteenth century as a decline, but we don't have a clear view of what was going on before that, so it is difficult to explain whether it was unusual.

The three main reasons for change were, by him:

1. Hygiene: Better access to water, knowledge of antiseptics. Gradual increase since the second half of the eighteenth century. Before this, you couldn't prevent this causes, for lack of knowledge. Gastro-intestinal illnesses might decrease because of these things.
    - The enlightenment in the eighteenth century was an elite phenomenon,
    - though it was spread during the nineteenth century by sanitation/hygiene campaigners
2. Nutritional conditions: Better access to food should cause improved conditions. However, Hollingsworth opposes this by the little social variations seen in early modern times, and Livi-Bacci opposes this by that most peaks in mortality are not related to grain crises and a mortality decline starts exactly at the time of nutritional crises.
3. Medical improvements: Pronounced time lag between medical advances and their effect. The second stage with the steep increase in life expectancy has all three factors present.

R workshop: Mortality

The necessary steps are five-fold:

1. Preparation
2. Import data from the hmd.

3. Plotting HMD life table measures. (From these steps, slides and videos are provided.)
4. Calculate CDR from crude HMD data.
5. Calculate life table measures from crude HMD data

The data downloaded from the HMD-database are in txt-format. Mind that these are not CSV, so make sure they are imported correctly. The data is imported with `read.data()`.

18–19 June 2016: Christa Matthys

Final assignment, Tor-Ivar Krogsæter (UiT, Tromsø, Norway): Different Perspectives on Servants' Social Mobility through Marriage

The research questions

Richard Wall examines the claim that past families were more cohesive, by examining several problems, particularly: 1) European family patterns has been so uniform that differences between countries have to a large extent been overlooked. 2) The sources discourage comparison by inconsistently defining households. He concludes that studies must investigate *la longue durée* to map actual changes, and that high-quality comparisons – within and without – rely on governments agreeing on standard definitions of households.

Bras and Neven study siblings' effect on the migrational patterns of women in industrialised areas in Pays de Herve (BG) and Zeeland (NL). In NL siblings channelled their sisters to work as maids in the cities, but in BG most women stayed at home, as households were small and domestic service uncommon. Key factors for both areas included parental death and parental or family migration. Siblings' action weren't determinative; rather they were additive, and in NL returning migrants increased the chance of women migrating.

Comparing the sources and methods

For Wall, the most important sources were statistical yearbooks coupled with the different governments' unpublished data; in other words, his data was aggregated cross-sectional censuses. This, unfortunately, caused problems, as different countries' authorities were looking for different information for different purposes. He discusses this, amongst others when considering the problem of how to define a household and its head, or when trying to create sub-populations for better analyses. To counter this problem, he suggests – as one method – to compare a country's economic level with the proportion of people living alone, thus studying whether that might be a decisive factor. One of the most important things he shows, is how public census data can be used to analyse when children left home, by finding how many children of which ages were registered under the head of the family; in pre-industrialised societies, however, the data shows that the process of leaving was very prolonged indeed, and the reasons were far more often than today not due to marriage. Repeatedly, he identifies the problem in comparison between countries, due to lack of international standards for defining the members of families, as shown e.g. in figures 10–11.

Bras and Neven, however, have access to much better data, with the population registers having been digitalised and made publicly available. They were interested in the migrational patterns of rural women, and therefore excluded those coming from towns larger than 5 000 inhabitants, unless the town had less than 20 000 inhabitants and at least 40 per cent of these were engaged in agriculture. For the Dutch, a half-percent sample of women had their life-courses reconstructed; for the Belgians, they had access to complete registers of the selected towns, out of which a sample of 5 000 women were selected. They left-censored the data by age 12, as this was the age most children finished schools; they right-censored the data by age 30, as this was the age most women would have been married. After all criteria had been adapted, they had a sample of 593 Dutch women and 3 442 Belgian women. Using a Cox-model, they could identify the key factors by birth-cohort, occupation of the family head, presence of parents, and parental and familial migration. The next part of the study, was to identify how the presence of different family members (for the women studied) at age 12 could be a determinant for her risk of migrating. The Cox exponentiated coefficient will identify this likelihood by giving a factorial number based on a reference value. Examples of the latter, are "Birth cohort: 1850–59", "Occupation: Farmer", or "Presence of parents: Both present" in table 2; as can be seen, all these numbers are exactly 1.00. For the other values, say "Mother absent, father present", one can see that there is a 3.26 times higher chance of the woman in question emigrating rural–urban. The

most important part of such analyses (*sīc*), though, is of course to identify the significance of each value. When coupled with the factorial number as shown early, this becomes a powerful explanatory tool.

Interpreting a covariate (table 2: bourgeoisie, Zeeland: 2.76@.05 significance)

The value tells us it is 2.76 times as likely for her to experience rural–urban migration than a farmer's daughter, and the significance is very good indeed, possibly due to a large sample size.