

Millennium Cohort Study Fertility Survey:

A Guide to the cleaning of the SPSS Dataset

November 2006

Edited by
Maggie Hancock, Jon Johnson,

Centre for Longitudinal Studies

Bedford Group for Lifecourse & Statistical Studies

Institute of Education, University of London



In collaboration with



**NPEU, University of Oxford
Old Road Campus, Oxford OX3 7LF**



Table of Contents

Introduction.....	2
Background.....	2
Methods.....	2
Summary of data cleaning	2
Data cleaning details by variable	3
Further information.....	11
Comments	11

Introduction

This document has been prepared to accompany the initial deposit, with the UK Data Archive at the University of Essex, of data from the Millennium Cohort Study (MCS) Fertility Study. There were 236 study participants.

This note explains the data cleaning and also where data has been grouped to reduce disclosure.

There are three related documents that should be read in conjunction with this note

- MCS Fertility Report which details the background to the survey;
- MCS Fertility Survey Questionnaire;
- MCS Fertility Survey Technical Report on Response

Available with this deposit, or from the CLS Website (<http://www.cls.ioe.ac.uk/mcs>)

Background

It is estimated that infertility affects one in seven couples in the UK. A large and growing number of babies are conceived through infertility treatment. Despite this, data on the characteristics and outcomes of such pregnancies are relatively scarce. The aims of this study are to describe the characteristics of women who delivered successfully following infertility treatment, their experiences in seeking treatment, and the outcomes for the child and the mother in a nationally representative sample of children born at the start of the 21st century.

Methods

The study participants were drawn from the Millennium Cohort Study (MCS), a nationally representative study of 18,819 infants born in the UK in 2000-1. Infants who were alive and living in the UK at age 9 months were drawn from child benefit registers. Disproportionately stratified sampling at electoral ward level ensured adequate representation of disadvantaged and ethnic minority areas. Parents were interviewed for the first time (sweep 1) when the infant was aged about 9-10 months, and detailed information was collected on a wide range of socio-economic and health factors. Consenting parents who reported having had any medical infertility treatment for this pregnancy were sent a postal questionnaire which asked more detail about their experiences of infertility treatment. Much of the statistical analysis involved comparing the women who had received different types of infertility treatment, either with each other or with women who had received no such treatment. All analyses allowed for the clustered, stratified sample design, with re-weighting where necessary to allow for the different sampling proportions.

Of the 18,482 MCS mothers included in the analysis, 460 (2.5%) reported that infertility treatment had led to the conception of the MCS baby. 230 women (50%) responded to the postal questionnaire. Among the 460 MCS mothers who reported having had infertility treatment, 203 (44%) had clomid only, 207 (45%) had ART and 50 (11%) had surgery. Further results will be made available once they are accepted for publication.

Summary of data cleaning

All variables have -2 assigned as a missing value.

Where flag variables (e.g. Q4) have been coded with their position value when questionnaire item ticked, these values have been recoded to 1.

Where no response to a question is expected depending on the value entered for a previous item, values of -2 (not completed) are recoded to -1 (not applicable). -1 is assigned as a missing value. Where no response is expected but a value is entered that indicates no response, these values are recoded to -1 (e.g. Q16).

Cost and mileage variables have been recoded and new value labels specified (e.g. Q32yes).

Other text answers for the following questions have not been saved to the data file due to disclosure issues.

Q27 'Details of private fertility treatment'

Q29 'Details of NHS care'

Q30 'Fertility drugs paid for'

Q31 'Drugs provided free by clinic or GP'

Q38 'Names of the clinics attended';

Q39 'Why went to more than one clinic'

Q50 'How disruptive was receiving fertility treatment'

Q51 'In what ways could experience have been improved'

Q52 'What would you like to tell policy makers in NHS or gov.'

Notes on back page

Data cleaning details by variable

Q4 'Did the GP advise or prescribe any treatment for you?'

q4_1 Did GP advise or prescribe treatment-No

q4_2 Did GP advise or prescribe treatment-Yes

q4_3 Did GP advise or prescribe treatment-Smoking

q4_4 Did GP advise or prescribe treatment-Drinking

q4_5 Did GP advise or prescribe treatment-Weight loss

q4_6 Did GP advise or prescribe treatment-Other

The q4 fields were originally coded with their position value if ticked (i.e. 2 if q4_2 ticked, 3 if q4_3 ticked).

All 6 q4_n recoded as follows:

-2 for not completed or 1 where advice or treatment indicated.

-2 specified as a missing value.

The q4_details (details of other treatment) is not saved with the data file (stored on a separate text file) due to disclosure issues.

Q7 'How many months between first visit to GP and referral'

No response to Q7 is expected if q5 (did GP refer you to a specialist clinic) has a value of 2 (no). Variable q7 has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q5=2.

4 cases responded 'no' (2) to q5 but have values for q7. These values have been left as entered.

Q8 How long did you wait for first visit to fertility clinic (months)

No response to Q8 is expected if q5 (did GP refer you to a specialist clinic) has a value of 2 (no). Variable q8 has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q5=2.

4 cases responded 'no' (2) to q5 but have values for q8. These values have been left as entered.

Q10 Who had tests and were they prescribed by GP or fertility clinic

q10who 'Who had tests'
q10gp1 'GP prescribed tests for partner and myself'
q10gp2 'GP prescribed tests for me'
q10gp3 'GP prescribed tests for partner'
q10fc1 'Fert. Clinic prescribed tests for partner and myself'
q10fc2 'Fert. Clinic prescribed tests for me'
q10fc3 'Fert. Clinic prescribed tests for partner'

Many respondents seemed to be confused when completing this question and have not always been consistent when completing the 'Who had test' and 'GP prescribed' and 'Fertility clinic prescribed' columns (with ticks on different rows for each column).

Two new derived variables have been added to the data file: q10gp and q10fc each set to:
1 where any of the GP prescribed or Fertility Clinic prescribed fields respectively have been ticked
-2 (not completed) where none have been ticked and q9 (did you or your partner have tests..) is set to 1 (yes)
-1 (not applicable) where none have been ticked and q9 (did you or your partner have tests...) is set to 2 (no) or 3 (unsure).

Variables q10who, q10gp1-3 and q10fc1-3, q10gp, q10fc recoded so that values of -2 (not completed) are recoded to -1 (not applicable) if q9 (did you or your partner have tests.....) is set to 2 (no) or 3 (unsure).

3 cases responded 'no' or 'unsure' to q9 but entered values for q10who. 2 cases responded 'no' or 'unsure' to q9 but ticked one of the q10GP (prescribed by GP) fields and 1 case responded 'no' or 'unsure' to q9 but ticked one of the q10fc (prescribed by fertility clinic) fields. These values left as entered.

Q11 'Did you and partner have tests at same referral?'

q11 has been recoded so that values of -2

No response to Q11 is expected if q9 (did you or your partner have tests to find out why difficulty getting pregnant) has a value of 2 (no) or 3 (unsure) or q10who (who had tests) has a value of 2 (only I had tests) or 3 (only my partner had tests). Variable q11 has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q9=2 or 3 or q10 =2 or 3.

3 cases responded 'no' to q9 but have values for q11. These values have been left as entered.
12 cases responded 2 (only I) or 3 (only partner) to q10 but entered values for q11. These values have been left entered.

Q12 'What did tests show as the likely cause of difficulty in getting pregnant?'

q12_1 'Test showed likely cause as - sperm '
q12_2 'Test showed likely cause as - fallopian '

q12_3 'Test showed likely cause as - ovulation '
q12_4 'Test showed likely cause as - endometriosis '
q12_5 'Test showed likely cause as - No reason '
q12_6 'Test showed likely cause as - other '

The q12 fields were originally coded with their position value if ticked (i.e. 2 if q12_2 ticked, 3 if q12_3 ticked).

-2 for not completed or 1 where advice or treatment indicated.

No response to Q12 is expected if q9 (did GP refer you to a specialist clinic) has a value of 2 (no) or 3 (unsure). Variable q12 has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q9=2 or 3.

3 cases responded 'no' (2) or 'unsure' (3) to q9 but have values for q12_3 (ovulation). These values have been left as entered.

1 case responded 'no' (2) or 'unsure' (3) to q9 but ticked box for q12_4 (endometriosis). These values have been left as entered.

Q16 Drugs received

No response to Q16 is expected if q15 (did you receive any fertility drugs to help you conceive in this pregnancy) has a value of 2 (no) or 3 (unsure). Variable q16 has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q15=2 or 3.

For each of the q16 variables a value of no (2) was entered for 6 cases who also responded 'no' (2) or 'unsure' (3) to question 15. All these values of no have been recoded to -1 (not applicable).

Q18 'Year treatment started'

No response to Q18 is expected if q17 (Have you ever had other fertility treatment in the past..) has a value of 2 (no). Variable q18 has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q17=2.

3 cases responded no to Q17 and entered a value in Q18 (all 2000 - which may possibly relate to the cohort child/children). These values have been left as entered.

Q19 'Fertility treatments received in the past'

q19clom Which treatments you received in past -Clomid
q19ov Which treatments you received in past -Ovulation
q19dia Which treatments you received in past -Diathermy
q19iui Which treatments you received in past -IUI
q19di Which treatments you received in past -DI
q19gift Which treatments you received in past -GIFT
q19ivf Which treatments you received in past -IVF
q19icsi Which treatments you received in past -ICSI
q19froz Which treatments you received in past -Frozen embryo
q19spec Which treatments you received in past -Specific treatment
q19oth Which treatments you received in past -Other

No response to Q19 is expected if q17 (have you ever had fertility treatment in the past...) has a value of 2 (no). Variables q19xxx have been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q17=2. #

For each of the q19 variables there is one response of 2 (no) where q17 = 2 (no). These values have All been recoded to -1 (Not applicable).
For each of the variables q19clom, q19ov And q19dia, there are two responses of 1 (yes) where q17 = 2(no). These values are left as entered.

Q20 'Have you received fertility drugs in past not related to CM birth '

No response to Q20 is expected if q17 (have you ever had fertility treatment in thepast...) has a value of 2 (no). Variables q19xxx have been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q17=2.

10 cases responded 'no' (2) to q17 but have values of 2 (no) for q20. These values have All been recoded To -1 (Not applicable). 2 cases responded 'no' (2) to q17 but have values of 1 (yes) for q20. These values have been left as entered.

Q21 'Drugs received and number of cycles

q21hmg Drug received - HMG
q21hmgno HMG cycles
q21fsh Drug received - FSH
q21fshno FSH cycles
q21rfsh Drug received - RFSH
q21rfshno RFSH cycles
q21hcg Drug received - HCG
q21hcgno HCG cycles
q21down Drug received - DRG
q21downno DRG cycles

No response to Q21 is expected if q17 (have you ever had fertility treatment in the past...) has a value of 2 (no). or q20 (Have you been prescribed fertility drugs In the past) has a value of 2 (no) or 3 (unsure).

Variables q21xxx have been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q17=2 or q20=2 or 3.

For each of the q21 variables there is one response of 2 (no) where q20 Is Set To 2 (no) or 3 (unsure). These values have all been recoded To -1 (Not applicable).

For each of the variables q21hmy, q21fsh, q21rfsh and q21hcg there is 1 responses of 1 (yes) where q17 = 2 (no). These values have been left as entered.

Q22 'Did earlier treatments lead to pregnancy '

No response to Q22 is expected if q17 (Have you ever had other fertility treatment In the past) has a value of 2 (no). Variable q22 has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q17=2.

7 cases responded 'no' (2) to q17 but have values of 2 (no) for q22. These values have All been recoded To -1 (Not applicable).

1 case responded 'yes' (2) to q17 but have values of 2 (no) for q22. These values have been left as entered.

Q23 'How many babies did you conceive '

No response to Q23 is expected if q17 (Have you ever had other fertility treatment In the past) has a value of 2 (no) or if q22 (Did these earlier treatment lead to a pregnancy) has a value of 2

(no) . Variable q23 has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q17=2 or q22=2.

1 case responded 'no' (2) to q17 but has a value of 1 (single baby) for q23. This value have been left as entered.

Q24 'Did pregnancy lead to live births '

No response to Q24 is expected if q17 (Have you ever had other fertility treatment In the past) has a value of 2 (no) or if

q22 (Did these earlier treatment lead to a pregnancy) has a value of 2 (no) .

Variable q24 has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q17=2 or q22=2.

1 case responded 'no' (2) to q17 but has a value of 2 (no) for q24. This values have been left as entered.

2 cases responded 'no' (2) to q22 but have values of 2 (no) for q24. These values have been left as entered.

Q25 'How many babies did you have'

No response to Q25 is expected if q17 (Have you ever had other fertility treatment In the past) has a value of 2 (no) or if q22 (Did these earlier treatment lead to a pregnancy) has a value of 2 (no) . Variable q25 has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q17=2 or q22=2.

1 case responded 'no' (2) to q17 but has a value of 1 (single) for q25. This values have been left as entered.

Q32yes 'Amount paid for drugs'

Variable q32yes recoded (1 thru 50=1)(51 thru 100=2) (101 thru 1000=3) (1000 thru highest=4).

Q32no 'How much did you pay for all care'

Variable q32no recoded (1 thru 100=1)(101 thru 1000=2)(1001 thru 5000=3)(5001 thru highest=4)

Q33 Who paid for treatment

qd33clom Clomid - Who paid
qd33clcst Clomid Cost
qd33ov Ovulation Induction - Who paid
qd33ovcst Ovulation InVction Cost
qd33dia Diathermy - Who paid
qd33dacst Diathermy Cost
qd33iui IUI - Who paid
qd33iucst IUI Cost
qd33di DI - Who paid
qd33dicst DI Cost
qd33gift GIFT - Who paid
qd33gicst GIFT Cost
qd33ivf IVF - Who paid
qd33ivcst IVF Cost

qd33icsi ICSI - Who paid
qd33iccst ICSI Cost
qd33froz Frozen Embryo - Who paid
qd33frcst Frozen Embryo Cost
qd33andro Andrology Test - Who paid
qd33ancst Andrology Test Cost
qd33part Treatment for Partner - Who paid
qd33pacst Treatment for Partner Cost

qd33xxx who paid variables categorised as follows:

- 1 Subject/ subject and partner/ partner/ family
- 2 GP/NHS/Clinic
- 3 Subject/partner & GP/NHS
- 4 Insurance/ BUPA
- 5 Military
- 6 Other (e.g. part of programme)

There could be some confusion over whether when responding that they had paid themselves, this could apply to the prescription cost only. Where prescription is mentioned NHS has been coded.

Where qd33xxx (who paid) is blank qd33xxx is set to -2. -2 set to a missing value.

All qd33xxxcost (cost) variables are recoded
(0=0)(1 thru 100=1) (101 thru 500=2)(501 thru 1000=3)(1001 thru 5000=4)(5001 thru Highest=5)
Note that recodes are the same for all qd33xxxcost field but range of costs vary.

For all qd33xxx and qd33xxxcost (who paid and cost) variables, -2 is set as a missing value.

Q37 'How many clinics did you go to for treatment'.

No response to Q37 is expected if q36 (have you received treatment at more than one clinic to try to get pregnant) has a value of 2 (no I only attended one fertility clinic) or 3 (No, I never went to a fertility clinic_). Variable q37 has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q36=2 or 3.

17 cases responded 'only received treatment at one clinic' (2) to q36 but have values of '1' for q7. These values have been recoded as -1.

2 cases responded 'received treatment at more than one clinic (1) to q36 but have values of '1' for q7. These values have been left as entered.

Q40 'How many appointment at fertility clinics in total

No response to Q40 is expected if q36 (have you received treatment at more than one clinic to try to get pregnant) has a value of 3 (No, I never went to a fertility clinic). Variable q40 has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q36=3.

Q41 'Miles travelled to most recent clinic appointment – yourself'

Variable qd41miles recoded (0=0)(1 thru 3=1) (4 thru 10=2) (11 thru 50=3) (51 thru Highest=4)

No response to qd41miles is expected if q36 (have you received treatment at more than one clinic to try to get pregnant) has a value of 3 (No , never went to a fertility clinic). Variable

q41miles has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q36=3.

Q41part 'Miles travelled to most recent clinic appointment – partner'

Variable qd41part recoded (0=0)(1 thru 3=1) (4 thru 10=2) (11 thru 50=3) (51 thru Highest=4)

No response to qd41part is expected if q36 (have you received treatment at more than one clinic to try to get pregnant) has a value of 3 (No , never went to a fertility clinic). Variable qd41part has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q36=3.

Q42 'How did you travel to the clinic?'

Flags for modes of transport set up as follows:
qd42walk How did you travel to clinic- walked
qd42car How did you travel to clinic- car
qd42train How did you travel to clinic- train/ tube
qd42bus How did you travel to clinic- bus
qd42ferry How did you travel to clinic- ferry
qd42plane How did you travel to clinic- plane

If the mode of transport is mentioned in q42, flag variable set to 1 else =set to -2(not completed)

No response to Q42 is expected if expected if q36 (have you received treatment at more than one clinic to try to get pregnant) has a value of 3 (No , never went to a fertility clinic). Variable qd42xxx has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q36=3.

1 case responded 'never went to fertility clinic' (3) to q36 but indicated 'car' as response to q42. This value left as entered.

Q43 'Total cost of most recent visit to clinic'

Variable qd43 recoded
(1 thru 10=1)(11 thru 20=2) (21 thru 100=3)(101 thru 1000=4) (1001 thru Highest=5)

No response to qd43 is expected if expected if q36 (Have you received treatment at more than one clinic to try to get pregnant?) has a value of 3 (No , never went to a fertility clinic). Variable qd43 has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q36=3.

Q45 'How many days taken off work to receive fertility treatment'

Variable q45 No response to Q45 is expected if q44 (need to take time off work for fertility treatment) has a value of 2 (n/a as not working), 3 (did not need to take time off) or 4 (not able to take time off). Variable q45 has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q44=2, 3 or 4.

3 cases responded 'did not need to take time off' (3) to q44 but have value of 0 for q45. These values have been recoded to -1.

Q46 'How time off was arranged'

Two flags had been previously set up q46paid (allowed time off work) and q46unpai (took unpaid or annual leave) – set to 1 if that value ticked on questionnaire.

No response to Q46 paid or Q46 unpai is expected if q44 (need to take time off work for fertility treatment) has a value of 2 (n/a as not working), 3 (did not need to take time off) or 4 (not able to take time off). Variables q46paid and q46unpai have been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q44=2, 3 or 4.

Q48 'How many days did partner take off work to receive fertility treatment'

No response to Q48 is expected if q47 (partner needed to take time off work for fertility treatment) has a value of 2 (n/a as not working), 3 (did not need to take time off) or 4 (not able to take time off). Variable q48 has been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q47=2, 3 or 4.

6 cases responded 'partner did not need to take time off' (3) to q47 but have value of 0 for q48. These values have been recoded to -1.

Q49 'How partner time off was arranged'

Two flags had been previously set up q49paid (allowed time off work) and q49unpai (took unpaid or annual leave) – set to 1 if that value ticked on questionnaire.

No response to Q49 paid or Q49 unpai is expected if q47 (need to take time off work for fertility treatment) has a value of 2 (n/a as not working), 3 (did not need to take time off) or 4 (not able to take time off). Variables q49paid and q49unpai have been recoded so that values of -2 (not completed) are recoded to -1 (not applicable) where q47=2, 3 or 4.

2 cases responded 'partner did not need to take time off' (3) to q47 but have value of 2 (partner took unpaid time off) 0 for q49. These values left as entered.

Further information

Further information on MCS is available from the CLS website (<http://www.cls.ioe.ac.uk/mcs>). CLS can also be contacted at the following email address: clsfeedback@ioe.ac.uk

Comments

We would value your comments on this document, or on any aspect of MCS. Please email: clsfeedback@ioe.ac.uk.