Education and voter turnout revisited: evidence from a Swedish twin sample with validated turnout data

Abstract

The association between education and voter turnout is well-established in almost a century of research. Whether this association reflects a causal effect, however, is still subject to substantial debate. This study uses a discordant twin design and variance decomposition methods with newly available validated turnout data in a large sample of Swedish twins to estimate 1) the effect of years of education on voter turnout in both national and European elections, net of common genetic and environmental factors, and 2) the extent to which the association between the two is biased by genetic confounding. The results show that education does not have an effect on national electoral turnout, but does have an effect on turnout for the European elections. Furthermore, the association between education and turnout is shown to be affected by substantial genetic confounding, underscoring the importance of taking this issue seriously in observational research.

1 Introduction

A vast range of research in political science has documented the various ways in which political participation is unevenly distributed across social strata and covaries with different types of individual resources. Income, wealth, socio-economic background, ethnicity and personality attributes are all examples of variables that to some degree predict one's propensity to partake in political activities or vote in elections.

Another such individual resource is education. The association between education and voter turnout is well-established in almost a century of political science literature (Merriam and Gosnell, 1924; Lewis-Beck et al, 2008). Whether this reflects a causal effect of education, or if there are other explanations, has been subject to substantial debate. Generally speaking, there are several reasons why such an association may not be causal. First and foremost, people with different levels of education also likely differ on many other social factors relevant for political participation that may not be causally intermediary. Second, individual factors such as personality traits or cognitive capacity may influence education and participation simultaneously. And third, there may be common underlying genetic factors that are associated with both.

In the last decade, several studies have addressed these issues using different types of causal or quasi-causal designs. For example, Sondheimer and Green (2010) present the results from three policy experiments and conclude that education does have a marked impact on electoral participation among lower socioeconomic groups in the United States. Dinesen et al (2016), further, investigate a broader variety of political participation using a discordant twin design, and find that the association holds in Denmark and the United States but not in Sweden. Others have found more consistent negative results. Examining the impact of exogenous variation in education induced by the Vietnam draft lottery in the United States, Berinsky and Lenz (2011) argue that there was no discernible effect on voter turnout. Persson (2014), further, uses longitudinal data and matching techniques in a British cohort and finds no effects of education on voter turnout. Most recently, Lindgren et al (2019) uses a Swedish education reform as credible exogenous variation and find no aggregate effects on voter turnout – but importantly also find that it did have a marked effect among the most socioeconomically disadvantaged groups.

Results thus vary from zero to positive in the previous literature. These differences can have many different explanations. Berinsky and Lenz (2011) argue that it may simply reflect methodological differences. It is also conceivable that they are to some extent a consequence of heterogeneous effects, i.e. that different groups, countries or types of elections are being studied.

This research letter extends on the twin design of Dinesen et al (2016) and makes use of newly available validated electoral participation data for a large sample of Swedish twins. Sweden is, when it comes to electoral turnout, a least likely case for finding an effect: the voting norm is strong, at least for national elections, and turnout is generally very high in an international comparison. This means that there is little variation to be explained. For this reason, we include participation data for both national parliamentary elections (in 2010), where turnout was high, and for elections to the European parliament (in 2009) where turnout was slightly lower.

I propose a two-pronged methodological approach: first, a discordant MZ twin design is used to test whether the association holds when controlling for genetic and common familial factors. This approach is used by Dinesen et al (2016), but have also been used to investigate, for example, the relationship between education and trust (Oskarsson et al, 2017), and education and political knowledge (Weinschenk and Dawes, 2018). Second, I present results from bivariate variance decomposition models to test to what extent the raw association can be attributed to genetic correlation, i.e. both traits sharing a genetic etiology. Results show that the effect is robust for elections to the European parliament, but not to the national parliament. Furthermore, there is evidence for substantial overlap in the genetic etiology of the traits. The large degree of genetic confounding underscores the liability of purely correlational approaches to severe bias even when a rich set of controls are used.

The unique contribution of this paper over previous similar studies is thus two-fold: first, I test the effect of education on voting with true voting data from different *types* of elections in a high-powered sample. Second, I am able to document the magnitude of genetic confounding in this relationship.

2 Methods and data

Observational designs always run the risk of capturing correlations that actually reflect non-causal relationships, such as underlying variables that affect both the exposure and the outcome, or reversed causation. Measures can be taken to minimize these risks, such as only using independent variables that logically precede the outcome of interest, and controlling for possible confounders. Even with such precautions, results hinge on a number of assumptions: for example, that all independent variables have been measured without error, that all confounders have been identified, that the functional forms are known and correctly specified etc. An issue of particular salience that has been highlighted in several recent studies in the field of political behavior (eg. Oskarsson et al, 2017) is genetic confounding: correlations between measured variables may also to some extent reflect a shared genetic etiology. This problem has no straight-forward solution unless genetically informative data is used.

One solution is to utilize twin data. In particular, when the objective is to control for genetic effects rather than investigate them directly, samples of identical twins are highly useful. The discordant MZ twin design employed here controls for unmeasurable confounders shared by monozygotic (i.e. identical) twins – that is, genetic effects and environmental effects shared by both twins – by restricting the analyzed variation to that within twin pairs. The logic is fairly simple: if the twin with higher education also has a higher propensity for voting, this association atleast cannot be attributed to all the factors shared with the other twin. In practice, this approach boils down to using data only for identical twins and including twin pair fixed effects.

To get a sense of the external validity of the results, regular (naive) models are also included for the entire population. This allows for a direct comparison between the naive effects for population-wide data and naive effects in the twin sample used.

In order to disentangle the amount of covariance between education and electoral participation that is captured by genetic correlation (that is, traits sharing genetic influences), I also leverage the variation from same-sex dizygotic (fraternal) twin pairs and use a bivariate ACE decomposition. Variance decomposition models are structural equation models designed to estimate the amount of variance in a given trait (and in this case, the covariance between multiple traits) that can be attributed to different sources. The variance components typically used in behavior genetic studies are an additive genetic component (A), common environmental

Table 1: Descriptive statistics

VARIABLE	Mean	SD	Min	Max	N
Nat, MZ	.96	.2	0	1	5,428
Nat, DZ	.96	.2	0	1	4,002
EU, MZ	.62	.48	0	1	5,420
EU, DZ	.63	.48	0	1	3,967
Education, MZ	12.39	2.64	7	19	5,601
Education, DZ	11.93	2.76	7	19	4,134
Birthyear, MZ	1955.66	12.28	1922	1979	5,618
Birthyear, DZ	1951.58	12.11	1921	1979	4,138
Nat, pop	.82	.39	0	1	5,776,821
EU, pop	.47	.50	0	1	5,607,660
Education, pop	11.56	2.82	7	19	5,885,351
Birthyear, pop	1953.49	15.98	1899	1979	5,607,660

Note: Nat=national election participation, EU=European election participation, MZ=monozygotic twins, DZ=dizygotic twins (same-sex pairs), pop=population data

factors shared within the twin pair (C), and a unique environmental factor (E). Intuitively, if the correlation between education in one twin and turnout in the other is higher among identical twins (who share 100% of their segregating DNA) than among fraternal twins (who share on average 50%), then education and electoral turnout must share some degree of genetic etiology.

Data comes from registry sources. First of all, the sample consists of all genotyped twin pairs in the Swedish Twin Registry, which delivered the zygosity (fraternal or identical) of the twins. This means that the sample is not fully representative of the twin population, since consent to participate in genotyping may correlate with other characteristics. The independent variable, years of education is taken from the LISA registry for 2009.

Validated turnout data from the general elections of 2010, and the European elections of 2009, are used. The reliance on actual turnout data rather than self-reported turnout has two major advantages. First, some selection bias is avoided. Second, self-reported turnout is known to be notoriously unrealiable both due to social desirability bias and misremembering (Karp and Brockington, 2005), which is bound to both attenuate effect estimates and decrease statistical precision. In this case relying on registry data instead of survey data also increases the sample size substantially, from roughly 2200 in Dinesen (2016) to more than 5400 twins (and even more when including also fraternal twins in the decomposition models).

The sample is restricted to twin pairs born before 1980 (who have consequently reached the age of 30 at the time of the elections 2009/2010) to ensure that they will have finished their formal education, and to avoid the risk of reverse causation. The same restriction is put on the population data.

Descriptive statistics are presented in table 1. As is evident, the turnout for the general elections is very high by international standards, and even higher in this sample of twins than in the general population. By comparison, the turnout for the European election is substantially lower, leaving a fair amount of variation to be leveraged.

3 Results

Table 2 presents OLS models for both the twin sample (naive and within-twin pair models) and the general population.

In the twin sample, effects are significant for both of the naive models. The effect is substantially larger for the European elections (.038 vs. .005, i.e. roughly four percentage points increased likelihood of voting per added year of education, vs. half a percentage point increase) – likely owing to the smaller amount of variation in general election turnout.

Table 2: OLS results

	(1)	(2)	(3)	(4)	(5)	(6)
	Naive	Within	Naive	Within	Population	Population
VARIABLES	Nat	Nat	EU	EU	Nat	EU
Education	0.00521***	-0.00174	0.0377***	0.0182*	0.0186***	0.0429***
	(0.00102)	(0.00363)	(0.00262)	(0.00791)	(5.56e-05)	(7.45e-05)
Constant	0.894***	0.980***	0.155***	0.397***	0.617***	-0.0158***
	(0.0139)	(0.0450)	(0.0343)	(0.0981)	(0.0.00063)	(0.000885)
Observations	5,414	5,414	5,408	5,408	5,591,016	5,411,102
R-squared	0.005	0.613	0.042	0.670	0.020	0.058
Twin pair FE	No	Yes	No	Yes	-	-

Robust standard errors in parentheses
*** $p<0.001$, ** $p<0.01$, * $p<0.05$, $^{+}$ $p<0.1$

Table 3: Bivariate ACE models

	Nat	EU	Education
h^2	.22	.30	.38
	[.11;.34]	[.19;.40]	[.32;.44]
c^2	-.03	.03	.36
	[-.12;.07]	[-.07;.12]	[.30;.41]
e^2	.80	.68	.26
	[.77;.84]	[.65;.71]	[.25;.28]
biv h^2	.06	.08	-
	[-.00;.12]	[.03;.14]	
$\%r_g$	80%	42%	-

The first and second columns show the bivariate heritabilities of the respective electoral participation with education years.

This can be compared to the results for the general population: the estimate for the EU election is just slightly higher (4.3 percentage points), whereas the estimate for the national election is almost four times higher (1.9 percentage points). The latter difference is also statistically significant. The reason why the effect size differs by such a large magnitude between the twin sample and the general population for the national election is most likely that the aggregate turnout is even higher in the twin sample than in the general population, leaving even less variation to be explained.

When restricting the model to the within-pair variation, both estimates are depressed. The effect for the European election decreases to .018, i.e. roughly two percentage points per added education year, but remains statistically significant. The effect for the general election disappears completely ($-.002$ and no longer significant).[1]

When turning to the bivariate ACE models presented in table 3, it is evident, first and foremost, that there is a moderate to high degree of heritability for both voting behavior and education years among Swedish twins. The heritability of education is 38%, whereas the heritability of electoral participation is 30% for the

[1]The variation within MZ twin pairs is an average of 1.2 education years, six percentage points in national election participation and 32 percentage points in European election participation. This should leave sufficient variation for the discordant analysis to make sense.

European election and 22% for the general election. The remaining variation is mainly captured by unique environment for voting behavior, and roughly equally split between common and unique environment for education.

Education and voter turnout also appears to share some degree of genetic etiology. The bivariate heritabilities are .06 for the national elections and .08 for the European elections, indicating genetic correlations of about .21 and .24, respectively. This means that the proportions of the naive correlations between education and electoral participation can be attributed to genetic overlap by about 80% and 40%, respectively. It should also be noted, however, that these overlap figures are only significant for the European elections.

4 Discussion

When moving from naive correlations to the stronger case of within-pair MZ twin comparisons, the association between education and electoral turnout is decreased in magnitude but holds for European elections, and disappears completely for domestic general elections. It thus appears that among Swedish twins, an increased level of education does not produce increased domestic electoral participation, but indeed seems to lead to increased participation in the European election.

There are several possible reasons for this discrepancy. First and foremost, it may just be a variation issue: the turnout in the general election, especially among the twins in our sample, is very high, meaning that there may not be enough variation to be explained. This is not the case for the European elections. It is also possible, as per Lindgren et al (2019), that the lack of an average effect for the national elections obscures a local effect among disadvantaged groups.

Another explanation for the discrepancy between national and European elections is that, owing to the nature of, and distance to, the political assembly in question, participating in the European elections is perceived to require a higher degree of political sofistication. Education, then, may simply have a larger effect for this case than for national elections, for which debate is prevalent in domestic media on a daily basis throughout the electoral cycle and where issues may be perceived to be more "close to home."

The second thing to note is the general depression of effect sizes when within twin-pair variation is used. The extended analyis using bivariate variance decomposition showed that this reduction to a large extent appears to reflect shared genetic influences on both education and voting. This study can not shed any light on what type of genetic effects, or underlying trait, is driving this, but it is clear that a substantial portion of naive correlations between education and voting behavior can be attributed to genetic correlation. Possible explanations are that the shared genetics reflect endophenotypes like cognitive capacity (see e.g. Denny and Boyle, 2008) or personality characteristics (e.g. Gerber et al, 2011 and Hakimi et al, 2011). This underscores the importance of taking genetic confounding seriously when studying associations of this nature.

It is also important to keep in mind that although a substantial amount of bias due to common genetic and environmental factors is removed, within twin-pair estimates may still be biased for several reasons. First, since twins in the same pair may exert influence on *each other*, statistical assumptions of independence may be violated. Such a violation could bias estimates in either way, depending on how such an influence looks. If twins in the same pair tend to mimic each other and thus become more alike, it will bias coefficients toward zero. If, on the other hand, twins tend to segregate into behavioral "niches" within families and thus become *less* alike, this will bias coefficients away from zero.

Second, variables that can broadly be categorized as unique environmental factors (i.e. E in the variance decomposition models) are still uncontrolled, as long as they happen before the age at which one finishes one's education. For example, if there is an influence on both the level of education reached in adulthood, and electoral turnout, of the social environment imposed by a particular school class, this may leave residual confounding if twins are put in separate classes. The same applies to other early-life experiences that differ between the twins. This category of confounders may bias the results in either direction. It is not, given the data at hand, possible to say if, or in which direction, the within twin-pair results in this study are biased.

Comparing the results from this Swedish twin sample to the existing literature, it becomes clear that a simple, singular answer to the question of whether education has a causal influence on voting behavior should not be expected. Rather, such an effect may be strong among groups where the voting norm is weak or where other individual resources are missing (i.e. the samples studied by Sondheimer and Green, 2010), and is less likely to manifest in more favourable conditions such as in high-turnout elections in affluent and comparatively egalitarian countries like Sweden. It therefore seems possible to reconcile results from several previous studies showing either positive or null results.

References

[1] Berinsky, AJ. and Lenz, GS. (2011). Education and Political Participation: Exploring the Causal Link. *Political Behavior*, 33:357–373.

[2] Denny, K. and Doyle, O. (2008). Political interest, cognitive ability and personality: Determinants of voterturnout in britain. *British Journal of Political Science*, 38(2), 291–310

[3] Dinesen, PT., Dawes, CT., Johannesson, M., Klemmensen, R., Magnusson, P., Norgaard, AS., Petersen, I. and Oskarsson, S. (2016). Estimating the Impact of Education on Political Participation : Evidence from Monozygotic Twins in the United States, Denmark and Sweden. *Political Behavior*, 38(3).

[4] Gerber, AS., Huber, GA., Doherty, D., Dowling, CM., Raso, C. and Ha, SE. (2011). Personality Traits and Participation in Political Processes. *The Journal of Politics*, 73(3).

[5] Hakimi, S., Hejazi, E. and Lavasani, MG. (2011). The Relationships Between Personality Traits and Students' Academic Achievement. *Procedia - Social and Behavioral Sciences*, 29:836-845.

[6] Karp, JA. and Brockington, D. (2005). Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries. *Journal of Politics*, 67(3).

[7] Lewis-Beck, MS., Jacoby, WG., Norpoth, H and Weisberg, HF. (2008). *The American Voter Revisited*. Ann Arbor: University of Michigan Press.

[8] Lindgren, KO., Oskarsson, S. and Persson, M. (2019). Enhancing Electoral Equality: Can Education Compensate for Family Background Differences in Voting Participation? *American Political Science Review*, 113(1):108-122.

[9] Oskarsson, S., Dinesen, PT., Dawes, CT., Johannesson, M. and Magnusson, P. (2017). Education and Social Trust: Testing a Causal Hypothesis Using the Discordant Twin Design. *Political Psychology*, 38(3).

[10] Persson, M. (2015). Education and Political Participation. *British Journal of Political Science*, 45(3).

[11] Sondheimer, RM and Green, DP. (2010). Using Experiments to Estimate the Effects of Education on Voter Turnout. *American Journal of Political Science*, 54:174–189.

[12] Weinschenk, AC and Dawes, CT. (2018). The Effect of Education on Political Knowledge: Evidence From Monozygotic Twins. *American Politics Research*, 46(1).