Computer analysis of life histories from Swedish church records: a case study from the demographic data base at Umea University

Sandström, Göran; Sundin, Jan

Göran Sandström Jan Sundin

Computer Analysis of Life Histories from Swedish Church Records: A Case Study from the Demographic Data Base at Umeå University

There has been a steady increase in the use of collective life histories as tools for the investigation of the conditions and behavior of man in the past. It has been observed that cross-sectional data often give too little information about social change in society, stressing structures instead of processes as determinant factors.

Sometimes these changes can be traced through the study of patterns in different age groups at a given time. In other cases we need more continuous data about generations at an individual level. Usually this can be done only for a selective group in a society. The use of census-type material combined with parochial records will result in a loss of many people due to out-migration. Some researchers have tried to circumvent the problem by anticipating that on average, the out-migrants had the same characteristics as those who remained. Others have tried to make intelligent guesses about the impact of the losses, but no method has been found to correct completely for this important drawback.

From an international point of view the work which is done at the Demographic Data Base at Umeå University may have its greatest value in creating a laboratory where these alternative methods may be developed and tested. The reason for this is the design of Swedish population records from the end of the 17th century onwards. In principle the so-called church examination registers provide a continuous record for each individual with central demographic and social information.

Sources

Parishioners were registered in Church examination registers by villages, households and families. The younger children were excluded from the earliest records, but later on in the 18th century they were also recorded. In addition to names, there is also information about the occupation of the head of the family, the dates and places of birth of all family members and the dates of marriages and deaths. There are also notes about when individuals had moved into their particular households (alone or with their families), where they came from (from another houshold in the

parish or from another parish), when they left the household and where they went. Absence or presence at examinations and at Holy Communion were noted annually. (Communion was often held only twice a year.) Each person who was examined received graded marks for reading from a book and for comprehension of the doctrines. Finally, any lapses in morality were also often noted, e. g., breaking any of the Ten Commandments and particularly bearing illegitimate children.

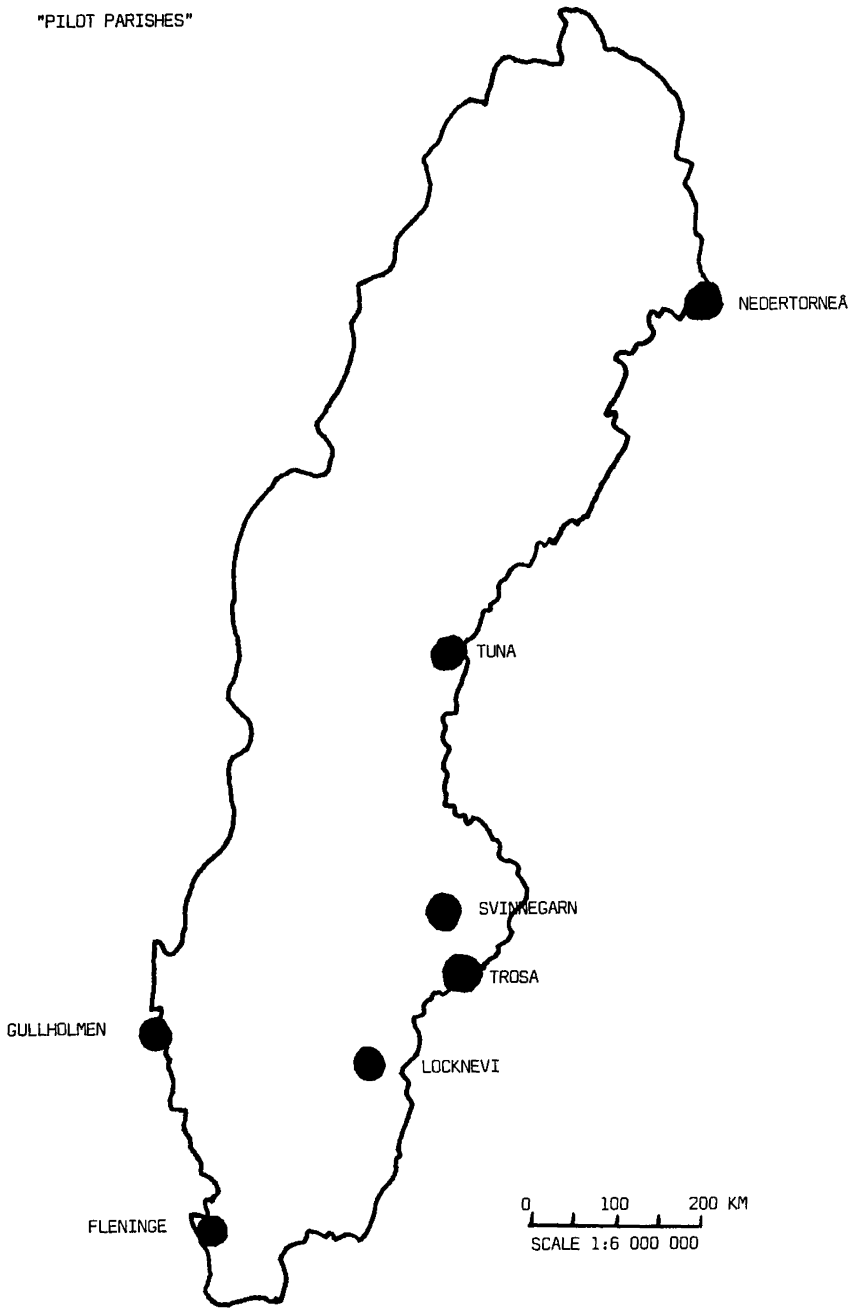
The demands from the old Swedish Central Bureau of Statistics, set up in 1749, meant that the clergy had to maintain the parish registers even more carefully. In fact, a general improvement in quality can be observed during the latter part of the 18th century. After 1800 the sources are almost always of high quality, and most of the registers have been preserved.

In addition to the examination registers, the church kept registers of births, marriages and deaths. As early as the 18th century the clergy were supposed to keep records of migrations to and from the parish. However, even if these registers ever existed on a wide scale they have mostly failed to survive. The reorganization of the Bureau of Statistics in 1860 led to the demand for more exact information about migrations which resulted in the keeping and preservation of migration registers for almost all Swedish parishes from this year onwards.

The annual registration makes it possible to trace an individual as long as he stays in a parish, and to be sure that he has been there all the time. When the source material is complete – by about 1800/1810 – it is possible to trace him from parish to parish until he dies. The possibilities for individual longitudinal studies are therefore extremely good.

Content of the Data Base

The Data Base has, for the time being, concentrated on church records from a period between 1815 and 1895. Five parishes are now entirely at the disposal of researchers at the Data Base and two more will be available in the near future. (The seven parishes are indicated on the following map.) There are, however, some drawbacks to this strategy. For example, the subsamples of one single parish may be too small for analysis and a researcher will have to follow his migrants without computer aid. Therefore the Data Base has chosen to register several parishes forming a region around the town of Sundsvall. One of those parishes, Tuna, is on the map. Since most migrants moved within a rather small area, the majority of the migrants in the central part of the Sundsvall region may be traced by the computer. Some work may be needed to combine data for the one person who lived in various parishes, but our goal is to reduce that work as much as possible. Coding of the Sundsvall region will take another three years. In order to create a general framework for the


"PILOT PARISHES"
NEDERTORNEÅ
TUNA
SVINNEGARN
TROSA
GULLHOLMEN
LOCKNEVI
FLENINGE
0 100 200 KM
SCALE 1:6 000 000

parishes and the region, the Data Base has also started a registration of aggregate statistics at the parish level for the seven parishes and at a somewhat broader level for the rest of Sweden (between diocese and parish). These statistics provide information about the social and demographic structure and events in five-year periods and about births, deaths and marriages each year. Death figures are given for each sex and age group separately.

No coding of other sources on either an individual or aggregate level has yet been made by the Data Base itself. In some cases independent research projects are collecting such material in cooperation with the Data Base. Two examples will be given: for one of the parishes (Locknevi) human geographers in Stockholm have created a system for the exact mapping of each house in the parish during the 19th century. Historians in the same city have been coding taxation poll data for each person who was not able to pay taxes during his or her lifetime. These data are linked on the individual level with data at the Data Base about the population of the parish of Fleninge.

As mentioned above three more years of work must be completed before the first region is fully coded. During that time different strategies must be prepared for the future. For example:

1. The Sundsvall region could be expanded further. This would make it even easier to observe the individuals as part of a larger system and at the same time make it easier to follow migrants.
2. A new region could be chosen from another part of Sweden to be coded in the same way as the first one. This would increase the possibilities for making comparisons.
3. The coding of church records could be stopped and parallel sources on the individual level could be integrated into the parishes and region already coded.
4. Migrants could be followed into their new parishes in order to create cohorts of people who had ever been within the borders of the parishes in the Data Base system.

No final choice has been made from among these priorities. There are researchers in favour of or in opposition to each of the strategies. One probable solution is that a combination of some of these approaches will be pursued.

System of Registration and Manual Linking

All coding is done at the „production unit" in Haparanda on preprinted cards. Each time a person appears in a source one card will be written out with the relevant data about him. 10 % of the cards are checked for errors. If the number of errors is above certain limits of acceptability the problematic variables will be corrected on all the cards from which the sample was drawn. The critical values, of course, vary depending on the character of the variable. Variables which are essential for linking are given very low critical values for error checking.

When one volume of the examination record is registered in this way the cards are sorted by hand according to given data of birth and for each person chronologically. Migration within the parish during the actual period of 5 to 10 years is represented by one card for each individual's place of residence. This operation is compared with the notes about destination of migration given on each page on which the individual has appeared. A second check is made by using information about attendance at examination and Holy Communion, so that the chronological order of the person's residence is ascertained.

A similar sorting takes place when cards from different volumes are put together, forming a chronological listing of the person's residence while in the parish. Finally, the card about birth, marriage, death and in- or out-migration are placed in chronological order within the set of cards for each person. It should be noted that normally the names are only used as a check on the other linkage variables or in the few cases when there are logical contradictions between the other variables. Cases of making one person out of two are extremely rare. Theoretically one can make two persons out of one, but since there are so many linkage variables such cases will appear very rarely. Finally, a „guide-card" is put in front of the card file for every individual, containing a „unique identity" of the individual, his parents and husband/wife. The unique identities are chosen from the birth register and, if this information is not available, from the first examination register. This identity is duplicated by the tape-puncher on each card belonging to the same individual.

No corrections are made of any source during this process. The manual linking gives a „proposed" life line. Any researcher can carry out the linkage in other ways, by hand or by machine. The manual linking can be checked by the computer according to the needs of the investigator. The work on creating this system of checks has started at the research unit in Umeå. Since the work is not complete, documentation of this process has not yet published.

When the cards are sorted by hand the result is entered on tapes checked by programs in order to ensure the best possible logic, especially concerning linkage variables. Afterwards the tapes are sent to the Computer Center at Umeå University, where they are stored. The computer breaks down the material into source-oriented files. From these files one will be able to reproduce the source as it is in its primary form. Some cross-sectional studies can also be made.

The second step is to create new „individually-oriented files of events" with the help of the unique identities. These files do not contain all the data within the source-files. Priorities have been made concerning selection of information which is repeated in many sources. Dates of births are, for instance, taken first from the birth register, dates of marriages from the marriage register, etc. The rules of priority may vary slightly from one period or from one parish to another, depending on the quality and accuracy of information given in different volumes. Since the priorities are made by computer they can always be documented. Should a researcher choose another system of priorities, he can do so if he is willing to pay the extra costs for programming and testing. The result is a file which is standardized for every parish. It can be used for standard programming to create statistics and listings of life histories.

The next step is to investigate the possibilities of making family life histories. Information on the „guide card" about parents and husband/wife is checked by the computer. When the family reconstruction is accepted, new files can be created in a variety of ways according to the needs of the researcher. Genealogical searching can be done in order to describe patterns of kinship, completed family size, and age specific fertility rates can be computed, inter-generational mobility can be traced, etc.

Information Included

The following is a list of the major information items that will be found in the so-called „individually-oriented files of events".

1 Name
2 Date and place of birth
3 Parents names, dates of birth and occupations
4 Legitimacy of births
5 Migrations within the parish (years of migration and destinations)
6 Years of attendance at the examinations and Holy Communion
7 Marks for reading and comprehension
8 Occupational statuses
9 Dates of marriages and dissolution of marriage
10 Civil status
11 Dates and causes of death
12 In- or out-migrations

Of these data, those relating to occupation are the most difficult to attach to a specific year. A person would start as a „son" of somebody. Later he could acquire an occupational title in a volume, but it is sometimes hard to decide which year the priest thought that his occupational career started. Changes of occupation appearing on the same page were recorded by crossing out the old occupation and writing the new one above. The frequency of this may be hard to ascertain. Occupational changes were often recorded when the priest was writing a person into a new page. The question is, however, whether the change always took place in connection with the person's migration.

Any detailed studies of the processes of intra-generational mobility will have to take these circumstances into consideration. By using other sources, e. g. taxation poll registers, the matter can often be solved in a more precise way.

Purely demographic data are generally of a very high quality and are present for the overwhelming majority of cases. Errors can therefore be detected and corrected by combining different variables and repeated information from different sources.

Years of attendance should be regarded as very accurate. The registers were kept, after all, as a result of need for this information the quality and meaning of marks are discussed by Egil Johansson in „Educational Reports, Umeå, No. 12, 1977. The History of Literacy in Sweden in Comparison with some Other Countries".

Causes of death are of course given according to the contemporary medical knowledge of the priest. His instructions contained a list of the usual causes from which he should try to select the appropriate term for each circumstance. Generally this information is least valuable for aged persons: Their death was often classified as „because of old age".

File Organization

The input process has been so designed as to give a result which should contain the same set of information as the sources, even information which does not have its own data representation in the sources. When the data are organized in files, the first step therefore is to store data in so-called *source files*. These files are copies of the excerpted material as it was formed at the compilation process. The files are physically split up into five different files, one for each of the sources, but can logically be processed as one, as all the data concerning the same person have been stored using the unique identity as the key.

Since the records in the source files contain the data in the same way as it has been excerpted, all changes in forms and sources will change the contents of the records in these files. Of course this is an advantage in so far as one processes the data in order to give results based only on the information as it is represented in one of

An example of family reconstruction using the event file:

```
 01924 WF SONESDOTTER, PERNILLA
 01553 HB PAHLSSON, HANS
 02730 HB JONASSON, JOHANNES
 04027    , PETTER
 04172    HANSDOTTER, MARIA
 04353    HANSDR, SOPHIA
 04522    HANSDOTTER, JOHANNA
 04658    , CAROLINA
 04781    HANSDOTTER, ANNA STINA
 05238    JONASSON, EMMA KAROLINA
 05671    , HANS PETTER

KEY   YEAR 1820      1830      1840      1850      1860      1870      1880      1890
           v.........v.........v.........v.........v.........v.........v.........v
01924 1822   B--------------------M----------W--M--------------------D
01553 1817                       M----------D
02730 1833             B                   I-C-CM--------------------W------
04027 1846                       B------D
04172 1848                         B-------------------MC----------C---------
04353 1850                           B------------D
04522 1851                           B-----------------------M
04658 1853                             D
04781 1854                             B----------------------D         I---
05238 1859                                   M----------------------C----
05671 1864                                         D
```

– = living in the parish at the end of the year
O = out-migration
I = in-migration
C = circular migration within the parish
B = birth
M = marriage
W = widow/er
D = death

The tests have been performed for each of the years in the time interval 1820 to 1889 and in the same order as given above. When more than one event has occurred during the same year the result form the *last* test has been printed out.

the sources. But experience shows that the researcher is only rarely satisfied with this kind of data processing. Instead he wants all possible information about the individuals to be used. For example he wants information about migration from the register of migration if it is available. Otherwise he wants the same information from the church examination register. The researcher thus demands that the input data be read by given priority rules. These rules are however, often obvious. The registers of birth, migration, marriage and death are for instance always considered as more reliable concerning the central information they are intended to give.

In order to use all the information in the best way we have developed a system in which all information in the sources is selected according to a well-defined priority rule. Using the five source files as input the process gives as output one file called *event file* where all demographic events and genealogical relations are represented.

Each person is represented by one data record containing „the most probable description of that person". When this file is created the examination register is used as the basis for the records, which are then complemented by the other sources using the priority rule. This event file of course describes the individuals at different quality levels for different parishes depending on the completeness and the quality of the source material (and the quality of the excerption) but the structure is the same for all parishes. The implication of this is that it will be possible to make standard programs workable for any parish, which will be of great importance in saving time for clients who want information from the Data Base. We must, however, stress the fact that the event file is only an alternative to the source files, which is why any researcher may choose input data in the way most suitable for him.

The final structure of the event file is not yet decided. When this decision is made a complete documentation will be provided. The development work is aiming to have a file containing information to satisfy the majority of Data Base clients.

Possibilities and Drawbacks of the System

As seen from the examples above, the system can be used in different kinds of investigations. Data could be accumulated for each individual separately, for the family, the household, the village or the parish. The methodological problems concerning the first two categories are somewhat easier to solve than those of the third. In principle, each page in the examination register should contain one household. Sometimes, however, the priest had to use the page for two households, and in these cases there may be difficulties in separating the units. In some books the system did not work according to the normal standards. Servants could be entered in one row at the end of the volume without any indication of the household to which they belonged. Another problem involves deciding what the priest meant to be a household. We are not always sure that the members lived in the same house or had their meals at the same table. The older generation could have retired to a small cottage beside the main building on the farm and yet may sometimes be registered as belonging to the same household as the head of the farm.

A careful researcher therefore often has to consult other sources, mainly the taxation poll registers. In this source the unit comparable to a household was to begin with what was called a „smoke", indicating that they cooked their meal at the same hearth. Sometimes it could, however, be used in a more legalistic sense, meaning only that the head of the „smoke" was responsible for paying the taxes of the other members.

Because of these problems, the Data Base has chosen not to give a fixed definition of the concept „household". Every row of the page has been given a number,

so that the researcher is able to use this information in making his own rules about the composition of households. Once he has made these rules and perhaps made completions with the help of other sources this new information could be put into a new file.

Until now the resources at the disposal of the research branch of the project have been too small. Some work on the documentation of the linking process has therefore been postponed but more will be done during the coming autumn. The manual linkage will be rested by computer.

Attempts have been made to give an algorithm which tells researchers more about the procedure. This is even more important since it has been claimed that Swedish sources offer the best opportunities of making a safe linking.

Voices have also been raised saying that the linking process is either difficult and ambiguous or simple and unambiguous. If the latter is the case, the first step in the linking could be made by computer. To date, such proposals have been rejected mainly on the grounds of the work needed from programmers. A serious test will be made in order to discover the possibilities of machine-linkage. If the advantages are greater than the drawbacks, such a linking will of course be introduced.

The great amount of information for each individual in the Swedish sources is of course valuable in many aspects, though it also creates problems. The time spent on writing cards for each individual is considerable, and this reduces the possibilities of covering a long period of a large area. Every researcher must be aware of this problem if he wants to use the Data Base. In most instances the time saved by using the machine-readable data as a starting point will be considerable, even if one has to make completions by hand.

The advantages seem, on the other hand, to be obvious: populations will be almost complete. Data are at the disposal of researchers in one single machine-readable source. The standardized process affords opportunities for checking the errors carefully. Every researcher does not have to be an expert on the primary sources. Finally, the same set of data can be used by different investigators for different purposes.

Major Fields of Application

In order to discuss various possibilities offered by the Swedish church records and other parallel sources an international symposium was held in Umeå June 8th–11th 1977. Researchers representing different branches of science from Canada, United States, Great Britain, The Netherlands, Hungary and other countries met at the symposium. The need for the data was strongly supported by geneticists, human geographers, historians and statistical demographers.

One general impression which emerged from the symposium was that there is a

need for local studies on the individual level. Aggregate statistics were not considered to be sufficient for an understanding of the social processes which take place in history. One of the topics often mentioned was social change during the transition from an old rural to a modern industrialized and urbanized society. The impact of environment upon man and man's attempts to adjust his environment according to his needs are sometimes more easily discerned on a local level. This is one of the major fields of research for many human geographers and ethnologists.

Human migration, family building, occupational mobility, the development of popular education, among other things, seem to be hard to explain without individual longitudinal studies. Certainly, the reader can find a lot of other questions which are as important to answer as those mentioned, where the life histories of individuals can provide one piece of the puzzle. The Data Base at Umeå University was designed from the beginning to be a systematic project at the disposal of every researcher irrespective of his origin. Sometimes it can be used as a laboratory where methods are tested and comparisons made with findings in other investigations. Every scientist is therefore welcome to request our services. There are some minor restrictions concerning the use of data because of the Swedish law about individual privacy, but these have no important bearing on 19th century material.

The customer will pay for the time needed by the programmer to design the file according to his special needs and for the time consumed by the computer. Contact should be made as soon as possible, even if no promises can be given yet about the time of delivery. Since this report has been written for an international audience, it is important to stress the weight that is given to the attitudes of foreign researcher in the long-term planning of the Data Base.

Documentation

The Demographic Data Base: A Short Introduction.
The Demographic Data Base: Sources and Codes.
Further Documentation is under preparation describing the different error checks, the process of linkage and the file-system.

This documenation can be ordered from:
The Demographic Data Base
Management and Research Unit
Humanisthuset
University of Umeå
S–901 87 UMEÅ Sweden

Anyone wishing to be on the mailing-list of the Data Base is welcome to apply to the same address.