DISCREPANCY VARIABLES

Introduction

The ONS Longitudinal Study (LS) contains two sets of discrepancy variables. The variables indicate whether the date of birth or sex information from an LS member's initial entry to the LS differs from any subsequent date of birth or sex information that is held in the LS system. This guide is designed to introduce you to these discrepancy variables. After reading it you should:

- Have a better understanding of the discrepancy variables in the ONS Longitudinal Study.
- Decide whether they are relevant for your research.
- Make decisions around how to use them in your research.

What are the discrepancy variables?

The CORE file of the ONS Longitudinal Study (LS) includes two sets of six discrepancy indicator variables. One set focuses on date of birth (DOBD001, DOBD002, DOBD003, DOBD004, DOBD005, DOBD006) and the other on sex (SEXD001, SEXD002, SEXD003, SEXD004, SEXD005, SEXD006). For date of birth, the discrepancy variables indicate if the date of birth in the CORE file (i.e. the information from an LS member's initial entry into the LS) differs from other dates of birth for the LS member that are held in the system. Similarly the sex discrepancy variables indicate if the sex in the CORE file differs from the other information on their sex held in the LS system.

The six date of birth discrepancy variables and six sex discrepancy variables are coded identically[1]:

- -9 – No discrepant date of birth/ sex.
- 1 – 2001 Census LS member.
- 3 – 2001 Census multiply-enumerated LS member.
- 11 – 2011 Census LS member.
- 13 – 2011 Census multiply enumerated LS member.
- 31 – 1981 Census LS member.
- 33 – 1981 Census Absent usually Resident LS member.
- 41 – 1991 Census LS member.
- 43 – 1991 Census absent but usually resident LS member.
- 52 – Live birth to sample mother.
- 53 – Still birth to sample mother.
- 54 – Live birth to sample father.
- 55 – Still birth to sample father.
- 56 – Death of LS member.
- 57 – Widow(er)hood of LS member.

[1] When subsequent data is linked to an LS member, if there is a discrepancy and it is related to an event (e.g. death, cancer registration, live birth) then the discrepancy value will populate the first discrepancy variable and the other discrepancies will be moved to populate the later discrepancy variables. Census related discrepancies tend to populate the later discrepancy variables. Therefore, if the first discrepancy registered for an LS member is with the date of birth given in their 1991 Census and then a live birth is registered with another date of birth discrepancy for the LS member, DOBD001 will be coded 52 and DOBD002 will be coded 41. Then if the next date of birth discrepancy is on the LS member's death registration, DOBD001 will be coded 56, DOB002 will be coded 52 and DOBD003 will be coded 41.

58 – Infant death to sample mother.

59 – Cancer registration of LS member.

60 – Enlistment of LS member.

61 – Embarkation of LS member.

62 – Re-entry to sample of LS member.

63 – Immigration of LS member to NHS.

64 – Long stay psychiatric.

65 – Internal migration Inter-FPC moves.

66 – Stillbirth to sample mother.

67 – Infant death to sample father.

68 – Stillbirth to sample father.

81 – 1971 Census Ward data.

83 – 1981 Census Absent usually resident LS member.

91 – 1991 Census LS member (initially 41 was used but it is now standardised to 91).

93 – 1991 Census Absent usually resident LS member.

There is no discrepancy code for a discrepancy with the 1971 Census because the 1971 Census was the source of the date of birth and sex information that is contained on the LS system and the information in the CORE file. Therefore no LS members will have a discrepancy with the information in the 1971 Census file.

There are also two binary date of birth discrepancy variables in the members' files for 2001 and 2011, DOBDISC0 and DOBDISC11 respectively. These variables indicate if the date of birth given in the relevant Census is different to the date of birth recorded in the NHS Central Registry. They are coded:

0 – Non discrepancy.

1 – Discrepancy in date of birth.

DOBDISC11 has an additional missing category:

-9 – Untraced on original tracing exercise but may have been traced subsequently.

How many LS members have a discrepancy in their date of birth or sex information?

Date of birth

LS members can only have one entry in each of the six date of birth discrepancy variables. Across the six date of birth discrepancy variables the majority of LS members (>92%) have no discrepancy in the date of birth information that is held for them in the LS system, approximately 6% have one occurrence of a discrepancy in their date of birth information and 0.15% have six discrepancies (see Table 1 of the Date of birth sheet in the associated Discrepant variables spreadsheet). The most frequent discrepancy (3%) is with the date of birth that is given in an LS member's death registration (see Tables 2a – 2f of the Date of birth sheet in the associated Discrepant variables file).

Sex

LS members can only have one entry in each of the six sex discrepancy variables. Across the six sex discrepancy variables the majority of LS members (>99%) have no discrepancy in the sex information that is held for them in the LS system, <1% has one or more occurrences of a discrepancy (see Table 3 of the Sex sheet in the associated Discrepant variables spreadsheet). The most frequent discrepancies are with the information in the 2011 Census member file (0.14%), the 1971 Census Ward file (0.14%) and the category that is for a discrepancy with the 1991 Census (3%) (see Tables 4a – 4f of the Sex sheet in the associated Discrepant variables file).

How should I use the discrepancy indicator variables?

We suggest that all LS researchers include the six discrepant date of birth indicator variables and six discrepant sex indicator variables in the list of variables that they request in the LS supplementary form. To make use of them it is vital to include the year of birth variable (DOBYR) and sex variable (SEX) from the CORE file and the corresponding variables from each of the other files that variables are being drawn from.

For example, if a researcher is using variables from the ME91 ME01, ME11 and the LBSM files, they will need to include the following in their variables list:

-   ME91 – SEX9 and DOBYR9
-   ME01 – SEX0 and DOBYR0 (and DOBDISC0)
-   ME11 – SEX11 and DOBYR11 (and DOBDISC11)
-   LBSM – MOBIYRBM[2]

This will mean that if a discrepancy is indicated with the sex or date of birth variable in one of these files, the researcher can examine other variables in their dataset to assess which date of birth or sex information is the most plausible in order to determine how they want to address the discrepancy. Alternatively, researchers may choose to exclude the LS members with a discrepancy from their study sample.

When assessing what to do if there is a discrepancy between an LS member's initial date of birth or sex information and subsequent information that is held in the system the researcher should ask themselves the following questions:

-   Where is the discrepancy? Is the discrepancy with the Census year information that I am using as my baseline?
-   Does this discrepancy matter for my research?
-   Is there other information in the data that I can use to assess which information is correct?

---

[2] There is no sex indicator variable for the LS member in the LBSM file because it is for live birth to sample <u>mothers</u>. Researchers will need to decide whether to assume that the sex information in the CORE file is wrong and/ or whether they want to exclude LS members with a record in the LBSM file who are categorised as men in the CORE file.