COMMENTARY

Genotyping the dead: Using offspring as proxy to estimate the genetic correlation of education and longevity

Dalton Conley[a,1] and Ramina Sotoudeh[a]

In the work on the social determinants of health, it has long been recognized that, among the strongest (if not the strongest) predictor of morbidity and mortality, is educational attainment—that is, the number of years of formal schooling (1). To what extent the relationship between education and longevity is a causal one running from schooling to health has been of intense interest to social scientists. Of course, ill health can truncate a schooling career, but there is also adequate reason to suspect that formal schooling does indeed improve health and well-being through a number of channels ranging from improved impulse control (2) to better ability to understand health risks (3) to improved economic circumstances (4). The extent to which we can isolate the mechanisms behind this strong relationship will help us design better interventions to promote health and health equity (not to mention, to mitigate the adverse effects of ill health on education and its related outcomes).

Furthermore, it is critical to know the extent to which the observed association is driven by an underlying genetic correlation. Quantifying the extent to which mortality risk and education share a common genetic component is not just important for "differencing" out genetics to better understand environmental influences that affect both health and education, it also informs theories of evolution and selection. Namely, recent research has shown that alleles that positively predict educational attainment—and its associated endophenotypes—tend to be under positive selection (5). The extent to which the relationship between longevity (and, critically, fertility) and education is genetic suggests a future direction for the human population—not just in terms of inequality in life expectancy, but also in the overall rate of increase in our life spans, which has been slowing in recent decades in wealthy countries such as the United States.

Against this backdrop, Marioni et al. (6) set out to assess the extent to which the relationship between educational attainment and longevity has a genetic basis. Prior research has suggested that the link between years of formal schooling and mortality risk does have a strong genetic link (7); however, much of this work is based on old twin and adoption studies (8, 9) that rest on a number of controversial assumptions (10). Thus, it is a welcome addition to the literature to deploy actual molecular genetic markers to examine this question. To do so, the authors pool three large datasets to test whether genetic variants known to be predictive of educational attainment also predict the length of one's life span.

Such datasets often suffer from a large degree of right censoring because most people who have been genotyped are still living. Furthermore, older samples for whom genetic data were obtained later in life, and for whom prospective mortality data are more likely to be available, are shown to suffer from mortality selection bias. Domingue et al. (11) show that more educated, socioeconomically advantaged, and healthier individuals (based both on phenotypes and polygenic scores) are more likely to survive to be eligible to participate in the genetic sample of Health and Retirement Study. These biases are problematic for studies that seek to understand how genetics influence later life outcomes, including mortality and its various causes. Such right censoring can be treated as a missing data problem. In the case of Marioni et al.'s data, the age at time of mortality for offspring is unobserved, because most respondents are still alive, and the genotypes of parents were not measured. Marioni et al. overcome this problem by taking advantage of the fact that children inherit 50% of each of their parents' genomes. The authors treat offspring genotype as a proxy measure for parental genotype, which they then use to predict parental longevity. Using this approach, Marioni et al. find that the specific polygenic score that predicts educational attainment in the offspring (study participants) also predicts parental longevity. This is a novel approach to address a phenotype that, by definition, cannot be measured among the living.

[a]Department of Sociology, Princeton University, Princeton, NJ 08544

Author contributions: D.C. and R.S. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 13366.

[1]To whom correspondence should be addressed. Email: dconley@princeton.edu.

Limitations of the Paper

Although Marioni et al. successfully overcome the challenges of the data structure, their methods are not completely free of bias. Because an individual's inclusion in the analysis depends on their having had children, individuals without children have no chance of inclusion into the study and those with few children have a attenuated chance of inclusion. This systematically biases the estimate of educational polygenic scores on longevity. We know from previous research that educational attainment, and even polygenic score for educational attainment, is negatively associated with fertility and that this relationship is getting stronger in more recent years (12). This would cause the sample of genotyped offspring to underrepresent parents with more education. There is also the aforementioned positive association between education and health. The combination of these two factors—an underrepresentation of educated parents and a lower level of health associated with lower levels of education—could result in a biased estimate of the relationship between mortality and education if the slope of the relationship differs by education level.

To address this problem, Marioni et al. conduct a separate sensitivity analysis (only using the UK Biobank sample) in which they control for the number of siblings of the respondent, as a way to control for mother's fertility. This adjusts for differences in selection into the sample for people who have children, but, it cannot account for people who had no children and therefore have no chance of inclusion in the sample.

One remaining worry is the trend over time toward increased rates of educational attainment, which might have caused the institutional landscapes for parents and respondents to differ. This study rests on the assumption that the heritability of educational attainment remains constant over time because it uses polygenic score estimates from recent genome-wide association (GWA) studies. It is possible that the genes that influence educational attainment for one cohort are different from those for another because the institutional landscape changes such that different phenotypic attributes are favored. Thus, there could be a problem using the same polygenic score—calculated using modern GWA studies—for both respondents and their parents. This would imply that we cannot accurately capture the effect of a polygenic score for education on the educational attainments of the parents. That is, the alleles that predict educational attainment, and the extent to which they predict it, could be different for parents and offspring. Luckily, recent evidence suggests such a dynamic of different genetic architectures across cohorts is not likely to be a major factor (13).

Potential Mechanisms

As Marioni et al. discuss, there are many potential reasons for why genetic variants that predict education also predict longevity. One possible pathway, extensively explored in the social scientific literature, is that educational attainment affords individuals higher socioeconomic positions, access to healthcare, as well as health-improving resources and environments. The polygenic score for education, in this case, might increase educational attainment, which then, through purely environmental pathways, decreases mortality risk. Another possible mechanism for the effect of education polygenic score on parents' longevity is that offspring genotype by benefitting offspring's educational attainment leads offspring to attain resources that are then used to care for the aging parents. This would be an indirect effect of education polygenic score on parents' longevity that does not work through parents' own educational attainment. Is there evidence of this in their results? The authors find that parental longevity was predicted much more strongly by the phenotype of offspring educational attainment than the education polygenic score. Although this might be due to the low statistical power of the original GWA studies as the authors suggest, it could also be evidence of an environmental pathway through offspring's resources.

Two results reassure us that this is not the case. First, the authors compare the genetic profile for education to genetic profiles of other, more directly health-related, mortality risk factors such as cardiovascular disease and body mass index, and they note that the effects are similar in magnitude. In addition, intergenerational pathways from parent to child that are socially mediated tend to have a stronger effect for mothers than for fathers (14). Furthermore, research on filial support has shown that mothers are supported more by their children than are fathers (15, 16); therefore, if filial support (or parental cultural transmission) was a driving mechanism of Marioni et al.'s findings, we would expect there to be different effects of polygenic score for education on mothers' longevity compared with that of fathers. However, Marioni et al. find an almost identical effect size for mothers and fathers.

Marioni et al. have overcome a missing data problem and paved the way for future research to determine the mechanisms that link the genetic architecture of education, on the one hand, and life expectancy, on the other.

Alternatively, education might not even be a mediator. Another factor such as a genetically induced long-term illness could be leading individuals to drop out of educational institutions as well as incur a higher risk of mortality. The fact that other health-related mortality risk factors have a similar magnitude to the effect of education, supports this possibility. Finally, as the authors point out, biological pleiotropy might be at play. Genetic variants may be responsible for a general state of well-being, which leads to both a higher physical and neural health, and therefore a lower mortality risk and higher educational attainment respectively. Bulik-Sullivan et al. (17) find a genetic correlation between education and a number of important health measures, which provides some evidence for biological pleiotropy.

It is hard to parse out these competing mechanisms, especially when the fine-grained data necessary to test them are not available, and perhaps do not even exist. This study would have greatly benefited from parent-level phenotypic data that could be used to test the competing mechanisms. Given the available data, however, the present study is an important contribution to the study of the genetic underpinnings of two important phenomena. Marioni et al. have overcome a missing data problem and paved the way for future research to determine the mechanisms that link the genetic architecture of education, on the one hand, and life expectancy, on the other.

Future Research

This paper informs not only the substantive issue of education, genes, and health. It also makes an important methodological contribution. Namely, building on earlier work (18), this research group demonstrates the efficacy of using pedigrees to analyze phenotypes that would otherwise not be measureable. That is, linking genotypes of study participants to their relatives' phenotypes provides a tool for researchers to study previously unreachable populations through their relatives. Populations that are systematically missing from any of these Biobank samples (such as

prisoners or other institutionalized populations or the deceased) could be sampled through their relatives.

As with all record linkage, there are many associated ethical concerns. Although participants in a genetic study give their consent to participate, the people studied via these respondents are not consulted, nor have they given consent. Would it be ethical to infer, albeit probabilistically, a person's genotype without acquiring their informed consent? The answer to this question arguably depends on how the genetic data are used. It might not be problematic in the case of Marioni et al. where the quantity of interest is not detrimental to the individuals or their parents and where many parents were deceased. In the wrong hands, however, this method could do harm to vulnerable populations. For example, insurance companies could infer one's genetic propensity to acquire a costly disease without the consent of the patient if they had data on the genotype of the patient's family member(s). They could then deny the patient coverage on this basis. As with all exciting, novel scientific tools, this one should not be blindly used without debate over the social implications of its application.

1 Christenson BA, Johnson NE (1995) Educational inequality in adult mortality: An assessment with death certificate data from Michigan. *Demography* 32(2): 215–229.
2 Kenkel DS (1991) Health behavior, health knowledge, and schooling. *J Polit Econ* 99(2):287–305.
3 Cutler DM, Lleras-Muney A (2010) Understanding differences in health behaviors by education. *J Health Econ* 29(1):1–28.
4 Seeman TE, et al. (2004) Cumulative biological risk and socio-economic differences in mortality: MacArthur studies of successful aging. *Soc Sci Med* 58(10): 1985–1997.
5 Domingue BW, Belsky DW, Conley D, Harris KM, Boardman JD (2015) Polygenic influence on educational attainment. *AERA Open* 1(3):2332858415599972.
6 Marioni RE, et al.; Social Science Genetic Association Consortium (2016) Genetic variants linked to education predict longevity. *Proc Natl Acad Sci USA* 113:13366–13371.
7 Krapohl E, Plomin R (2016) Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Mol Psychiatry* 21(3):437–443.
8 Plomin R, Fulker DW, Corley R, DeFries JC (1997) Nature, nurture, and cognitive development from 1 to 16 years: A parent-offspring adoption study. *Psychol Sci* 8(6):442–447.
9 Turkheimer E, Haley A, Waldron M, D'Onofrio B, Gottesman II (2003) Socioeconomic status modifies heritability of IQ in young children. *Psychol Sci* 14(6): 623–628.
10 Conley D, Rauscher E, Dawes C, Magnusson PK, Siegal ML (2013) Heritability and the equal environments assumption: Evidence from multiple samples of misclassified twins. *Behav Genet* 43(5):415–426.
11 Domingue BW, et al. (2016) Mortality selection in a genetic sample and implications for association studies. bioRxiv:049635.
12 Conley D, et al. (2016) Assortative mating and differential fertility by phenotype and genotype across the 20th century. *Proc Natl Acad Sci USA* 113(24): 6647–6652.
13 Conley D, Laidley TM, Boardman JD, Domingue BW (2016) Changing polygenic penetrance on phenotypes in the 20th century among adults in the US population. *Sci Rep* 6:30348.
14 Conley D, et al. (2015) Is the effect of parental education on offspring biased or moderated by genotype? *Sociol Sci* 2:82–105.
15 Rossi AS, Rossi PPH (1990) *Of Human Bonding: Parent Child Relations Across the Life Course* (Aldine De Gryuter, New York).
16 Silverstein M, Bengtson VL (1997) Intergenerational solidarity and the structure of adult child-parent relationships in american families. *Am J Sociol* 103(2): 429–460.
17 Bulik-Sullivan B, et al.; ReproGen Consortium; Psychiatric Genomics Consortium; Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3 (2015) An atlas of genetic correlations across human diseases and traits. *Nat Genet* 47(11):1236–1241.
18 Joshi PK, et al. (2016) Variants near CHRNA3/5 and APOE have age- and sex-related effects on human lifespan. *Nat Commun* 7:11174.