UK Data Service

National Survey of Sexual Attitudes and Lifestyles, 2010-2012:
Open Teaching Dataset

User Guide

Author: Jennifer Buckley and Alle Bloom

Version: 1.0

Date: February 2020

Contents

Introduction

The [British National Surveys of Sexual Attitudes and Lifestyles (Natsal)](#) are some of the largest and most detailed studies of sexual attitudes and behaviour in the world.

This open teaching dataset contains data from Natsal-3, which interviewed 15,162 adults aged 16-74 in 2010-2012. The data all comes from the [original study accessible via the UK Data Service](#). To make the data accessible for teaching via an open data license we have selected a subset of variables and cases. We have also recoded some demographic variables and created a new continuous variable measuring attitudes towards sexual behaviours.

This user guide provides

- a background to Natsal including details of the survey design.
- details of this open teaching dataset including a variable list, syntax used to create the dataset and full codebook.

Background to Natsal

The National Surveys of Sexual Attitudes and Lifestyles ran first in 1990-1991. Natsal originated in the mid-1980s as a response to the HIV epidemic and the need for population-based data to understand transmission. Further studies have taken place every 10 years since with Natsal-2 in 1999-2001 and Natsal-3 in 2010-2012. Natsal-4 was under development at the time of writing.

Data from these studies have improved our understanding of sexual behaviour and provided evidence to inform policy and practice. [Initial findings from Natsal-3 are available to read online via the Lancet](#).

Survey design

Natsal-3 involved face-to-face interviews, using Computer-Assisted Personal Interviewing (CAPI), and a self-completion part, using Computer-Assisted Self Interviewing (CASI).

The face-to-face interview included questions on general health, family when growing up, learning about sex, first sexual experience, contraception, periods and menopause, attitudes to different kinds of relationships and sociodemographic details including cohabitation history.

More sensitive questions came in the self-completion part to reassure respondents of confidentiality and to avoid embarrassment. Topics covered in this part include questions about sexual histories and recreational drug use.

Not all respondents received the self-completion questions, it depended on how they answered previous questions (survey routing). Those with no sexual experience did not get the CASI questionnaire, while those with some sexual experiences got a shortened version. This design feature means that some cases are recorded as 'not applicable' for variables coming from the CASI part of the survey. Overall, 1.9% of respondents were not eligible for this module at all, and of the 98.1% that were eligible for any form of the CASI, 3.9% were given the shortened version and 2.2% refused to complete any of the CASI.

Response rate

The sample design for Natsal-3 required that one randomly selected adult aged 16-74 (16-34 in the boost samples) be interviewed in each eligible sampled address. In total 59,412 addresses were sampled, 24924 in the core sample and 18,537 and 15,951 in the boost samples.

Once removing those who were ineligible (based on building type or being out of the age range), those who's eligibility was unknown (through lack of contact or refusing information) and the estimated ineligibles, 26,274 estimated eligible addresses remained.

In total, interviews were completed with 15, 162 respondents.

Using AAPOR's formula for calculating RR3 (Response rate 3) NATSAL-3 achieved an overall response rate of 57.7%.

Overview of the Natsal teaching dataset

Sample

The open dataset contains a random subsample of cases from the original sample. The subsample was 25 percent, which is 3,799 cases.

Topics

The open dataset contains variables covering the following topics:

- attitudes towards sexual lifestyles and behaviours such as adultery, same sex relationships and sex in the media
- health/mental health and disability
- Religion and religious beliefs
- relationships status
- religious beliefs
- and many demographic variables

A full list of variables can be found below.

Additional variables

This teaching dataset contains additional variables created from the original variables:

Sexual conservatism

The variable *attconservative* measures sexual attitudes.

It has been derived using a statistical method called principal component analysis (PCA), which can be used to analyse patterns of responses to multiple questions. One benefit of methods such as PCA is that they can help measure complex concepts that can be hard to measure with a single survey question.

In this case, we used 7 questions measuring opinions relating to sexual behaviours and lifestyles, these were:

- What is your opinion about a person having one night stands? Right/wrong
- The rest of the questions asked respondents to say whether they agreed strongly, agreed, neither agreed nor disagreed, disagreed, disagreed strongly with the following statements:
- It's OK to have sex with someone without being in love with them
- People are under a lot of pressure to have sex nowadays
- It is natural for people to want sex less as they get older
- Men have a naturally higher sex drive than women
- There's too much sex in the media these days
- Young people today start having sex too early

Using PCA, we derived a variable where a higher response indicates a more conservative attitude towards sexual behaviour in general. The new variable is a continuous variable (with a distribution close to normal) and can be examined using techniques such as correlation and regression.

The SPSS syntax used to create the variable is in the syntax section below.

Age at birth of first child
The variable Dage1ch derives from age1ch (the original age at birth of first child variable in the dataset) but with top and bottom coded values removed. Variables that are top-coded group all observations with values above a set upper bound into one category. In age1ch, this upper bound is anyone who had their first child over the age of 40. Bottom-coded variables groups all observations with values below a lower limit. In age1ch the lower bound is anyone who had their first child under the age of 15.

Creating a new version without the top and bottom coding is designed to simplify analysis of the variable using methods such as OLS linear regression. The original variable age1ch remains in the dataset.

Ethnicity
The variable ethnicgrpr is a recoded version of the ethnicgrp variable, reduced into two categories – 'white' and 'not white'. This has been done to reduce the risk of disclosure.

Sexual orientation
The variable sexidr is a recoded version of the sexed variable, reduced into two categories – 'heterosexual/straight' and 'not heterosexual/straight'. This has been done to reduce the risk of disclosure.

Missing Values

Don't know and refusal

When asked a question, survey participants may respond 'do not know' or refuse to answer. Such responses are recorded using specific codes, often something distinctive compared to other values such 99 or 999.

To help with analysis of this dataset, most 'not answered' and 'don't know' values are pre-set as missing values in the SPSS and Stata versions. The main exception is the sexual attitudes variables where the category labelled 'Depends/Don't Know' is kept as a valid value.

Not applicable

Many variables in the dataset have a category of not applicable, generally coded as -1. The not applicable category applies when a specific case is not given a value, usually because they were not asked the question. For instance, the variable age1ch records the age of respondent at the birth of their 1st child; naturally, this variable is only applicable to those who have had children.

Additionally, not applicable categories apply for questions asked within the more sensitive self-completion part of the survey, which not all participants received (see section on survey design above for more details).

Survey methodology variables

Survey weight

Many survey datasets contain variables called survey weights. These variables are made by the data collectors for you to apply when analysing data. We use weights to make sample data better represent the population it's designed to reflect by adjusting for over and under-represented cases. Under and over-representation can result from the complex sampling methods used in large scale surveys and other issues including non-response.

In this teaching dataset, the survey weight variable **Total_wt** is for use when analysing the total sample (including the boosts of younger people). The weight accounts for unequal selection probabilities from the sample design (selection weight) and for non-response by adjusting the distribution of age, sex and regional profiles to match the general population (poststratification weight).

For more information about the weight variable, see the Natsal-3 documentation.

Variables in the dataset

The open teaching dataset contains 20 of the original 1224 variables in the dataset.

The survey questions can be found in the [questionnaire accessible with the main study data](#).

The documentation that accompanies this open dataset includes a codebook.

Variable list

Variable Name	Variable Label
agrp	Respondent's age at interview, years, grouped (6 categories)
rsex	Respondent's sex
ethnicgrpr	Ethnic group (recoded)
sexidr	Sexual identity (recoded)
rnssecgp_6	Respondent's NSSEC code (8 groups)
adj_imd_quintile	Quintile of adjusted IMD (For Great Britain)
rwcasual	Opinion about one night stands? right/wrong.
snnolov	Sex without love OK
snpres	People are under pressure to have sex
snold	Natural for people to want sex less as they get older
snsexdrv	Men have a naturally higher sex drive than women
snmedia	Too much sex in the media
snearly	Young people today start having sex too early
attconservative	Sexually conservative attitudes scale, higher score = more conservative attitudes
dage1ch	Age of respondent at birth of 1st child (no top or bottom coding)
disabil2	Longstanding illness, disability or infirmity
depscr	Screen positive for current depression (based on PHQ-2, mood)
religimp	Importance of religion and religious beliefs now
relstatr	Relationship status at the time of the interview (recoded)
total_wt	Total weight for Natsal 2010(core and boost samples)

Syntax used to create the dataset

```
**PCA to create a factor for sexually conservative attitudes.
missing values snnolov snpres snold snsexdrv snmedia snearly snsexed (8, 9).
FACTOR
   /VARIABLES rwcasual snnolov snpres snold
snsexdrv snmedia snearly
 /MISSING LISTWISE
 /ANALYSIS rwcasual snnolov snpres snold
snsexdrv snmedia snearly
 /PRINT INITIAL EXTRACTION FSCORE
 /PLOT EIGEN
 /CRITERIA MINEIGEN(1) ITERATE(25)
 /EXTRACTION PC
 /ROTATION NOROTATE
   /SAVE REG(ALL)
 /METHOD=CORRELATION.

*rename generated factor, only using the first one here (though others interesting too).
COMPUTE attconservative = (FAC1_1)*-1.
VARIABLE LABELS attconservative 'Sexually conservative attitudes scale, higher score = more conservative
attitudes'.
Execute.
DELETE VARIABLES FAC2_1 FAC3_1.

*Create another version of the age of first child variable without the top or bottom coding.
RECODE age1ch (14=-1) (41=-1) (ELSE=copy) into dage1ch.
VARIABLE LABELS dage1ch  'Age of respondent at birth of 1st child (no top or bottom coding)'.
VALUE LABELS dage1ch
-1 'Not applicable including top coded'
99 'Not answered'.

*Recoding variables due to disclosure.
**recoding the ethnic group variable.
RECODE ethnicgrp (1=1) (2=2) (3=2) (4=2) (5=2) (6=2) (9=9) INTO ethnicgrpr.
VARIABLE LABELS ethnicgrpr 'ethnic group (recoded)'.
VALUE LABELS ethnicgrpr
1 'White'
2 'Not white'
9 'Not answered'.
EXECUTE.

**recoding the sexual identity variable.
RECODE sexid (1=1) (2=2) (3=2) (4=2) (9=9) INTO sexidr.
VARIABLE LABELS  rsexid 'Sexual Identity (recoded) '.

VALUE LABELS sexidr
1 'Heterosexual/straight'
2 'Not heterosexual/straight'
9 'Not answered'.
EXECUTE.

*rename weight variable.
Variable labels total_wt 'Statistical weight'.
```

EXECUTE.

**recoding the relationship status variable.
RECODE relstat (1=1) (2=1) (3=2) (4=3) (5=4) (6=5) (7=6) (8=7) (9=9) into relstatr.

VARIABLE LABELS relstatr 'Relationship status at the time of the interview (recoded)'.
VALUE LABELS relstatr
1 'Married/In a civil partnership and living with spouse/partner'
2 'Living with partner'
3 'In a steady & ongoing relationship but not living with a partner in last year'
4 'Widowed and not in a steady & ongoing relationship in last year'
5 'Divorced/Separated and not in a steady & ongoing relationship in last year'
6 'Previously lived with partner as a non-married couple but not now in a steady & ongoing relationship in last year'
7 'Never married, or lived with partner and not in a steady & ongoing relationship in last year'
9 'Not answered'.
*Set missing values.
missing values ethnicgrp disabil2(9).
missing values depscr (-1, 9).
missing values rsex (-1, -8, -9).
missing values rnssecgp_6(99).
missing values dage1ch (-1, 99).
missing values rwcasual snnolov snpres snold snsexdrv snmedia snearly relstatr religimp (9).

*Adjust varable levels.
VARIABLE LEVEL ethnicgrp disabil2 depscr relstat rnssecgp_6 (NOMINAL).

*Save and order desired variables into a new file.
SAVE OUTFILE = 'natsal_3_teaching_open.sav'
 /Keep = agrp rsex ethnicgrpr sexidr rnssecgp_6 adj_imd_quintile rwcasual
 snnolov snpres snold snsexdrv snmedia snearly attconservative
 dage1ch disabil2 depscr religimp relstatr total_wt.