IQ and Family Background:
Are Associations Strong or Weak?

Anders Björklund
Karin Hederos Eriksson
Markus Jäntti

DISCUSSION PAPER SERIES

I Z A

IQ and Family Background:
Are Associations Strong or Weak?

Anders Björklund
*SOFI, Stockholm University
and IZA*

Karin Hederos Eriksson
SOFI, Stockholm University

Markus Jäntti
SOFI, Stockholm University

ABSTRACT

IQ and Family Background:
Are Associations Strong or Weak?[*]

For the purpose of understanding the underlying mechanisms behind intergenerational associations in income and education, recent studies have explored the intergenerational transmission of abilities. We use a large representative sample of Swedish men to examine both intergenerational and sibling correlations in IQ. Since siblings share both parental factors and neighbourhood influences, the sibling correlation is a broader measure of the importance of family background than the intergenerational correlation. We use IQ data from the Swedish military enlistment tests. The correlation in IQ between fathers (born 1951-1956) and sons (born 1966-1980) is estimated to 0.347. The corresponding estimate for brothers (born 1951-1968) is 0.473, suggesting that family background explains approximately 50% of a person's IQ. Estimating sibling correlations in IQ we thus find that family background has a substantially larger impact on IQ than has been indicated by previous studies examining only intergenerational correlations in IQ.

JEL Classification: J0, I0, J1

Keywords: ability, intergenerational mobility, family background

Corresponding author:

Anders Björklund
Swedish Institute for Social Research (SOFI)
Stockholm University
SE-10691 Stockholm
Sweden
E-mail: Anders.Bjorklund@sofi.su.se

---

1 Introduction

Recent years have seen an upsurge of studies on intergenerational associations in income and education. For example, recent surveys show estimates of intergenerational income elasticities for several countries, and corresponding estimates of years of schooling for a very large number of countries (Björklund & Jäntti 2009, Corak 2006, Hertz et al. 2007) . As a complement and in order to understand the mechanisms behind these associations, it is also of interest to learn about the intergenerational transmission of skills and abilities, for example such ones called IQ.

The literatures in various disciplines offer a number of estimates of intergenerational IQ correlations, but most of these stem from small and non-representative samples (Bowles & Gintis 2002, Bouchard & McGue 1981). Recently, however, Black et al. (2009) have used data covering the whole Norwegian male population to estimate the father-son correlation in the IQ scored in the compulsory enlistment tests to the country's military service. Using log scores, they obtain a precisely estimated correlation of 0.32.

Is such a number low or high? Does it motivate popular expressions in the intergenerational literature such as "Like Father, Like Son" and "The Apple Does Not Fall Far From The Tree"? In our view, this is a matter of judgement. On the one hand, as stressed by the authors, the number exceeds the corresponding ones for long-run earnings in Norway. Further, the interpretation is that a 10 percent differential in fathers' IQ at age 18 is associated with an expected differential of 3.2 percent for sons at the same age. Although this implies some "regression toward the mean", there is also substantial transmission from one generation to the next. On the other hand, the implication is that the explanatory power of father's IQ is quite low. A correlation of 0.32 implies an R^2 close to 10 percent, leaving 90 percent to factors uncorrelated with father's IQ.

A less recognized, but in our view important, literature has instead explored the role of family background by using measures of sibling similarity, such as the sibling correlation. It has long been known that a sibling correlation is a broader measure of the impact of family and community background than an intergenerational one.[2] The reason is that siblings share not only the observed parental characteristic that can be used in an intergenerational study – be it income, education, occupational class or IQ that is examined – but also many unobserved factors of great importance for their outcomes.

In this study, we estimate both intergenerational and sibling correlations in IQ on a large representative sample. We use a Swedish data set that is constructed in a similar way as the Norwegian data set but is complemented with information about brothers. We find an estimate of the intergenerational correlation that is very close the Norwegian one. However, when we use data on brothers, we find that close to half of the variation in IQ is accounted for

---

2 This insight goes back at least to Corcoran et al. (1976) . See also, for example, Hauser & Mossel (1985); Erikson (1987); Sieben & De Graaf (2003) for studies in sociology using occupational and educational variables. Solon (1999) offers a formal exposition of the interpretation of the sibling correlation and its relationship to the intergenerational correlation discussed here.

by family and community background factors. And this number is a lower bound on the role of family and community background because siblings unlikely share all of those inherited factors that influence the outcome.

We continue the paper in the next section in which we explain the interpretation of the sibling correlation and its relationship to the intergenerational coefficient. In the third section, we present the data. The main results are shown in the fourth section, followed by a brief concluding section.

2 The relationship between sibling and intergenerational correlations

Consider the following decomposition of an outcome, y, of interest – in our case IQ – for individual j in family i :

$$y_{ij} = a_i \quad b_{ij} \tag{1}$$

where a_i is a component common to all siblings in family i , and b_{ij} is a component unique to individual j in family i , which captures individual deviations from the family component. The two components are orthogonal by construction. Thus, the variance of y_{ij} is the sum of the variances of the family and individual components:

$$\sigma_y^2 = \sigma_a^2 \quad \sigma_b^2. \tag{2}$$

The share of the variance in the outcome variable, y_{ij} , which can be attributed to family background effects, is

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 \quad \sigma_b^2} \tag{3}$$

This share coincides with the correlation in the outcome variable between randomly drawn pairs of siblings, which is why ρ is called a sibling correlation.

A sibling correlation can thus be thought of as an omnibus measure of the importance of family background and community effects. It includes the effect of all parental factors shared by siblings, such as common genes, parental resources and typically unobserved parental influences such as aspirations and cultural inheritance. It further includes things not directly experienced in the home, such as school, church and other neighborhood effects. Interactions among the siblings are also likely to affect the common family component. On the other hand, genetic traits not shared by siblings, differential treatment of siblings, time-dependent changes in family resources and neighborhoods, are captured by the individual component b_{ij}. Because such factors are also part of family and community background, the sibling correlation is a lower bound on the importance of family and community background.

In addition to this appealing interpretation as an omnibus measure, the sibling correlation also has the following relationship to the corresponding intergenerational correlation (IGC):

Sibling correlation = IGC^2 + other shared factors that are uncorrelated with parental income. (4)

Because the square of the IGC is the explanatory power of the parental characteristic, it follows that the sibling correlation captures these factors plus others that are uncorrelated with the parental characteristic. It is an empirical issue, to which we soon turn, to determine the relative importance of the observed parental characteristic and the other factors shared by the siblings. A recent summary by Björklund & Jäntti (2009) of intergenerational and sibling correlations for long-run earnings suggests that less than half of the sibling correlation can be accounted for by father's earnings.

In order to calculate the sibling correlation in IQ, ρ, we need estimates of the within-family variation, σ^2_b, and the between-family variation, σ^2_a. These are obtained by estimating the parameters in equation 2 using a mixed-effect model, including cohort effects for both fathers and sons in the fixed part (Pinheiro & Bates 1999). We estimate the standard error of the sibling correlation by applying the so-called delta method to the estimated variance matrix of the variance components.

3 The Data

In order to construct our analysis data set, we exploit the valuable opportunity in the Nordic countries to use a unique personal identifier to match administrative register data from different sources.

The IQ data stem from the Swedish military enlistment tests. These tests, that are designed to measure general ability g, are compulsory for every young Swedish man, and they usually take place the year he turns 18. The purpose of the enlistment procedure is to assign the conscript to an appropriate duty position. The classification of conscripts to different positions is done according to requirement profiles concerning cognitive, personality, medical and physiological variables (Mårdberg & Carlstedt 1998). The IQ measure we use is a measure of general cognitive ability reported in stanine units (stanine is a contraction of standard and nine and it is a method of scaling test scores on a nine-point standard scale with a mean of five and a standard deviation of two). The measure of general cognitive ability that we obtain from the Swedish Military Archive and The National Service Administration is derived from a set of subtests evaluating different aspects of cognitive ability. The set of subtests has changed several times since it was first implemented in 1944. The men in our sample were subjected to three different sets of subtests depending on what year they enlisted in the military. Those who enlisted before 1980 performed four subtests: instructions, concept discrimination, technical comprehension and paper form board. Those who enlisted between 1980 and 1993 also performed four subtests, but concept discrimination was replaced by a synonym test, and paper form board by metal folding (Carlstedt 2000). Finally, those who enlisted in 1994 or later performed ten subtests with inductive reasoning, spatial, verbal and technical comprehension parts (Mårdberg & Carlstedt 1998).

We have access to IQ data for the 1951-1980 cohorts. In order to match brothers and father-son pairs, we exploit the so-called Multi-generational register developed by Statistics Sweden. Our starting point is a 35 percent random sample of the Swedish population. For this sample, we identify all full biological brothers. We focus on cohorts of brothers born 1951-68, but we also examine subsets of this sample. For our intergenerational estimates, we focus

on fathers born 1951-56 and identify their biological sons born through 1980. This implies that we are forced to use quite young fathers, and generally the sons were the first-born ones. We discuss the consequences of this in Section 4.

We describe our intergenerational sample in Table A1 in the Appendix and the sample of brothers in Table A2. Our father-son sample consists of 22,626 father-son pairs. The fathers averaged 24.2 years of age when the sons were born. The total sample of brothers born 1951-68 consists of 379,456 individuals from 264,639 families. Thus we follow the convention from the sibling correlation literature to include singletons in the estimations. When there are more than two brothers who satisfy the age criterion in a family, we include all.

4 Results

We first report our estimated intergenerational correlations in Table 1. Our results, which are obtained from models that also include birth year-indicators for both fathers and sons, are strikingly similar to the Norwegian estimates (Black et al. 2009). When we use a linear model, our estimate is 0.347 (compared to 0.38 for Norway), and when we use a log-log model the coefficient (elasticity) is 0.327 (0.32 for Norway). We also report the adjusted R^2, which are 0.132 and 0.120 respectively. Because a model with father's IQ entered as 8 dummy variables might have higher explanatory value in that it could capture possible non-linearities, we also estimated such a specification. The results reported in the last row of Table 1 show that the adjusted R^2 from such a specification is not higher than in the linear model.

Because of data availability, we were forced to use father-son pairs with rather young fathers. Although we have no specific reason to believe that this sample criterion will bias the results in any specific direction, it is a concern that the results might be sensitive to this property of our sample. It is comforting, however, that Black et al. (2009) found no significant interaction with "first son". We also experimented with our sample and included only father-son pairs in which the father was at least 24 years old the year the son was born. The estimates were virtually identical.

We report our brother correlation estimates in Table 2. When we include all brothers born 1951 to 1968, we get an estimate of 0.473 with a negligible standard error. In the next rows we have split the sample into cohorts of brothers born in 1951-56, 1957-62 and 1963-68. In this way, we examine closely spaced brothers and will be able to detect a trend for such brothers. We do find an increase from 0.489 (the first two cohort groups) to 0.513. This change is statistically significant but substantially insignificant.

The last two rows show estimates for brothers who are more widely spaced, being at least 5 calendar years apart. These estimates are only marginally lower than those for closely spaced brothers, 0.436 and 0.449 respectively. The difference in sibling similarity in years of schooling between siblings with small (four years or less) and large (more than four years) age spread has also been found to be small (Conley & Glauber 2008). Because more widely spaced siblings are likely to interact less and be exposed to more different "shocks", we infer that such factors are not very important for sibling similarity; permanent family and community characteristics are more likely determinants.

These results suggest that almost half of the variation in IQ can be attributed to factors shared by siblings. As emphasized above, this number is a lower bound on the importance of

family and community background factors. For example, all our genes are inherited from our parents, but we share on average only half of them with our full biological siblings. The exception is monozygotic (identical) twins, who have identical genes. In the second column, we report estimates for twins only. Because our register data does not contain information about zygosity, we have to stick to a mixture of identical and fraternal twins. Our estimates are clearly higher for such brothers, around 0.65 for all cohort groups. This number might be both an under- and an overestimate of the "true" impact of family and community background. Two arguments speak in favour of an underestimate. First, the sample contains around 50 percent fraternal twins who only share 50 percent of their genes with their brother. Second, even twins might be exposed to different "shocks" or differential treatment that are part of their family and community background but are not shared with their twin sibling. The argument in favour of an overestimate is that twins might interact very closely and in a way that has no counterpart among non-twin siblings and thus in the majority of the population. Our results above regarding the low importance of spacing suggest, however, that this argument might be weak.

5 Conclusions

We started out this inquiry by asking whether the association between IQ and family background is strong or weak. We first estimated the intergenerational correlation between fathers' and sons' IQ to 0.347. This estimate, very close to recent Norwegian results, strikes us as quite low. The implication of this estimate is that only 13 percent of the variation in sons' IQ can be statistically accounted for by fathers' IQ. Then we turned to the broader measure of the impact of parental and community background, namely the brother correlation. In addition to being broader, this measure can be interpreted as a lower bound on the fraction of total variation that can be attributed to factors that brothers share. Our estimates were very close to 0.5, suggesting that 50 percent is a lower bound of broad family background factors. For twin brothers – who likely consist of equal shares of MZ and DZ twins – the estimate was as high as 0.65. This estimate may overestimate but might also underestimate the role of family background. These numbers strike us as being quite large.

What is it then that brothers share and is important for their IQ but is uncorrelated with their father's IQ? An obvious candidate is the mother's IQ. Although spouses' IQ are most likely positively correlated and thus partly capture the same background factors, the combination of father's and mother's IQ is likely to raise the explanatory power in an intergenerational equation. Indeed, in a summary of previous estimates based on small and non-representative samples, Bowles & Gintis (2002) report the highest correlation from a study that applies the average of the two parents' IQ. We doubt though that simply adding mother's IQ would bring the explanatory power close to what the sibling similarity suggests. For example, attempts to account for the sibling similarity in long-run earnings by means of the education of both parents do not appear to capture much of the sibling similarity (Björklund et al. 2008). We hypothesize that very detailed information about parental aspirations, attitudes and parenting practices is needed to account for the large gap between

what sibling studies and intergenerational studies suggest about the role of family background factors.

Finally, we note that even though we have benefited from having access to a large and representative data set, and that the IQ test is taken from a situation in the "real life", our data set has some limitations. We have already noted the unfortunate gender bias that follows from using military enlistment data, namely that neither the role of mothers nor the outcomes for daughters can be taken into account. It might also be that the attitude to the military service has an impact on how well (or badly) young men do on the military enlistment tests. Such attitudes are most likely to "run in the family", so they might bias the family associations upwards. However, both the father-son and the brother correlation are likely to suffer from such an upward bias so we find it unlikely that our most important finding – the strikingly larger role for family background factors suggested by sibling similarity compared to the father-son association – can be explained by such bias.

References

Björklund, A. & Jäntti, M. (2009). Intergenerational income mobility and the role of family background. In W. Salverda, B. Nolan, & T. M. Smeeding (Eds.), Oxford Handbook of Economic Inequality chapter 20, (pp. 491–521). Oxford: Oxford University Press.

Björklund, A., Lindahl, L., & Lindquist, M. J. (2008). What more than parental income? An exploration of what swedish siblings get from their parents. IZA Discussion Paper 3735, IZA, Bonn.

Black, S., Devereaux, P. J., & Salvanes, K. G. (2009). Like father, like son? a note on the intergenerational transmission of IQ scores. Economics Letters, forthcoming.

Bouchard, T. J. & McGue, M. (1981). Familial studies of intelligence: A review. Science, 212, 1055–1059.

Bowles, S. & Gintis, H. (2002). The inheritance of inequality. Journal of Economic Perspectives, 16(3), 3–30.

Carlstedt, B. (2000). Cognitive abilities: aspects of structure, process and measurement. PhD thesis, Gothenburg University.

Conley, D. & Glauber, R. (2008). All in the family? family composition, resources, and sibling similarity in socioeconomic status. Research in Social Stratification and Mobility, 26, 297–306.

Corak, M. (2006). Do poor children become poor adults? Lessons for public policy from a cross country comparison of generational earnings mobility. Research on Economic Inequality, 13(1), 143–188.

Corcoran, M., Jencks, C., & Olneck, M. (21976). The effects of family background on earnings. American Economic Review, 66(2), 430–435.

Erikson, R. (1987). The long arm of the origin: The effects of family background on occupational and educational achievement. In U. Bergryd & C.-J. Janson (Eds.), Sociological Miscellany. Stockholm University.

Hauser, R. M. & Mossel, P. A. (1985). Fraternal resemblance in educational attainment and occupational status. The American Journal of Sociology, 91(3), 650–673.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N., & Verashchagina, A. (2007). The inheritance of educational inequality: International comparisons and fifty-year trends. The B.E. Journal of Economic Analysis & Policy, 7(2 (Advances)), Article 10.

Ihaka, R. & Gentleman, R. (1996). R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics, 5(3), 299–314.

Mårdberg, B. & Carlstedt, B. (1998). Swedish enlistment battery (SEB): Construct validity and latent variable estimation of cognitive abilities by the CAT-SEB. International Journal of Selection and Assessment, 6(2), 107–114.

Pinheiro, J. C. & Bates, D. M. (1999). Mixed-Effects Models in S and S-PLUS. Statistics and Computing. New York: Springer-Verlag.

Sieben, I. & De Graaf, P. M. (2003). The total impact of the family on educational attainment: A comparative sibling analysis. European Societies, 5(1), 33–68.

Solon, G. (1999). Intergenerational mobility in the labor market. In O. Ashenfelter & D. Card (Eds.), Handbook of Labor Economics, volume 3 (pp. 1761–1800). New York: Elsevier Science B.V.

Table 1. Estimated intergenerational IQ correlations. Standard errors within parentheses.

Dependent variable	IQ	Log(IQ)	Adjusted R-square
Father's IQ	0.347 (0.006)	-	0.132
Log (father's IQ)	-	0.327 (0.007)	0.120
Father's IQ in nine levels	Not reported	-	0.132

Note: The reported estimates are unstandardized regression coefficients, but since the standard deviations for fathers' and sons' IQ are almost the same (see Table A1), the estimates can be interpreted as correlations. The equations also include cohort controls for fathers and sons.

Table 2. Estimated brother correlations (standard errors in parentheses)

Years of birth and spacing	All	Twins	Only non-twins
All brothers born 1951-68	0.473 (0.002)	0.654 (0.036)	0.470 (0.003)
All brothers born 1951-56	0.489 (0.009)	0.664 (0.063)	0.480 (0.003)
All brothers born 1957-62	0.488 (0.003)	0.645 (0.065)	0.480 (0.003)
All brothers born 1963-68	0.513 (0.010)	0.653 (0.060)	0.507 (0.020)
Brothers born 1951-62, at least 5 years apart	0.450 (0.008)	Irrelevant	Irrelevant
Brothers born 1957-68, at least 5 years apart	0.454 (0.008)	Irrelevant	Irrelevant

Note: Estimates obtained using the lme function from package nlme in R(12, 17). Standard errors are computed using the delta-method from the estimated variance matrix of the variance components.

Appendix:

Table A1. Sample of fathers and sons. Means and standard deviations within parentheses. Number of father-son pairs: 22,626.

	Fathers	Sons
Year of birth	1952.9 (1.6)	1977.2 (2.4)
Father's age son's birth	24.2	NA
IQ	4.97 (1.89)	4.84 (1.91)

Table A2. Descriptive statistics of brother samples. Standard deviations within parentheses.

Years of birth and spacing	Number of families	Number of individuals	IQ
All brothers born 1951-68			
a) All	264 639	379 456	5.2 (2.0)
b) Twins	2 779	5 585	4.8 (2.0)
c) Non-twins	263 021	373 871	5.2 (2.0)
All brothers born 1951-56			
a) All	107 171	125 851	5.2 (2.0)
b) Twins	967	1 940	4.8 (2.0)
c) Non-twins	106 408	123 911	5.2 (2.0)
All brothers born 1957-62			
a) All	105 325	120 721	5.1 (1.9)
b) Twins	840	1 691	4.9 (1.9)
c) Non-twins	104 649	119 030	5.1 (1.9)
All brothers born 1963-68			
a) All	111 988	132 884	5.1 (1.9)
b) Twins	976	1 954	4.8 (1.9)
c) Non-twins	111 182	130 930	5.1 (1.9)
Brothers born 1951-62, at least 5 years apart	11 060	26 180	5.0 (2.0)
Brothers born 1957-68, at least 5 years apart	12 013	27 636	5.0 (2.0)