Research Report for RES-000-22-1545:


Background

This project developed a relatively underused aspect of the National Child Development Study (NCDS) and the British Cohort Study (BCS70) data sets, that of the data on twins. It attempted to develop better measures of zygosity (whether the twins were identical or not) as the previously available data was collected at birth as recorded by the medics, who most likely but not definitely based this decision on the number of placenta which can substantially underestimate the number of identical twins (Bryan 1992). Previous work using the twins of the cohort studies had either not required the data on the zygosity of twins or had been based on the unreliable available data (Annett 1987, Blanchflower and Elias 1999, Emanuel et al 1992)

As well as hoping to enhance the quality of the zygosity of the twins, this project proposed a hypothesis which could be used to verify the usefulness of the data. The hypothesis was, as suggested in the title, to test the relative importance of nature, nurture and peer effects on adult outcomes, focusing on education and labour market success. This hypothesis was proposed in response to two high profile publications such as Herrnstein and Murray (1996) which suggested educational success was mostly genetic and Harris (1999) which suggested educational success was mostly peer related. It was hoped that the data on twins from the cohort study would be able to shed some light on this topic.

Finally the additional benefit of the cohort studies compared to other twin data sources previously used by the PI was the availability of a range of peer measures available on the cohort studies (Hawkes 2003). Whilst data on peers could have been collected from twins of the St. Thomas’ Twin Register these would have been subject to recall bias as the collection of the data would have occurred now the twins are in adulthood. The longitudinal nature of the cohort studies means that data collected on
peers was collected at various points in childhood and therefore not subject to recall bias.

Objectives

In the original proposal the aims and objectives where:

4. To provide a complete and accurate data set which identifies the zygosity of the twins on the NCDS and the BCS70

5. To use the data held on the twins in the NCDS and BCS70 to consider the relative importance of nature, nurture and peer effects in childhood on adult outcomes such as education, employment and earnings

6. To identify the relative importance of nature, nurture and peer effects on each adult outcome across the two cohorts and within the cohorts across time in order to assess whether there are differences between cohorts over time and over age within cohorts.

For the first objective, to provide a complete and accurate data set identifying the twins on the cohort studies has been created and is in the process of being deposited at the data archive. This deposit will be completed once the internal team at CLS has had the opportunity to check the data. This is an important step as the quality of this small data set must conform with that of the cohort studies in general. Several academic users have contacted me to ask about the release of this data and this is likely to be a valuable addition to the cohort study data currently available through ESDS. The data note and data will be made available in due course through the ESDS and the availability of this data will be announced through the ESRC mini website and the KÖHORT magazine for users of the cohort study.

The second and third objectives have been written up into a paper which is currently in submission to several conferences and in due course will be submitted as a CLS working paper and an abridged version to an appropriate academic journal (Hawkes 2010). This paper will be uploaded to ESRC Today in due course. The focus on the education outcome was driven by the incomplete nature of the employment data from many of the twins at various points in time, especially for female twins. Missing data is a large problem with twin research as the estimated models rely on the comparison
between the twins for their identification. Therefore if one twin’s data is missing this means the loss of that pair of twins to the sample. Unlike the data on education, missing employment data could not be as easily derived from past and future employment data without introducing to the models a lot of measurement error. With such a small data set and the fact that twin models amplify the issues of measurement error, given the comparative nature of the models Griliches (1979), it was thought wisest to focus exclusively on the education variables.

Methods

1. Data Collection

As mentioned in the report above a lot of work had to be undertaken before the questionnaire could be sent out to confirm which twin was paired with which and which cohort members were actually twins. This is because the address database does not hold survey data on it for confidentially purposes. A twin pair indicator has now been added to allow the tracing team at CLS to use this information to contact missing twins. Once the sample was confirmed and identified and the questionnaire ethically approved (see below), it was possible to send out the questionnaire.

In the appendix below are copies of the questionnaire and covering letters sent out to the twins of the NCDS and BCS70. During the data collection period two reminders were sent out to those who had not responded in an attempt to optimise the response to this postal questionnaire. The response rates were good for a postal questionnaire with 78% for the NCDS and 65% for the BCS70. The lower BCS70 response rate is as expected given the age of the respondents as they are more likely to be mobile than the NCDS respondents, which is confirmed by the figures of those reported as gone away (1.61% or 3 cases for the NCDS compared to 7.05% or 11 cases for the BCS70).

The table below sets out the sample development. As we can see many more twins were originally on both cohort studies (366 NCDS, 316 BCS70) but due to high rates of infant mortality of twins, subsequent attrition and incomplete address information only 236 NCDS and 224 on the BCS70 individual twins could be mailed to. The
table then sets out the development of the sample before and after each reminder. As we can see from the first two rows contacting the twins in adulthood meant that only 64% of the NCDS and 70% of the BCS70 sample was contactable. Therefore this added to the desire to ask the parents of the twins on the Millennium Cohort Study about their zygosity earlier (at sweep 4, age 7), an additional benefit of this project.

Sample Returns for the NCDS/BCS70 Twin Survey

<table>
<thead>
<tr>
<th></th>
<th>NCDS Twin Sample</th>
<th>BCS70 Twin Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of twins available</td>
<td>366</td>
<td>316</td>
</tr>
<tr>
<td>Confirmed Addresses</td>
<td>236</td>
<td>224</td>
</tr>
<tr>
<td>(contacted sample)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of response</td>
<td>184</td>
<td>145</td>
</tr>
<tr>
<td>Responses before</td>
<td>143</td>
<td>110</td>
</tr>
<tr>
<td>reminders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional responses after 1st</td>
<td>34</td>
<td>27</td>
</tr>
<tr>
<td>reminder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional responses after 2nd</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>reminder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of twins with useable</td>
<td>242</td>
<td>202</td>
</tr>
<tr>
<td>zygosity</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The final row of the table highlights the number of twins for whom updated zygosity is available. This is larger than the number of responses because as long as one of the twin pair responses to the questionnaire the information supplied can help to generate the identification for both of the twins. Where both twins respond the information can help to countercheck the information supplied. The deposited data set will contain a range of twin identification variables including: three derived from the three distinct sets of questions on the questionnaire, a composite variable from these three variables and a complete variable which draws on the information from the birth variables where a questionnaire has not been returned.
2. Analysis

The analysis as outlined in the proposal used the methodology as proposed by DeFries and Fulker (1985) (henceforth called DFF). The twins are ranked by the dependent variable (in this case educational attainment). The one with the higher level of education is called the co-twin (i) while the weaker one is the proband twin (-i). These twins were raised within the same family (f). The following equation is estimated

\[ E_{if} = \alpha + \varphi E_{-if} + \lambda R_f + \pi X_{if} + \beta Y_f + \gamma f + \delta G_{if} + \varepsilon_{if} \]

where E is the education level of the twin i, R is the genetic relationship between the twins (1 for identical 0.5 for fraternal), X is a vector of variables affecting education, for example type of school, class size, peers, etc which are individual to twin i, Y is a vector of variables that twins 1 and 2 have in common that influence education, for example parents’ education, number of siblings, household income etc, G is their genetic make up and ε is an error term. Work by Miller et al (1996) and Hawkes (2003) confirmed that the results of the within-twin method favoured in economics are comparable to this behavioural genetics method which controls directly for genetic and family background effects. Therefore the DFF approach is sufficient. In addition the relative importance of nature and nurture on education can be assessed (controlling for peer effects) using the DFF model. Given the assumptions as set out in Behrman and Taubman (1976) the parameter \( \varphi \) provides a measure of common environment (\( c^2 \)) and the parameter \( \lambda \) provides a measure of heritability (\( h^2 \)). Therefore if \( \varphi > \lambda \) then common environment is more important than heritability.

Results

The analysis undertaken was based on the sample obtained from the questionnaire. At present work is being undertaken to see if this has caused any form of a selection bias in the twin sample. Of the 242 NCDS twins (121 pairs) for whom questionnaire data is available 23% were identical twins whilst of the 202 BCS70 twins (101 pairs) 31% were identical twins.
Depending on the educational attainment variable selected for analysis between 65% and 87% of twins actually have the same level of education. The variable with the most variability between the twins was chosen as the main dependent variable. In addition the educational variable chosen was selected as it was comparable across the two studies, NCDS and BCS70. The educational attainment variable selected that matched these criteria was the educational qualification held at age 33/34. Given the limited sample size only a handful of additional variables were included in the final regression models, although many different variables were explored but most were excluded as they did not improve the model fit. Those included in the final model, in addition to the variables discussed below, were parents’ social class at 10/11, parents’ interest in the child’s education as reported by the teacher at 10/11 and number of siblings in the household at 10/11.

### Summary Table of Results

<table>
<thead>
<tr>
<th></th>
<th>NCDS</th>
<th>BCS70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Environment ($c^2$)</td>
<td>0.67-0.88</td>
<td>0.72-0.89</td>
</tr>
<tr>
<td>Heritability ($h^2$)</td>
<td>0.05-0.12</td>
<td>0.02-0.14</td>
</tr>
<tr>
<td>Peers Effects</td>
<td>0.04-0.21</td>
<td>0.09-0.18</td>
</tr>
<tr>
<td>Sample Size in pairs</td>
<td>119</td>
<td>97</td>
</tr>
</tbody>
</table>

In all cases $c^2 > h^2$ suggesting a strong common environment effect on obtaining educational qualifications as gathered by age 33/34 for those born in the UK in 1958 and 1970. In most cases the heritability effect was insignificant. This suggests that for both cohorts the common environment is a stronger determinant of educational attainment at age 33/34 than heritability.

Peer effects were measured with two derived variables constructed with the variables in the table below. These two sets of variables were the child is in the top stream at school and if the child took part in group activities outside of school. These two variables were chosen because of the completeness of the data and the comparability.
across the two surveys. Other peer variables were included which were not comparable to try to find a peer effect, although no significant effect was found. Therefore the final models only included these two comparable derived variables. Both of these variables were measured at 11 years old and neither was statistically significant even at 10%.

**Selected Peer Effect Measures from the NCDS and BCS70**

<table>
<thead>
<tr>
<th>NCDS 2</th>
<th>Age 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Reported</td>
<td>Age group streamed by ability n861</td>
</tr>
<tr>
<td>Cohort Member Reported</td>
<td>Going to clubs outside school (inc Scouts and Guides) n938</td>
</tr>
<tr>
<td></td>
<td>Playing outdoor games or taking part in sports outside school hours n941</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BCS70 2</th>
<th>Age 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Reported</td>
<td>Is the child in a streamed class j055</td>
</tr>
<tr>
<td>Parent Reported</td>
<td>Plays sports in spare time m84</td>
</tr>
</tbody>
</table>

The overall conclusion of this work is that nurture the most important determinant of educational attainment at age 33/34 for those born in 1958 and 1970. This is a much more important effect than that of nature or peers providing a rebuttal for the UK data to the work by Herrnstein and Murray (1996) which proposed nature and Harris (1999) which proposed peers to be the key determinant of educational success. In addition there appears to be a very slight and marginally significant increase in the importance of the common environment between the two cohorts. No effect over time within a cohort was found.

This paper has been submitted to a range of conference for summer 2010 and will be submitted to an appropriate journal by the end of the summer.
Activities

Before any data could be collected it was necessary to draft and ask for peer comments on the questionnaire and covering letter. Members of the internal cohort team (Peter Shepherd, Jane Elliott, Lisa Calderwood, Denise Brown, Heather Joshi, John Bynner and Samantha Parsons), other Institute staff with extensive experience of the cohort studies (Ingrid Schoon and Eirini Flouri) and experienced twin researchers from outside of the Institute (Alison MacDonald a professional psychologist, Alison Macfarlane in the Department of Midwifery at City University who had used the cohort twin data before and Lynn Cherkas from the St. Thomas’ Twin Research Unit, Kings College London) provided their comments and suggestions on the questionnaire and covering letter. This invaluable peer review substantially improved the questionnaire and covering letter in terms of content and presentation. It also opened up conversations with potential interested users of the data. As already noted above, joint work on the twins of the cohort studies is planned with Eirini Flouri and Ingrid Schoon. In addition Alison Macfarlane has been in contact regarding the data from this project and its potential use.

A range of other activities have been used to seek peer comment on the results of the research and to promote the twin data of the cohort studies. Working on the NCDS/BCS70 workshops provided the opportunity to promote the twin data with potential NCDS/BCS70 users. The designing of an ESRC mini site has been invaluable to promoting the project and has lead to many enquiries from other academics regarding the project and access to the data. Discussions about the project with Ian Walker, a fellow twin researcher, were possible following his presentation at the Institute of Education Bedford Square Seminar Series on his own work using twin data.

In addition the Institute has promoted the project to external sources through its press office, although I have been cautious on discussing the results of the analysis of the data until appropriate peer review has been undertaken. Two press releases are arranged to coincide with the publication of the data and the publication of the results of the research respectively.
Beyond the informal, formal presentations have been undertaken at seminars at the Institute of Education and the University of Greenwich and an additional seminar at the University of Greenwich is plan for June 2010 in preparation for this summer’s conference season. The proposal also contained funds for one conference during the project. A sole authored paper entitled “Do we really need identical twins data? Considering the return to education using data on identical twins, fraternal twins and siblings for the UK.” was accepted and presented at the 2nd International Conference on Educational Economics organized by the Faculty of Economics at the National and Kapodistrian University of Athens, Greece in Athens, Greece on 27-30 August 2008. The paper outlined some results based on the old measure of zygosity for the cohort studies as well as some results based on data from the St. Twin Result Unit. It then highlighted this project and the need for better measures of twin identification in this type of work. The paper generated some debate and the conference participants provided many useful ideas for the paper.

**Outputs**

In the proposal there were six outputs outlined:

1. **Academic papers to be sent to prestigious peer review journals on the relative importance of nature, nurture and peer effects on adult outcomes such as education, employment and earnings would be produced.**

   Drafts of the paper considering the relative importance of nature, nurture and peer effects on educational achievement are in submission for conferences this summer (2010) and will be submitted to an appropriate journal after the conference period. There has been some delay in finishing this publication due to the maternity leave of the P.I.

2. **A data set consisting of the final identification, the zygosity scores and the data collected from the questionnaire would be made available to other researchers by depositing it at the data archive. This data set could also contain the peer effect derived variables.**
The data set is completed and in due course be deposited at the Data Archive once the internal CLS review process has been completed.

3. A data note on the zygosity scores which would provide details of how the twins were identified and the numbers of identical and fraternal twin pairs and singletons.

   The data note to accompany the dataset is also completed and will be deposited at the data archive at the same time as the data set following internal review at CLS. This internal review process is essential as the data contributes to the large publicly funded data sets of the NCDS and BCS70.

4. A summary briefing paper on the results to be put into the CLS series about findings. This will be put into a downloadable form and will be available from the CLS website.

   This will be completed following the summer conference session. The maternity leave spells of the PI have lead to some delay in the production of some outputs but these are now in hand and will be completed by the end of the summer (2010).

5. A press release of the findings when the above is available.

   There are two press releases planned: one to users on the publication of the data by the data archive and one more general press release on the publication of the briefing paper.

6. Presentations at an academic conference and external seminars in other institutions.

   These have been undertaken and the comments have been invaluable to the development of the project. This summer will be another round of presentations before submission to an appropriate journal, CLS working paper series and other outputs. The planned conference visits for last summer had to be delayed due to the maternity leave of the PI.
Impacts

During the course of the project many academic researchers have contacted me regarding the project. One in particular is very keen to have access to the data once it is published by the Data Archive, Nicola Pensiero who is a PhD Researcher at the European University Institute. Most have found out about the project from the ESRC mini web page that the PI has created. Therefore once the data is released by the Data Archive this will be announced on the project’s mini ESRC website as well as to cohort study users by the KOHORT magazine.

Future Research Priorities

A series of papers with the P.I., Eirini Flouri and Ingrid Schoon are planned and are in progress. These papers will look at the role of: aspirations, social background and parenting on educational attainment. In addition, a summary paper is proposed for an economics journal that would help to further the discussion about the usefulness of the within twin methodology in estimating the return to education.

1.7 Ethics

Before the questionnaire was able to enter the field ethical approval was sort from the NCDS/BCDS70 scientific committee and the Institute of Education research ethics committee. The approval from the NCDS/BCS70 scientific committee was essential as the twins surveyed are part of the cohort studies and therefore issues of respondent burden and appropriateness of the survey had to be considered. In addition as the project involved human participation by questionnaire the university research ethics committee had to review the project and materials to be sent into the field. Both committees provided invaluable feedback on the materials, these comments were incorporated into the documents sent out to the twins and approval was obtained.
Bibliography


Additional Materials: Annexes
Annex 1: Questionnaire

NCDS/BCS70 Twins Questionnaire

1. Are you one of a twin pair?
   - [ ] Yes ➔ Please answer question 2
   - [ ] Yes, but twin died at birth ➔ Please skip to the end of the questionnaire
   - [ ] No ➔ Please skip to the end of the questionnaire

2. Are you a same-sex twin pair?
   - [ ] Yes, both female ➔ Please answer question 3
   - [ ] Yes, both male ➔ Please answer question 3
   - [ ] No, one female, one male ➔ Please skip to question 12

3. Are you an identical twin pair?
   - [ ] Yes
   - [ ] No
   - [ ] Don’t know

4. At school, did teachers have trouble telling you apart?
   - [ ] Yes
   - [ ] No
   - [ ] Don’t know

5. Were your parents able to tell you apart when you were of school age?
   - [ ] Yes
   - [ ] No
   - [ ] Don’t know

6. Were your brothers/sisters able to tell you apart when you were of school age?
   - [ ] Yes
   - [ ] No
   - [ ] Don’t know
   - [ ] Did not have any brothers/sisters

7. Were your close friends able to tell you apart when you were of school age?
   - [ ] Yes

To cite this output:

REFERENCE No. RES-000-22-1545
8. Were strangers able to tell you apart when you were of school age?

☐ Yes
☐ No
☐ Don’t know

9. In childhood, which of the following would best describe you and your twin (please tick one only)?

☐ As alike as peas in a pod
☐ Of ordinary sibling likeness
☐ Don’t know

10. Do you have the same eye colour?

☐ Yes
☐ No
☐ Don’t know

11. Do you have similar height, weight and natural hair colour and texture?

Height ☐ Yes ☐ No ☐ Don’t Know
Weight ☐ Yes ☐ No ☐ Don’t Know
Natural Hair Colour ☐ Yes ☐ No ☐ Don’t Know
Hair Texture ☐ Yes ☐ No ☐ Don’t Know

12. Did you and your twin get on well with each other in childhood?

☐ Yes
☐ No
☐ Don’t know

13. Are you still in contact with your twin? If so how frequent is this contact?

☐ No
☐ Yes – daily
☐ Yes – a few times a week
☐ Yes – a few times a month
☐ Yes – less than a few times a month

Thank you for completing this questionnaire.
Please return it in the pre-paid envelope enclosed together with the updated address information for you and your twin.
Annex 2: Covering Letter for BCS70

Dear XXXXX

You may remember that in the past you have helped with the surveys of the British Cohort Study (BCS70) – an important ongoing research study of all those in Great Britain who were born in a single week in 1970. Your help with earlier BCS70 surveys was very important and much appreciated - it helped make the study a success.

We are writing because our records show that you are one of a twin pair and we would like to invite you to take part in a special survey of twins. As one of a twin pair you are a very special member of the study and we hope you will agree to help make our current survey of twins a success. Indeed, as there are only 200 sets of twins in the study, you are crucial to the project as we cannot replace or substitute another person’s information for yours.

All we would like you to do is to complete the enclosed questionnaire and return it in the envelope provided together with the updated address information for both you and your twin. The questionnaire should only take you around 15 minutes to complete. Your answers, when combined with information you have given during earlier surveys, will help us learn more about the impact of childhood experiences on employment and wages in adult life.

As a twin, you may have already taken part been in other projects which seek to gather information about your experiences but, as a very special part of the BCS70, I hope you will agree to help with this survey.

As always, the information that you give will be treated as confidential, and will not be reported in any way which enables you to be identified. We will provide a summary of what we find when the project is completed.

If you have any queries about the survey, please contact my colleague, Denise Hawkes – you can telephone 020 7612 6881, or email d.hawkes@ioe.ac.uk. Alternatively, you can call the usual freephone number - 0500 600 616 - or email bcs70@ioe.ac.uk.

I would like to thank you in advance for your time.
Yours sincerely,

Professor Heather Joshi
Annex 3: Covering Letter for NCDS

Dear XXXXX

You may remember that in the past you have helped with the surveys of the National Child Development Study (NCDS) – an important ongoing research study of all those in Great Britain who were born in a single week in 1958. Your help with earlier NCDS surveys was very important and much appreciated - it helped make the study a success.

We are writing because our records show that you are one of a twin pair and we would like to invite you to take part in a special survey of twins. As one of a twin pair you are a very special member of the study and we hope you will agree to help make our current survey of twins a success. Indeed, as there are only 250 sets of twins in the study, you are crucial to the project as we cannot replace or substitute another person’s information for yours.

All we would like you to do is to complete the enclosed questionnaire and return it in the envelope provided together with the updated address information for both you and your twin. The questionnaire should only take you around 15 minutes to complete. Your answers, when combined with information you have given during earlier surveys, will help us learn more about the impact of childhood experiences on employment and wages in adult life.

As a twin, you may have already taken part in other projects which seek to gather information about your experiences but, as a very special part of the NCDS, I hope you will agree to help with this survey.

As always, the information that you give will be treated as confidential, and will not be reported in any way which enables you to be identified. We will provide a summary of what we find when the project is completed.

If you have any queries about the survey, please contact my colleague, Denise Hawkes – you can telephone 020 7612 6881, or email d.hawkes@ioe.ac.uk. Alternatively, you can call the usual freephone number - 0500 600 616 - or email ncds@ioe.ac.uk.

I would like to thank you in advance for your time.
Yours sincerely,

Professor Heather Joshi