

Mads Meier Jæger

02:2015 WORKING PAPER

DANISH LONGITUDINAL SURVEY OF YOUTH (DLSY) - CUMULATIVE
1968-2004 FILE

SFI – THE DANISH NATIONAL CENTRE FOR SOCIAL RESEARCH

DANISH LONGITUDINAL SURVEY OF YOUTH (DLSY) - CUMULATIVE 1968-2004 FILE

Mads Meier Jæger

THE DANISH NATIONAL CENTRE FOR SOCIAL RESEARCH, COPENHAGEN,
DENMARK

Working Paper 02:2015

The Working Paper Series of The Danish National Centre for Social Research contain interim results of research and preparatory studies. The Working Paper Series provide a basis for professional discussion as part of the research process. Readers should note that results and interpretations in the final report or article may differ from the present Working Paper. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including ©-notice, is given to the source.

Danish Longitudinal Survey of Youth (DLSY) - Cumulative 1968-2004 File

Prepared by

Mads Meier Jæger

The Danish National Centre for Social Research

July 2015

Acknowledgments:

The DLSY was initiated by Professor Erik Jørgen Hansen, a pioneer in Danish social science research. Hansen designed and implemented the 1968-1992 data collections. Over the years the DLSY has been financed by The Danish National Centre for Social Research (DK: SFI – Det Nationale Forskningscenter for Velfærd, previously: *Socialforskningsinstituttet*), the Ministry of Education, the Ministry of Social Affairs, and the Danish Council for Independent Research | Social Sciences. The DLSY would not have been possible without funding from these agencies. All code work on the DLSY – Cumulative 1968-2004 File was carried out by Jakob Pindstrup Andersen, Ida Thyrring, Stine Møllegård Pedersen, and Nynne Rasmussen.

1. Introduction

This report presents the Danish Longitudinal Survey of Youth (DLSY; in Danish: *Ungdomsforløbsundersøgelsen*) Cumulative 1968-2004 File. The report describes the background of the DLSY, the content, the data collection procedures, and the available data. The report is intended as a reference for users who want to work with the DLSY Cumulative 1968-2004 file. More information on the DLSY, including how to get the data, questionnaires, documentation, publications, and contact information is available at the DLSY website: www.sfi.dk/dlsy.

2. Disclaimer

It is important that users know what data from the DLSY are available and how they were coded in the DLSY 1968-2004 Cumulative File. We do not have all the data from all DLSY waves. The DLSY 1968-2004 Cumulative File includes all the variables from the 1968-2004 waves that the DLSY team (headed by Professor Mads Meier Jæger) could identify from the original data files. Unfortunately, for some variables (and in particular those included in the early DLSY waves) the available documentation was inadequate or the responses in the original data files did not match those in the questionnaires or documentation. Variables that the DLSY team were unable to document are not included in the DLSY 1968-2004 Cumulative File. They may be included in future updates to the DLSY 1968-2004 Cumulative File if new information becomes available.

It is also important to note that although the DLSY team has done its utmost to ensure that the variables included in the DLSY 1968-2004 Cumulative File are correct, errors may exist. The DLSY team does not assume any liability for any type of damage, whether direct or indirect, arising from the use of the DLSY 1968-2004 Cumulative File. Please help us to improve the DLSY 1968-2004 Cumulative File by reporting errors and inconsistencies to dlsy@sfi.dk.

3. Background

3.1 The DLSY

The DLSY is an ongoing longitudinal study of a nationally representative sample of 3,151 Danish respondents who were born in or around 1954. The main respondents were first interviewed in 1968 when they were around 14 years old and attended the 7th grade of elementary school. The main respondents have since been interviewed in 1970, 1971, 1973, 1976, 1992, 2001, and finally

in 2004 when they were around 50 years old. Furthermore, one of the respondent's parents was interviewed in 1969, and the class teacher in the respondent's school class was interviewed in 1968. All survey waves are described below.

The DLSY was initiated in 1968 by Professor Erik Jørgen Hansen as a follow-up on several earlier studies of educational outcomes among young people in the mid-1960s (e.g., Hansen 1968). The purpose of the 1968 DLSY survey (at the time it was not known that the DLSY would become a long-running study) was to analyze the determinants and consequences of educational achievement and attainment. In particular, the survey was designed to facilitate analyses of the impact of family background (parents' socioeconomic position, family type, housing conditions, aspirations, peer influences, etc.) on educational achievement (measured by a series of IQ tests) and educational attainment (measured by later educational attainment). The class teacher in respondents' school class also provided information on the class in the 1968 survey. Follow-up surveys in 1970 and 1971 (age 16-17) tracked, among other things, respondents educational choices at the end of elementary school, and the 1973 and 1976 surveys (at age 19 and 22) tracked educational careers after secondary school (typically tertiary education). Since then the DLSY respondents have been interviewed in 1992 (age 38), in 2001 (age 47), and in 2004 (age 50), thereby providing rich longitudinal data on issues such educational and occupational careers, family formation and fertility, attitudes and aspirations, health, social contacts and relationships, and leisure time activities. The questions and themes in each wave vary.

The DLSY is particularly suited for analyses of social stratification, intergenerational social mobility and transfers, intergenerational social relations, and life course research. The DLSY is similar in structure, and can be used for comparative analysis with, other longitudinal studies such as:

- *The British National Child Development Study* (which samples the 1958 birth cohort) and the *Birth Cohort Study* (which samples the 1970 birth cohort, see <http://www.cls.ioe.ac.uk/page.aspx?&sitesectionid=724>).
- *The Wisconsin Longitudinal Study* (1939 birth cohort, see <http://www.ssc.wisc.edu/wlsresearch/>).
- *The National Longitudinal Survey of Youth 1979 and 1997* (1957-1965 and 1980-1984 cohorts, see <http://www.bls.gov/nls/>).

4. Sampling and Data Collection Methods

This section describes the sampling method and mode of interview used in each DLSY wave. The section also briefly describes the substantive content of each survey wave. For a detailed description of the variables in each survey wave please see <http://www.sfi.dk/documentation-6785.aspx>. For a summary view of the DLSY data structure see the DLSY flowchart below (Figure 1).

4.1 The 1968 Survey

4.1.1 Sampling

The first 1968 DLSY survey used cluster based sampling. Among the 3,540 7th grade elementary school classes in Denmark registered by the Ministry of Education in the school year 1967/68, a nationally representative sample of 152 (or just over 4 percent of all) school classes was selected and all pupils in these classes were included in the DLSY. The DLSY sampling framework consisted in stratifying all 7th grade classes in Denmark into 45 strata according to, first, the level of urbanization in the area in which the elementary school was located (eight urbanization categories were used) and, second, according to the type of tracking in the class (ordinary track, academic track, not tracked). The non-tracked classes were furthermore stratified according to school type (whether or not the school offered 8-10th grade and whether it offered lower secondary education) and whether the school was located in a parish (i.e., smaller rural) municipality (Jæger et al. 2003). The 152 school classes in the DLSY were sampled proportionally to the number of (total population) classes in each of the 45 strata, with the additional requirement that at least one class from each stratum should be included in the DLSY. This sampling framework makes the DLSY sample largely nationally representative all 7th grade pupils.¹ Of the 3,304 eligible students 3,151 were interviewed giving a response rate of 95 percent (143 students were not in school and

¹ The requirement that at least one class from each of the 45 strata be selected, and the fact that the sampling framework did not take into account that class sizes varied somewhat across Denmark, means that some of the 45 strata are over- or overrepresented in the final DLSY sample. The problems were minor, however.

10 refused to participate) (Jæger et al. 2003: 38). The variable A1 identifies which school DLSY respondents belonged to in 1968.

Respondents in the DLSY sample were born in or around 1954 and, consequently, they were around 14 years old at the time of the 1968 survey. Specifically, in the sample of 3,151 respondents 20 were born in 1952 or earlier (1 percent), 387 were born in 1953 (12 percent), 2,595 (82 percent) were born in 1954, and 139 were born in 1955 or later (4 percent). Birth year is missing for 10 respondents (0 percent).²

4.1.2 Content

The 1968 survey included three questionnaires:

Pupil Questionnaire A: This questionnaire contains questions about family structure, housing conditions, parents' education, occupation, and income, number of siblings, educational and occupational attainment for older siblings, pocket money, time spent on homework, attitudes towards school, work, and leisure time, time spent with friends, and attitudes towards future place of residence and work, occupation, and education. In total there were 73 questions in the questionnaire, of which 48 are represented in the available DLSY data. This questionnaire was completed in class with teachers reading out questions (group enquete).

Pupil Questionnaire B: This questionnaire contains several batteries of questions on respondents' attitudes towards gender roles, work, adults, parents, school, moral issues, and their expectations of the future. In total there were 100 questions in the questionnaire, all of which are included in the available DLSY data. This questionnaire was completed by respondents while in class (group enquete).

Questionnaire C: This questionnaire includes a range of cognitive ability tests designed to test three aspects of respondents' cognitive ability. The first set of tests deals with respondents' verbal ability and includes several linguistic comprehension and power of abstraction

² Frequencies are based on the currently available data. Most of the birthdates are from the 1968 survey, but for 70 respondents the birthdate from the 1968 survey was consistently different from the information given in later waves. In these cases we used the most consistent information from the later waves. For 153 respondents the birthdate was missing in the 1968 survey, and for these respondents we were able to impute the birthdate for 143 respondents (using information from later waves).

tests. The second set of tests deals with respondents' *spatial ability* and includes tests of three-dimensional and geometric comprehension. Finally, the third set of tests deals with respondents' *inductive ability* and includes a series of mathematical and logical tests. The three tests are described in Ørum (1971: 25-28) and Hansen (1995: 67-68, 101-102). Respondents' total score in each of the three tests is available in the DLSY dataset. Respondents' test scores are highly correlated, with a single latent "IQ" factor explaining more than 60 percent of the variance in the three tests.

4.2 The 1968 Class Teacher Survey

4.2.1 Sampling

The class teachers in the 152 school classes sampled for the DLSY filled out a postal questionnaire as part of the 1968 data collection.

4.2.2 Content

The Class Teacher Survey includes descriptive questions about the teacher (gender, age, experience, no. hours taught, etc.), and the teacher's assessment of the class along several dimensions (academic, social, etc.). As the documentation for this survey is unfortunately not available, it is unknown how many questions there were in the original questionnaire. In the available DLSY data file eight questions from the class teacher questionnaire are included.

4.3 The 1969 Parental Survey

4.3.1 Sampling

The 1969 Parental Survey sampled one of the DLSY respondent's parents (mother, father or legal guardian). The parent filled out a postal questionnaire.

4.3.2 Content

The Parental Survey included questions on parents' demographics, education, occupation, contact with the DLSY respondent's school (parent-teacher meetings etc.), the extent to which parents help the DLSY respondent with homework, and attitudes towards schooling. Information on parents' income in the year 1967/1968 was also obtained from the tax authorities. There were 31 questions in the questionnaire of which 27 are included in the available DLSY data.

4.4 The 1970 Survey

4.4.1 Sampling

The 1970 survey consisted of two components: 1) a group enquete questionnaire with the vast majority of DLSY respondents (by then around 16 years old) who were still in school and 2) individual face-to-face interviews with respondents who had left school. In the available DLSY data only the second survey is included.

4.4.2 Content

The 1970 survey of the (small group of) DLSY respondents who were not in school were asked about leisure time activities, friends, attitudes towards work and the future, and reasons for leaving school. These respondents were typically engaged in some type of work or vocational training. There were 141 questions in the questionnaire of which 125 are included in the available DLSY data.

4.5 The 1971 Survey

4.5.1 Sampling

The 1971 survey sampled the 1,124 DLSY respondents who were not currently in school (interviews were carried out with 1,040 of these respondents). These respondents were interviewed face-to-face.

4.5.2 Content

The 1971 survey contained questions relating to respondents' family, housing, friends, and attitudes towards work and the future. There were 142 questions in the questionnaire of which 121 are included in the available DLSY data.

4.6 The 1973 Survey

4.6.1 Sampling

The 1973 survey sampled all DLSY respondents for face-to-face interviews. By 1973 respondents were around 19-20 years old and were attending or had completed secondary education.

4.6.2 Content

The 1973 survey focused on educational choices and labor market participation, economic situation, family formation, and attitudes towards work, politics, and a range of general moral issues. There were 121 questions in the questionnaire of which 111 are included in the available DLSY data.

4.7 The 1976 Survey

4.7.1 Sampling

By 1976 the respondents had reached their early 20s and were for the most part engaged either in higher education (about 30 percent) or in the labor market. The 1976 survey used face-to-face interviews.

4.7.2 Content

The 1976 survey included questions on educational attainment, labor market status, economy, housing conditions, family situation, fertility, and attitudes towards politics and a range of general moral issues. There were 82 questions in the questionnaire of which 69 are included in the available DLSY data.

4.8 The 1992 Survey

4.8.1 Sampling

The 1992 survey was the first survey of the DLSY respondents in “adulthood” (at around age 38). The 1992 survey used face-to-face interviews.

4.8.2 Content

The 1992 survey asked respondents to report on their socioeconomic position, completed educations, income, working conditions, family situation, health, leisure time activities, psychological well-being, attitudes toward child rearing, reading habits, language skills, and contact with other family members. There were 73 questions in the questionnaire of which 69 are included in the available DLSY data.

4.9 The 2001 Survey

4.9.1 Sampling

The 2001 survey was carried out via face-to-face interviews when respondents were approximately 47 years old.

4.9.2 Content

The 2001 survey asked respondents to report both on themselves, their parents, and their children. Information on respondents included detailed recording of marital and fertility histories, current occupational status, income and wealth, family background (economic hardship in the home when respondent was growing up, geographical mobility), social connections/networks, social and educational attitudes, social contacts with children and other family members, economic transfers to children, and health. Respondents were also asked to report on parents’ mortality and health and their socioeconomic position, working conditions, and wealth (ownership of home, car, second home etc.) when respondents still lived at home. Finally, respondents were asked to report on their children’s educational attainment and labor market position. There were 94 questions in the questionnaire of which 92 are included in the available DLSY data.

4.10 The 2004 Survey

4.10.1 Sampling

The 2004 survey was carried out via face-to-face interviews when respondents were approximately 50 years old.

4.10.1 Content

The 2004 survey was much smaller in scope than the previous waves and included only 7 questions (all of which are included in the available DLSY data). These questions concerned parents' social connections when respondents were growing up and whether these social connections have helped respondents later in life.

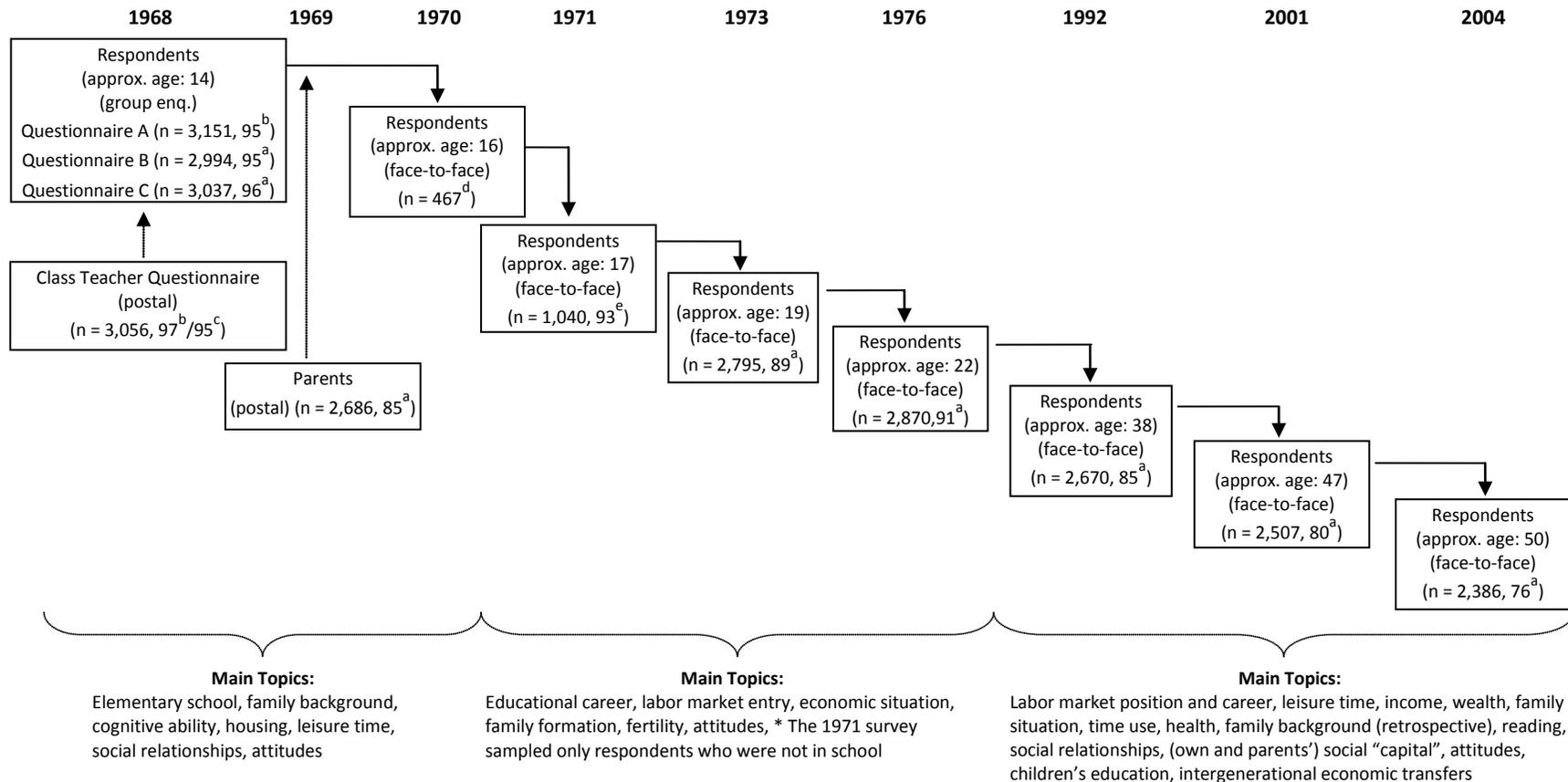
5. Response and Refusal Rates

Figure 1 summarizes data collection methods, topics included, approximate age of respondents, sample size, and response rates in each DLSY wave. Response rates in the DLSY Cumulative 1968-2004 file are calculated on the basis of the available DLSY data as the share of the 3,151 original participants who have provided at least one valid response to a question. The response rate for the 1968 Questionnaire A is calculated as the share of the total sample of 3,304 eligible students in the original sample to be interviewed (as we noted earlier, not everyone in the 7th grade class sampled for the DLSY completed the 1968 questionnaires). For the teacher questionnaire we calculate the response rate both on the basis of the number of respondents and the number of classes, as teachers completed one questionnaire for the entire class. In the case of the 1971 wave, the response rate is calculated on the basis of the number of respondents who had left school (1,124 respondents, representing 36 percent of the original 3,151 respondents). As it is unknown how many respondents had left school in 1970, we were not able to report a response rate for this wave.

The response rates in the DLSY have consistently remained very high over the 35+ years of data collection. The drop in response rates over the period is partially due to respondents moving out of the country or dying. In the latest 2004 survey around 76 percent of the original 3,151 sample members were successfully interviewed.

Please note that because the survey data from the DLSY Cumulative 1968-2004 file can be linked with register data from Statistics Denmark, a considerable amount of information (for example, data on income, marital status, unemployment history, health care use; typically from 1980 onward) exists for all 3,151 original sample members even though they have dropped out of the DLSY survey. Due to data security legislation, the register data are not included in the public use DLSY file, but they can be accessed with special permission from Statistics Denmark.

Figure 1: Flowchart for the Danish Longitudinal Survey of Youth 1968-2004 (DLSY Cumulative 1968-2004 file)



Notes: ^a Calculations are based on the available DLSY data and represents the share of the original sample of 3,151 respondents who have answered at least one question in the survey wave. ^b The response rate for the 1968 Questionnaire A is the share of the total number of respondents in the school classes sampled for the DLSY who actually participated. ^c This response rate is calculated as the percentage of classes out of the 152 classes originally sampled for which the teacher answered at least one question in the questionnaire. ^d In the available data only the survey of the respondents who had left school is included. As it is unknown how many had left school, a response rate cannot be reported. The 467 individuals interviewed represents 15 percent of the original sample of 3,151. ^e The 1971 sample only included the 1,124 respondents who had left school. The reported response rate is calculated on the basis of this sample size (which represents 33 percent of the original sample).

6. Missing values

Table 1 summarizes missing value codes used in the DLSY Cumulative 1968-2004 file to distinguish different types of non-response. Please note that in some cases it was difficult to determine why a question that was supposed to be answered was in fact not answered. In these cases the DLSY team used available information from other parts of the survey (for example, by looking at the respondent's response to questions before and after the question that was not completed) to assign missing values.

Table 1: The use of missing values

Value	Meaning	Use
-1	"Don't know"	This code is only used in the 1976-2001 waves. In the 1968 and 2004 waves a "don't know" code was only used in four questions in which a "don't know" response was explicitly assigned in the questionnaire.
-2	"Irrelevant"	Non-response resulting from the respondent not being eligible to respond to a question (due to, for example, being filtered out earlier in the questionnaire). Note that cases in which respondents have erroneously answered a question that they were not supposed to answer have not been changed.
-3	"Non-Response"	Default missing value. Used when respondents who were supposed to respond to a question did not provide a valid response.
-4	"Non-Participant"	The respondent did not participate in the DLSY wave
-5	"Coding Error"	The reported response code is inconsistent with the response categories in the questionnaire

7. Analysis of Representativeness and Panel Attrition

Erik Jørgen Hansen carried out a detailed analysis of representativeness and panel attrition in the DLSY up to, and including, the 1992 wave (Hansen 1995). He concludes that there is some evidence of selectivity and selective attrition, but also that these problems are minor. We now summarize his main findings.

In the original sample of 3,151 students, Hansen (1995) found a slight overrepresentation of students from Western Jutland and Bornholm – and likewise a slight underrepresentation of students from the rest of Denmark. This situation was in part caused by the sampling procedure and in part by geographic differences in response rates (Hansen 1995: 61, 77). He also found that in 1992 the geographical imbalance in the original sample had not led to major bias relating to educational attainment, employment status, or sector of occupation (ibid: 62f).

In the 1992 wave Hansen found skewness in the geographic distribution of response rates, and although the response rate was generally high, it was somewhat lower in Copenhagen, the surrounding suburbs and on Funen (ibid: 73). Furthermore, he found differences in response rates according to the parents' socioeconomic position and the respondent's verbal test score and educational attainment. In general, the response rate was lower among respondents from less advantaged family backgrounds (defined as fathers belonging to social class V in 1968, i.e., manual working class), among those with a low score on the 1968 verbal test, and among those who had not completed any post-compulsory education after the 7th grade in the 1976 wave (ibid: 79ff).

References

- Hansen, Erik Jørgen 1968. *De 14-20-åriges uddannelsessituation 1965, bind I + II* [The state of education among 14-20-year-olds]. Copenhagen: The Danish National Centre for Social Research, Publication 31.
- Hansen, Erik Jørgen 1995. *En generation blev voksen* [A generation grew up]. Copenhagen: The Danish National Centre for Social Research, Report 95:08.
- Jæger, Mads Meier, Ploug, Niels, and Munk, Martin D. 2003. *Ulighed og livsløb. Analyser af betydningen af social baggrund* [Inequality and the life course. Analyses of the impact of family background]. Copenhagen: The Danish National Centre for Social Research, Report 03:10 (downloadable from www.sfi.dk).
- Ørum, Bente 1971. *Social baggrund, intellektuelt niveau og placering i skolesystemet* [Family background, intellectual level, and position in the schooling system]. Copenhagen: The Danish National Centre for Social Research, Publication 59.