3 Methodology and organisation of the Longitudinal Study

3.1 STUDY DESIGN

The Longitudinal Study (LS) was designed as a continuous, multi-cohort study. The initial study population was selected as a random 1 per cent sample of the total population of England and Wales clustered by date of birth.[1] The original sampling frame included every person resident and enumerated (including visitors) in the 1971 Census of Population. The sample population was selected if their birthdates coincided with one of four dates of birth (month and day). The same four dates of birth were used to extract sample members from the 1981 and 1991 Censuses. Between censuses new members (with LS dates of birth) enter the sample via birth or immigration and exit from it through death or emigration. By recording entrances and exits to the study, the sample population should accurately reflect the demography of the parent population over time. However, the recording of both emigration and immigration is known to be incomplete (see Chapter 7, sections 7.1.2 and 7.2.2) and to this extent the LS departs from a true population sample.

The study was designed so that data about LS members would be acquired from the answers to census questions and from linking administrative data routinely collected by OPCS to the sample. It was therefore essential that the linkage of events occurring to members (e.g. births, deaths, cancer registrations, etc.), the addition of new members into the sample and the linkage of individual records from census to census should be achieved with a very high degree of accuracy. To achieve this level of accuracy all linkage in the LS depends crucially on the National Health Service Central Register (NHSCR) which acts as the avenue through which the linkage of events and census data to LS members takes place (see section 3.3).

3.2 ORIGINAL SAMPLE SELECTION

The original sample was selected from those persons enumerated in England and Wales on 25 April 1971 (Census day). The record of each person for whom one of the LS dates had been recorded as their birthdate was extracted from the 1971 Census computer files. A unique, eight digit serial number was then assigned to each LS member. A set of punched index cards was created containing essential census data for the identification of each sample member. This included information such as ward and enumeration district number (to allow the census schedule to be identified), sex, marital status, usual residence, date of birth and the allocated LS serial number. From this information each relevant census schedule was identified and name, usual address and enumeration address details were written onto the index card. The first three characters of the sample member's surname were punched onto each card and after sorting into partial alphabetical order they were sent to the NHSCR at Southport. There they were used to identify and flag the record of each sample member in the Central Register.

At the same time as the Longitudinal Study in England and Wales was set up, a parallel study was established in Scotland. All cards for members of the Scottish sample were sent to the General Register Office Scotland (GRO(S)). The Scottish LS was discontinued soon after the 1981 Census, following the Rayner Review of the Government Statistical Service. The final work done on this data set was to complete the linkage with the 1981 Census. The arguments in support of suspension of the study were a perceived lack of demand for and usage of the data compared with the high cost of collection and processing. There were a number of reasons for this. Little analysis of the data was carried out in GRO(S), with the result that the existence and potential of the study was not so widely known as the England and Wales LS. Also, as with a number of other data sources, the sample size of approximately 50,000 members created problems with analysis at some levels of aggregation. In 1991 this policy was reviewed and potential academic customers were approached to gauge possible demand. The response was not encouraging. No linkage with the 1991 Census was carried out and the Scottish LS was effectively abandoned.[2]

3.3 THE ROLE OF THE NHSCR

NHSCR is central to the successful running of any linkage study involving vital events occurring to the population of England and Wales. Population registers and personal identification numbers are not legal requirements in the UK. As a result, tracing members of the population and linking events to them has to be done using the only commonly held means of identification, the NHS number. The NHS number, unlike other means of identification, such as the National Insurance number, is issued at birth registration and thus allows linkage of events to children as well as adults.

The Central Register (originally known as the National Register) was set up in September 1939 when a full enumeration of the population of England and Wales was carried out at the outbreak of the Second World War. The Register was structured geographically with persons being recorded within their families and households, within registration districts. Each individual was given an identification number which was also used on food rationing

cards. This number became the individual's NHS number in 1948 when the National Health Service was created. The register had a dual purpose until 1952 when both national registration and food rationing were abolished.

Each person enumerated on 29 September 1939 was issued with a four-letter code assigned to them by the enumerator in their enumeration district. The National Registration number was made up of the person's enumeration district code followed by the number of their enumeration schedule and the number of the person on the schedule.

The entries on the schedules were subsequently transcribed into registers, each containing approximately 7,000 entries in code number order. Each entry contains the surname, forenames, address, sex, date of birth, marital status and occupation of each person as given on the schedule.

In 1948, after the creation of the NHS, identification numbers allocated in 1939, or ration card numbers given to those born after 1939, were used to form NHS numbers. A block of NHS numbers was specially allocated for those members of the armed services who were demobbed in 1948 after the end of the Second World War. Subsequently, NHS numbers have been allocated in sequence within registration districts, at birth registration. Immigrants to the country are given an NHS number by NHSCR only after registering with a general practitioner (GP).

It is planned to allocate new NHS numbers to the population in 1995 when individuals present themselves for medical treatment. Although a new number will be issued, the old number will be retained as well. Only new births and immigrants will have a single NHS number after the introduction date.

3.3.1 NHSCR — pre-computerisation

Until 1991 all routine events were clerically recorded in the registers. These events included births, deaths, cancer registrations, enlistments into the armed forces, embarkations, entry into long-stay psychiatric hospitals, re-entries to the NHS, and internal migration. Each register contained one line per person in NHS number order and a record was thus held for all the population with the exception of immigrants not registering with a GP.

In the space allotted for each individual in the register each Family Health Service Authority (FHSA) (previously NHS Executive Committee, then Family Practitioner Committee) posting was entered together with a cipher denoting any notifiable events, such as enlistment or cancer registration. If a person was included as a member of a medical research study a cipher or flag denoting entry to the particular study was also entered in the register. Approximately 60 per cent of event notifications include the NHS number but the remainder either do not quote the number or have it entered incorrectly. Various alphabetical indexes are available to assist in tracing the NHS number where it is missing or incorrect.

The alphabetical indexes contain name, date of birth and NHS number. Annual indexes of births registered since 1939 in England and Wales, containing details of date of birth and NHS number, are also kept.

3.3.2 The creation of the initial LS indexes

When the 1971 based LS index cards for England and Wales arrived at NHSCR each sample member had to be identified (traced) within the Central Register. The alphabetical indexes were used to identify the member, and the LS member's entry in the Central Register was flagged. The NHS number of the member was then transcribed onto the LS index card which was filed alphabetically under the member's full name in the LS alphabetical index.

Not all 1971 Census LS index cards could be traced (i.e. the NHS number found). By the end of 1976, when the files were 'frozen' and the LS alphabetical index created, 3.2 per cent of LS members (16,883 cards) could not be traced to an entry in the Central Register. These untraced cards were then used to create a 'no trace' LS index. It should be noted that some of these cases were subsequently traced after 1976 and are discussed further in Chapter 8.

The reasons behind this inability to trace the NHS number included:

(a) The date of birth given on the census form as an LS date of birth was inconsistent with that held in NHSCR records. There were two reasons for the discrepancies between the birthdates given at census and those on record at NHSCR:

  (i) The head of household completing the census schedule may have given an incorrect birth date for the LS member.

  or (ii) The birthdate given on the records at NHSCR may have been incorrect or different from that given on the census schedule.

(b) Women who had married but had not notified their doctors of a change of name, or who had moved on marriage and not yet registered with a new doctor, would still be held on the registers at NHSCR under their maiden names. When enumerated in the census they would, in the majority of cases, have given their married name. Any person who was inconsistent in quoting their name, so that one name was held at NHSCR and another quoted at census, may also have been untraced.

(c) An immigrant to England and Wales in the period prior to the census who had not registered with a doctor by the census date would not hold an NHS number.

(d) If the address given on the census form was not known to NHSCR because that person had moved area, but not yet registered with a new doctor, then they may not have been traced.

Staff at NHSCR were (and are) expert at resolving difficult tracing problems and used a number of methods to attempt to find the correct NHS number or changed name or address for untraced cases. For LS members who could not be found in the registers the usual residence as given at enumeration was used to identify the relevant local NHS Executive Council (EC). The ECs (later Family Practitioner Committees, now Family Health Service Authorities) were contacted by letter and asked if they could supply the NHS numbers for the untraced LS members. In 70 per cent of cases this proved successful. If the NHS number was identified as Scottish, the LS index cards were sent to NHSCR Edinburgh where the LS members were flagged in the Scottish registers and the cards returned to Southport.

Where the cards were still untraced the original census schedules were checked for further details, including year of entry to the country (if born overseas) and the names of other family members. Registration records and electoral registers were also used for this checking and date of marriage was also included in the search criteria. These extra checks frequently produced enough information to allow NHSCR to trace the person successfully.

3.3.2.1 1981 Census–LS link

An overview of the 1981 Census–LS link is given below. Full details (including details of the follow-up of LS members resident in Scotland are given in Chapter 5, section 5.1.3).

After the 1981 Census the role of NHSCR expanded to include the forward linkage of LS members from census to census as well as tracing. OPCS again produced index cards containing census data to identify LS members and each card was allocated a 1981 series LS number. As with the 1971 cards, name and address were added to them manually. On receipt of the cards at NHSCR a matching and linking exercise was undertaken.

If a card already existed in the LS alphabetical index, the 1971 LS number was added to the 1981 card. The card in the LS index may have been created as the result of the original sampling from the 1971 Census or may have been added during the decade if it was for a new birth or an immigrant. Where the LS member had been present at the 1971 Census, or an intercensal LS number had been allocated, the cards were said to be matched and the linkage was provided by the addition of the 1971 LS number. New entrants at the 1981 Census (not new births or immigrants who had entered the LS intercensally) were allocated a 1981 series number and had no previous LS number.

Matching of cards was done using both the trace and no trace indexes and new entrants were checked on the Central Register for an existing NHS number. Ninety-seven per cent of LS members enumerated in the 1981 Census were traced at NHSCR and 91 per cent of those LS members traced and flagged at the 1971 Census who had not died or emigrated intercensally, were successfully forward linked to the 1981 Census[3]. After the linkage exercise was completed the cards, carrying the 1971 LS number where available, were processed further (see Chapter 5).

3.3.3 NHSCR — post-computerisation

Computerisation of the records at NHSCR occurred in March 1991 when a new computerised Central Index was created. This was known as the Central Health Register Inquiry System or CHRIS database. The database was manufactured from the merging of 98 Family Health Service Authorities' computerised files and it is automatically updated with data from the FHSAs. Data are downloaded daily from the FHSAs to the CHRIS system via a direct communications link and turnaround of information occurs over a period of two to five days. It should be noted that automatic updating can only take place where a match is found, in approximately 70 per cent of cases. The remaining 30 per cent of cases where there is no match still require clerical updating. All LS members are now flagged on the new Central Index and the initial load of the LS index onto the CHRIS system provided a massive revalidation of the LS index at NHSCR.

3.3.3.1 1991 Census–LS link

An overview of the 1991 Census–LS link is given below. Full details are given in Chapter 5, section 5.3.

After the 1991 Census, index cards were again produced from the census data at OPCS for NHSCR to use to trace the LS members and link the 1991 LS sample over the censuses. The tracing and linkage were done using the Central Index database where possible and the original index of LS members where not. In 1991, 98 per cent of LS members' cards were successfully traced at NHSCR and of these, 79 per cent had also been enumerated at the 1981 Census. Fourteen per cent were new births or immigrants between 1981 and 1991, 4 per cent were new entrants at census who were neither new births or immigrants, and 3 per cent were LS members who had been sampled in 1971, missed the 1981 Census and were resampled in 1991. Forward linkage of eligible LS members from the 1981 Census sample (that is those who had neither died nor emigrated intercensally) stood at 90 per cent. As in 1981 the cards were processed further in order to merge the census data into the LS dataset.

3.4 UPDATING THE LS SAMPLE

An overview of the updating of the LS sample with events data is given below. A detailed account of the events linked to the LS sample and the methods used for that linkage is given in Chapter 6.

The LS was designed to be a sample with continuous updating of the population. To achieve this the components of population change, births, immigrations into England and Wales, deaths and emigrations out of England and Wales are added or subtracted from the sample as necessary. These are all events that are routinely notified to, and recorded at, NHSCR. Although the recording of births and deaths is almost 100 per cent complete, the notification of immigrations and emigrations is substantially lower. It should be noted that an 'exit' via death or emigration does

not mean that the record of the LS member is removed from the files, rather that the exit event is recorded for that member. The LS database is updated with an indicator to note that a death or emigration has occurred and in the case of a death no further events can normally be notified. The one exception to this is when a cancer registration is notified to NHSCR after the date of the person's death. A time lag in the system of cancer registration means that the cancer date may be received several years after an individual has died. An emigrant may re-enter the country (and the NHS) at a later date and be reactivated within the study.

The LS is also updated with other events occurring to the sample members. These events include births to LS sample members, infant mortality to children of sample mothers, deaths of the spouses of LS members, cancer registrations, enlistments into the armed forces, entries into long-stay psychiatric hospitals (until 1984), and re-entries into the NHS from either the forces or from psychiatric hospitals (only for original admissions prior to 1984).

Prior to the computerisation of the records held at NHSCR, the LS indexes were updated manually with deaths and any new entrants were given an LS index card. Information on immigrations, emigrations, enlistments, entries and exits from long-stay psychiatric hospitals, re-entries, deaths and cancer registrations was entered in the NHSCR registers and in the OPCS computer files. After 1991 all LS members were flagged on the new CHRIS database and the methods of linking events were changed.

3.4.1 New entrants

New entrants to the LS comprise births occurring in England and Wales on LS dates and immigrants whose birthdate is an LS date. The addition of new births to the sample is relatively straightforward, as all births in England and Wales must be registered by law. A draft entry form which carries full details of each birth is sent to OPCS by the Registrar of Births and Deaths for statistical processing. NHSCR Southport receives a short form from the Registrar carrying details of the baby's name, date of birth and NHS number which is used for updating the registers.

The addition of immigrants is more difficult as they are only issued with an NHS number and entered into the registers at NHSCR when they register with an NHS doctor. The category of immigrant includes not only those individuals who describe themselves to their general practitioners as such, but also those who, having quoted a previous address abroad, cannot be matched to an existing NHS number. This category includes people moving from Scotland, Northern Ireland, the Isle of Man or the Channel Islands into England and Wales as well as those entering from outside the UK.

3.4.2 Events data

The event data recorded in the LS include not only exits from, and entries to the study but also events that occur to existing LS members. Some of the event data are acquired from routine notifications to NHSCR. Most event data however, are generated from routine statistical processing by OPCS and are then linked to the LS members at NHSCR. Deaths and cancer registrations are linked using a dual system which helps to improve completeness. (For further details of linking event data see Chapter 6.)

Some events are detected from the date of birth of the LS member as stated on the event notification document. A listing is generated from the annual computer file giving details of the events of interest occurring to persons with an LS date of birth. Because no names are included on the file the event notification document is then extracted and used to identify the person. Drafts and, in some cases, listings are then sent to NHSCR for linkage to be performed. Events identified in this way are (with approximate annual numbers):

(a) New births of LS members (7,100);

(b) Immigrants and re-entrants from abroad (2,500);

(c) Births to sample members (6,000 each to LS mothers and LS fathers. Births to LS fathers were only collected between 1971 and 1978, and from Census day 1981 to the end of 1981);

(d) Widow(er)hood of sample members (2,300);

(e) Infant mortality (200).

A dual system is used for deaths and cancer registrations. The event notifications are crosschecked at NHSCR using both routine notifications and the OPCS annual computer files. The approximate annual numbers of these events are as follows:

(a) Deaths of LS members (6,100);

(b) Cancer registrations (2,100).

Some events are linked to the LS member when the registers at NHSCR are updated from routine notifications and the LS flag in the register identifies the person to whom the event has occurred as a member of the study. Events included in the LS using this method are as follows (with approximate annual numbers):

(a) Internal migration, movement between FPCs (16,000) (included for the years 1971-74);

(b) Enlistments into the armed forces (300);

(c) Embarkations (1,200);

(d) Entries into long stay psychiatric hospitals (50) (included for the years 1971-83);

(e) Re-entries into the NHS from (b), (c) and (d) (500).

3.5 CREATION OF THE ICL COMPUTER FILES

When the selection and tracing of the original 1971 LS sample was complete, a set of ICL computer files was produced. These files were created to hold data acquired from the 1971 Census schedules on the LS members and their households. A similar set of files was created after the 1981 Census to hold data from the 1981 Census schedules. A set of events files was created for each decade, with one file per event per decade. These files were held separately from the LS files holding census data.

The requirements of the LS were such that all data for LS members and their households had to be fully coded for inclusion in the appropriate LS personal and household files. The Census Office do not fully code all items on census schedules during the census processing stages. 'Hard-to-code' items such as economic activity, migration, travel to work, etc., are fully coded for only 10 per cent of the main census records. In 1971 these items were fully coded specifically for all LS members and others in their households. In 1981 and 1991 the 'hard-to-code' items were again fully coded (with a few minor exceptions noted later) for the LS. Full details of the census data coded at each census for the LS members are given in Chapter 4.

Because two different forms are used to elicit information from households and institutions (hotels, hospitals, prisons, etc.), the main census records differ in detail according to where persons were enumerated. For people enumerated in households, the record for the entire household of the LS member was extracted. For those LS members residing and enumerated in institutions, only personal information is available.

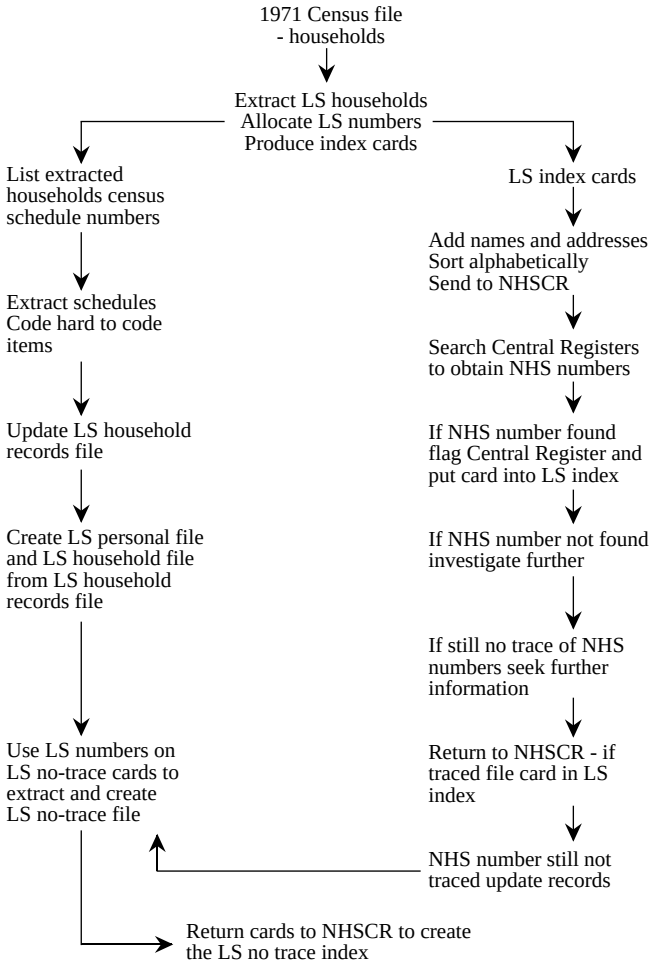
Access to the LS data files was not direct. Certain data (date of birth of LS member, LS serial number, etc.) were not available for any analysis due to confidentiality issues. The other files used by the LS, containing events data, were only linked to the census personal and household files when information on events occurring to LS members was required. Data were extracted from these files (known as source files) in the form of flat files (known as datastreams). The datastreams also contained derived variables which were created, using algorithms, from variables in the source files. Datastreams were used until early 1994 as the only datasets for analysis, and are still used for some types of analysis not requiring 1991 Census data.

3.5.1 1971 computer files

Information from the LS index cards was used to create two sets of computer files. The first file, the LS personal file, consisted of the anonymised census records for all LS members (both traced and untraced). It contained all the information recorded in the 1971 Census for each LS member, together with some details of the household in which the LS member was enumerated. The second file, the LS no trace file, was of exactly the same format as the first, but carried information relating to the untraced cases only. A further file was created containing data on all the members of households where at least one person was included in the LS. This file was called the LS household file.

Figure 3.1 shows the processes involved in the creation of original 1971 LS files from the 1971 Census files.

Figure 3.1 Creation of the 1971 LS computer files



3.5.2 1981 computer files

After the 1981 Census a single LS personal and household file was created containing data from the 1981 Census. The information in the 1971 and 1981 files could only be linked together via the LS serial numbers used to identify LS members. Details of the process of matching and linking LS member records from census to census is covered in Chapter 5, section 5.1.

3.6 TRANSFER TO THE MODEL 204 DATABASE

In 1987 a decision was made to move OPCS's work from VME files to a database management system. The database

management system chosen was Model 204 which runs under the MVS operating system. The Model 204 database could not run on the ICL machine and a new Amdahl machine was chosen to replace it. This particular database management system was considered to be the most suitable for dealing with the large amount of processing involved in both the day-to-day statistical functions of OPCS and in specialist applications such as the Census and the Longitudinal Study. Transferring the LS to Model 204 therefore required a complete redevelopment of the computer system from one based on flat files to database.

Accessing data from the ICL LS files, for extraction of a datastream, involved using a unique key (the LS number) to read data from individual records. This method was slow and involved sequentially reading through a set of records to extract the required data. With the increasing amount of data to be held in the LS files, access was likely to become slower. Model 204 was one of the few database management systems which could handle the size and complexity of the redeveloped LS system (37 logical entities stored on 31 separate physical files, occupying approximately three gigabytes of storage).

Model 204 is an inverted list database which means that information contained in one or more of a file's records can be accessed using a single field name (or variable name) such as 'age' where that field name equals a value such as '25'. All records where age equals 25 would be selected. By ensuring that the most commonly used field name-value pairs are stored together in an index giving the location of all occurrences of the pair, search times are shortened. Only the most commonly used search criteria are indexed in this way because of the need for large amounts of storage space. However, this method of indexing data does allow for both the use of complex search criteria and easier linkage of data. Other indexing and search methods (hashing algorithms, B trees and bit mapping) are also used to access data. The multiple indexing systems and the use of primary and secondary indexes add to the speed of retrieval.

As well as providing fast methods of handling and retrieving data, Model 204 has other data management advantages, particularly in the area of security. Data within the database can be protected with different levels of security down to the level of a single field. In addition to this, the operating system, MVS, provides entry level security utilising IBM's Access Control Facility software (ACF2).

3.6.1 Structure of the LS database

Unlike the ICL files, where 1971 and 1981 data were held in physically separate source files which could only be linked together using a Cobol program to extract a rectangular datastream file, the new database links these files (and the 1991 files) together using field name-value pairs. Field names used in this way are defined as keys; thus in the LS, specifying a key field of 'sex' with the value of 'male' would select all LS member records where sex is male. To link all other data relating to those males selected using the field name-value pair, the LS number would then be used. Data are accessed using a utility designed to extract subsets of data either in the form of SAS datasets or SPSS system files for further analysis. An outline of the structure of the M204 LS database is shown in Figure 3.2.

3.6.2 Transfer and loading of data

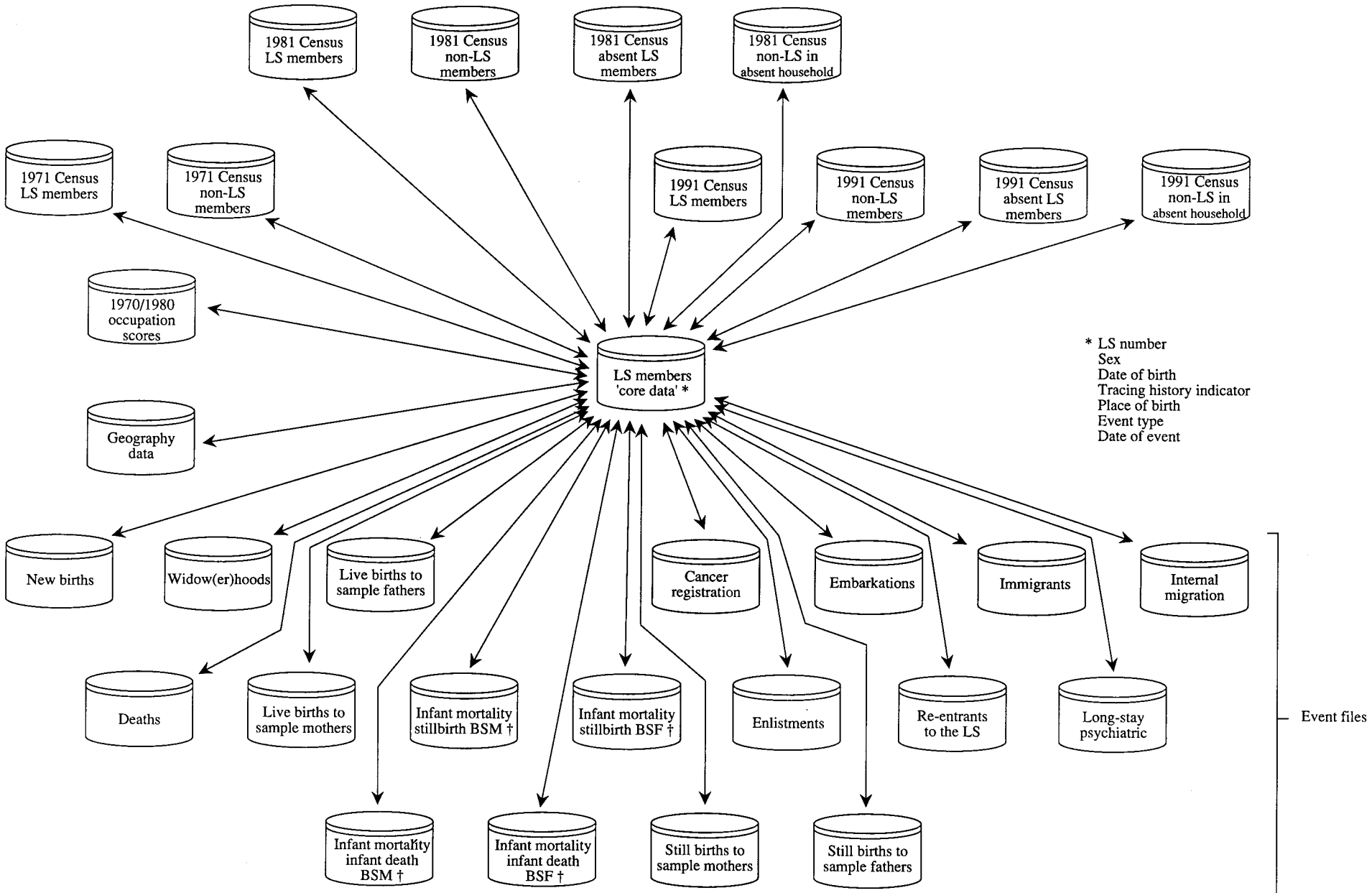
Once the new database had been created it was initially loaded with historical (pre-1991 Census) data previously held in VME files on the ICL machine. The files that had to be transferred included the 1971 and 1981 Census–LS files and the event files for events occurring to LS members from 1971 up to the 1991 Census. After the historical data were loaded, the 1991 LS Census data were linked into the database. It should be noted that the 1971 Census–LS data held in the database are structured around LS households, allowing duplication of data if two LS members are present in a single household. This single household is therefore represented as two households in the database. 1981 and 1991 Census–LS data have been structured using the concept of the Census household rather than the LS household. As a result if there are two or more LS members present in the same Census household this will be represented as two LS households but one Census household in the database. There will therefore be data duplication if two or more LS members reside in the same Census household.

3.6.2.1 Transfer of historical data

Before the historical data files could be transferred it was essential to perform a number of procedures on them. First, the data had to be cleaned to ensure that any files taken across to the new environment were as up to date and error-free as possible. Second, reformatting of some files in VME was needed to ensure that they would not only transfer to MVS but would also load efficiently from there into the M204 database. Once these procedures had been completed the files were transferred onto the Amdahl as MVS files. The MVS files were then loaded into Model 204 to begin the creation of the M204 LS database.

The first load into MVS was of data held in the three 1971 Census–LS files. These files were reformatted into three MVS files: the 1971 Census–LS core data file, the 1971 LS members file and the 1971 non LS members file. The 1971 LS members core data included the LS number, sex and date of birth of each LS member and the MVS file was used to load the M204 LS members core database file directly. The 1971 MVS LS members file and the 1971 LS non-members file were both loaded from MVS into a single M204 1971 Census file. The separate first decade events source files (covering the years 1971 post-Census to 1981 pre-Census) were then loaded into the 18 LS events database files. First decade new births (on LS dates) and immigrant entry event files were loaded before other first decade events files. The new birth and immigrant entry files were used to update the LS members core file to ensure that other events

Figure 3.2 Structure of the M204 LS Database



† BSM/BSF = Births to sample mothers/Births to sample fathers

(e.g. deaths, embarkations, etc.) would link correctly when they were loaded. The order of loading also ensured that the database would contain the correct rolled-forward population base when the 1981 LS-Census data files were loaded.

After the 1971 data load, the 1981 LS-Census file was loaded into the 1981 LS-Census database file. The LS members core data file was updated on the load from the reformatted MVS 1981 LS-Census files with new records containing the 1981 LS number of any new member added to the sample at the 1981 Census. Any date of birth discrepancies found for existing LS members were flagged on the core records. The second decade (covering the years 1981 post-Census to 1991 pre-Census) events files were then loaded into the existing LS database file with the data from entry event files being input first.

3.6.2.2 The creation of the 1991 LS-Census database file

The loading of data on 1991 Census households containing LS members was done using the 1991 LS–Census household extract file directly in the Model 204 environment. The LS-Census file was then used to update the LS members core data file with new records for the 1991 LS numbers. The process of linking 1991 LS data with data from 1981 and 1971 is discussed in Chapter 5.

3.6.3 Confidentiality and access

The LS is an ongoing study of over half a million members of the population of England and Wales. The sources of data used in this study are protected by two acts of Parliament, the Census Act and the Population Statistics Act. The Census Act covers the confidentiality of data collected at census and does not allow access to individual returns for 100 years. However, access to anonymised data is allowed through certain data sets such as the Samples of Anonymised Records and the LS. The Population Statistics Act covers access to the confidential particulars supplied at birth and death registration. These data can only be used for statistical purposes and can never be released on an individual basis.

Because of these statutory constraints the maintenance of confidentiality and limitations on direct access to data held in the LS are of paramount importance. The data are held in an anonymised form, no names or addresses are held on the database and a number of variables are not accessible for analysis. These variables are essential to the selection of the study population and the linkage of both census and event data to the LS members in that population.

In order to comply with the terms of the Census Act, automatic linkage of data at the time of Census is not allowed. Although names and addresses are used to perform the link they are never entered onto the LS database and are held separately at NHSCR. No census data are held at NHSCR, only the name and address, NHS number, the LS number if a person is a member of the LS, and a variable number of flags denoting FHSA, LS member, etc. The only link between data on an LS member received at NHSCR and that received at OPCS is through the LS number which is never made available for analysis.

Access to data held within the LS database is closely supervised and though individual data are held they can only be released in a grouped form. Primary analysis of data can only be done within the confines of OPCS by persons who have signed the Official Secrets Act. Data may only be released to users for further analysis in the form of statistics or tabulations which do not allow the identification of individuals.

References

1. Fox A J and Goldblatt P O. *Longitudinal Study: socio-demographic mortality differentials.* OPCS Series LS no. 1, HMSO (London 1982).
2. Arrundale J. General Register Office Scotland. Personal communication, 1994.
3. Office of Population Censuses and Surveys. *Census 1971–1981, the Longitudinal Study: linked Census data, England and Wales*. OPCS Series CEN81 LS, HMSO (London 1988).