View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Explore Bristol Research

Munafò, M., Davies, N. M., & Davey Smith, G. (2020). Can genetics reveal the causes and consequences of educational attainment? *Journal of the Royal Statistical Society: Series A*, *183*(2), 681-688. https://doi.org/10.1111/rssa.12543

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
10.1111/rssa.12543

Link to publication record in Explore Bristol Research
PDF-document

This is the final published version of the article (version of record). It first appeared online via Wiley at https://rss.onlinelibrary.wiley.com/doi/epdf/10.1111/rssa.12543 . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistics in Society
Series A

J. R. Statist. Soc. A (2020)
183, *Part* 2, *pp.* 681–688

Can genetics reveal the causes and consequences of educational attainment?

Marcus Munafò, Neil M. Davies and George Davey Smith

University of Bristol, UK

[Received February 2019. Revised October 2019]

Summary. There is an extensive literature on the causes of educational inequalities, and the life course consequences of educational attainment. Mendelian randomization, where genetic variants associated with exposures of interest are used as proxies for those exposures, often within an instrumental variables framework, has proven highly effective at elucidating the causal effects of several risk factors in the biomedical sciences. We discuss the potential for this approach to be used in the context of social and socio-economic exposures and outcomes, such as educational attainment.

Keywords: Causal inference; Education; Genetic structure; Mendelian randomization

Socio-economic inequalities in cognitive and social skills are apparent before children even start school (Blanden *et al.*, 2007; Kautz *et al.*, 2014). These differences are large and pervasive, and their effects persist over the life course, subsequently crystallizing in differences in educational attainment. The concept 'educational attainment' is intuitive and seemingly straightforward, but its measurement is not. There is variation in the number of years that are spent in formal education, but also differences in attainment between those who spend a similar number of years in education (e.g. the number and grade of secondary school qualifications), and even more fine-grained differences between those who achieve similar levels of attainment (e.g. similar university grades at more or less prestigious universities). This is a particular challenge in large-scale genomewide association studies, which typically combine data from multiple samples, drawn from different birth cohorts, different educational systems, and so on. The solution to this data harmonization problem is to rely on a simple, universal but relatively crude metric—the number of years spent in education. In some education systems this measure may be misleading, e.g. in education systems that routinely hold back students who have failed a grade.

There is a broad literature exploring different potential causes of educational inequalities at different ages. This includes studies in economics and the social sciences (e.g. economics, policy studies, sociology and education), and biomedical sciences (e.g. psychology, epidemiology, genetics and epigenetics) (Eccles, 2005; Feinstein, 2003; Jerrim and Vignoles, 2013; Krapohl and Plomin, 2016; Little *et al.*, 2019). These studies have explored a large set of potential inputs into educational attainment, such as the *in utero* environment, and parental, family, school and teacher factors (Burgess, 2016; Karlsson Linnér *et al.*, 2017; Lee *et al.*, 2018; Slater *et al.*, 2012). However, determining causal links from these factors to educational attainment is challeng-

Address for correspondence: Marcus Munafò, School of Psychological Science, University of Bristol, 12a Priory Road, Bristol, BS8 1TU, UK.
E-mail: marcus.munafo@bristol.ac.uk

© 2019 The Authors Journal of the Royal Statistical Society: Series A (Statistics in Society) 0964–1998/20/183681
Published by John Wiley & Sons Ltd on behalf of the Royal Statistical Society.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

ing. Researchers have used a variety of approaches to estimate the contributions to education of the child's cognitive abilities and psychological traits (or essential life skills—often called 'non-cognitive traits') as well as parental inputs of time and money (Smithers *et al.*, 2018).

Another large interdisciplinary literature explores the life course consequences of educational attainment, such as for health and wellbeing, earnings, life satisfaction, fertility, parenting and partnership formation. From the original work of Grossman (1972) and others (e.g. Mincer (1974) and Becker (1975)) onwards, economists have estimated the causal effect of education on earnings and other outcomes. Many different approaches have been used to estimate causal effects of education, including instrumental variables, regression discontinuity designs and comparing outcomes between monozygotic and dizygotic twins. This literature suggests that schooling causes a substantial increase in individual earnings (see for example Card (2001) and Heckman *et al.* (2006)), as well as a reduction in workplace injury rates, unemployment, welfare receipt and quicker entry into the labour market (Oreopoulos and Salvanes, 2011). There is also a considerable literature showing evidence of causal linkages between education and health and other non-pecuniary outcomes (Grossman, 2006; Oreopoulos and Salvanes, 2011). However, there is also uncertainty regarding whether these differences are due to a causal effect of education. Many studies suggest that differences in educational attainment are large and important for later life outcomes, and estimated causal effects on some outcomes are relatively robust across different samples and empirical approaches (e.g. the effect of education on earnings), but there is conflicting evidence on the effects of education on later life health and mortality (Albouy and Lequien, 2009; Clark and Royer, 2013; Davies *et al.*, 2018a; Galama *et al.*, 2018; Lleras-Muney, 2005; van Kippersluis *et al.*, 2011). These differences in health outcomes between educated and less educated groups could be due to pre-existing differences that occur before education (e.g. in diet, physical activity and socio-economic position).

Recent insights into the genetic influences on a range of socio-economic, biological, behavioural and health phenotypes now enable us to apply Mendelian randomization to these questions (Davey Smith and Ebrahim, 2003; Davies *et al.*, 2018b). This is an instrumental variable approach, which uses genetic variants (typically single-nucleotide polymorphisms (SNPs), as proxies for potentially modifiable exposures of interest). It requires SNPs that

(a) are associated with the exposure of interest,
(b) share no causes with the outcome of interest and
(c) do not directly affect the outcome via mechanisms that are not mediated via the exposure (Fig. 1(a)).

The first assumption is directly testable. The second assumption is likely to hold in many applications because of the random inheritance of alleles from parents at conception, and it can be falsified by examining the association of SNPs of interest with measured confounders. The third assumption is not directly testable, but there are an increasing number of instrumental variable estimators that are robust to violations of this assumption (Bowden *et al.*, 2016; Davies *et al.*, 2018b; DiPrete *et al.*, 2018; Hartwig *et al.*, 2017; Hemani *et al.*, 2018; Windmeijer *et al.*, 2019). Furthermore, it is important to distinguish between vertical pleiotropy (Fig. 1(b)), where a variant affects a phenotype such as cognition which goes on to affect education and in turn a health outcome, and horizontal pleiotropy (Fig. 1(c)), for example, where a variant affects health outcomes without being mediated via education. Only horizontal pleiotropy causes bias in a Mendelian randomization analysis and is a threat to both the second and the third assumptions described above. So Mendelian randomization will only be biased if SNPs affect an outcome, and education does not fully mediate the effects of the SNPs. It is worth noting that horizontal pleiotropy is likely to be a particular concern in the context of complex and distal phenotypes

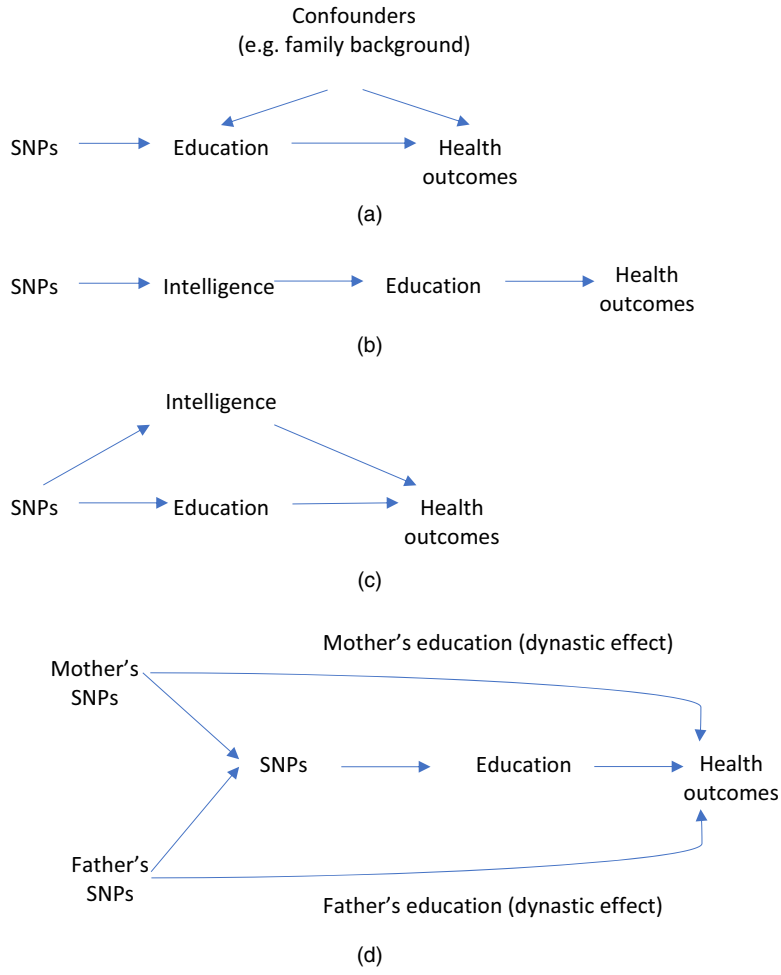

Fig. 1. Causal diagram indicating possible relationships between education and later health outcomes, and how Mendelian randomization can be informative: (a) genetic variants as instruments for education; (b) vertical pleiotropy; (c) horizontal pleiotropy; (d) dynastic effects

such as educational attainment, where the effects of genetic variants are likely to operate via a range of biological, behavioural and social pathways to influence the outcome.

Other threats to Mendelian randomization studies that are particularly relevant in the context of educational research include assortative mating and dynastic effects. Assortative mating occurs when individuals that are phenotypically similar—and therefore genotypically similar—are more likely to have offspring together (Hartwig *et al.*, 2018). Dynastic effects occur, for example, when the parents' education-associated SNPs affect their offspring's outcomes because of their effect on the parents' own educational attainment (Fig. 1(d)). Both assortative mating and dynastic effects can cause bias and false positive findings in Mendelian randomization studies of the effect of education. Cohort studies that include genetic information on both parents and children or samples of siblings enable these potential threats to be investigated and overcome. For example, this enables examination of the effects of parental transmitted *versus* non-transmitted alleles on offspring outcomes. Indeed, the ability to investigate transmitted and

non-transmitted alleles in principle enables us to examine effects of parental education *versus* offspring education on a range of outcomes—something that Mendelian randomization using data from unrelated individuals cannot do because of the correlation between offspring and parental genotype.

Mendelian randomization has proven highly effective at elucidating the causal effects of several risk factors in the biomedical sciences, as well as providing an indication of when risk factors may not affect an outcome. This approach has several potential advantages that are relevant to education research, including robustness to measurement error, reverse causation, endogeneity and confounding, and provides a strong basis for causal inference, particularly when results from this approach are triangulated with those from other approaches that rely on differing assumptions. It has successfully predicted the findings from several randomized controlled trials before the trials were completed and is transforming how drug targets are validated (Walker *et al.*, 2017). The approach has also clarified the health consequences of behaviours (such as alcohol use, smoking, physical activity and obesity) and the relationship between psychosocial indicators such as a sense of wellbeing and health (Wootton *et al.*, 2018). These techniques can potentially be used to address critical questions relating to the determinants and consequences of educational attainment. Researchers have conducted genomewide association studies of educational attainment by using very large samples of data ($N = 1.1$ million) from around the world (Lee *et al.*, 2018). These studies reported 1271 SNPs associated with educational attainment at the genomewide significance threshold of $p < 5 \times 10^{-8}$. Across the entire genome, common SNPs explain 11–13% of the variation in educational attainment, meaning that they strongly predict educational attainment and can provide instruments for use in other samples. Recently this approach has been used to demonstrate a substantial protective effect of education on coronary heart disease, with partial mediation of these effects being through health-related behaviours such as smoking and obesity-related traits (Tillmann *et al.*, 2017). These links can be further interrogated through multivariable Mendelian randomization, e.g. demonstrating that the causal effect of education on smoking does not simply reflect cognitive ability (Sanderson *et al.*, 2019a). It has been used to study antenatal maternal and paternal influences (Davey Smith, 2008; Lawlor *et al.*, 2017), suggesting that maternal alcohol use during pregnancy has detrimental effects on educational outcomes, even when drinking is in the light to moderate range and without binge drinking (von Hinke Kessler Scholder *et al.*, 2014; Zuccolo *et al.*, 2013).

A striking example of how genetic studies can identify the direction of causation is provided by evidence from recent studies investigating the relationship between educational attainment and myopia (Mountjoy *et al.*, 2018). It is possible that myopia could lead to worse educational attainment, e.g. if pupils with uncorrected myopia cannot read the blackboard, or improved attainment, e.g. if myopia leads to 'bookish' children who spend more time reading and therefore learn more. However, there is little evidence that SNPs that are associated with myopia at genomewide levels (Pickrell *et al.*, 2016) associate with educational attainment. Conversely, there is evidence that SNPs that associate with educational attainment at genomewide levels (Okbay *et al.*, 2016) also associate with myopia. These findings suggest that something related to the educational environment influences rates of myopia in the population, but that myopia does not impact on educational outcomes, at least in the context where glasses are an effective and ubiquitous intervention. More research is needed to determine which social or policy interventions can help to break the effects of education on myopia. Further evidence that the effects of education on myopia are likely to be due to an aspect of the environment is provided by changes to rates of myopia in certain east Asian countries, which have increased as the intensity of the education systems in those countries has increased. The underlying distribution of genetic variation in the population cannot have changed sufficiently quickly to explain these changes. The

direction of causation between myopia and education has been the subject of research for many years, but other than through randomized controlled trials there is no way to obtain definitive causal evidence of the effects of the educational environment on myopia. However, although a recent randomized controlled trial provided evidence that interventions during education can reduce myopia (He *et al.*, 2015), these are typically challenging in this context, often expensive, time consuming, underpowered, unrepresentative and not always possible.

Genetics, through the application of Mendelian randomization methods, can therefore provide a valuable source of evidence to address questions regarding the causes and consequences of educational attainment. Its underlying assumptions and potential sources of bias are potentially quite different from other current methods, so it is likely to be most powerful when used in combination with these other non-genetic approaches, such as natural experiments, within a triangulation framework (Lawlor *et al.*, 2016). The increasing availability of genomewide association studies summary data, and the development of multiple methods for interrogating these, offers considerable scope for rapidly and cost-effectively generating valuable causal evidence. Mendelian randomization can also be used to identify intermediate pathways (i.e. the specific aspects of modern educational practices). This requires genomewide association studies of potential mediators (Sanderson *et al.*, 2019b). However, as with any methods, there are limitations. In particular, horizontal pleiotropy is a threat to the assumptions of Mendelian randomization, and these assumptions can either not be tested, or tested only imperfectly. This necessitates the use of multiple methods, including a range of pleiotropy robust and within-family methods with different assumptions and sources of bias (Pingault *et al.*, 2018). Although the field continues to evolve, guidelines now exist for the reporting of Mendelian randomization studies (Davies *et al.*, 2018b). Moreover, interpretation may need to be cautious—it is highly likely that true causal pathways to and from educational attainment are context dependent. In the past higher educational attainment was associated with higher body mass index in the UK, whereas now the opposite is true (Davey Smith, 2003). Similarly, in many parts of India higher educational attainment is associated with high body mass index today (Subramanian *et al.*, 2013). For upstream causes like education, causation will be context dependent but is no less 'causal' because of this. Identifying what are the causes, correlates and consequences of educational attainment (including both positive and negative outcomes) is clearly of profound societal importance and policy relevance. Determining the direction of causality is notoriously difficult for education studies, but evidence from genomewide association studies, and the application of this knowledge in Mendelian randomization, provides a powerful new tool from a perhaps unexpected quarter.

Acknowledgements

The Medical Research Council and the University of Bristol support the Medical Research Council Integrative Epidemiology Unit (grants MC_UU_00011/1 and MC_UU_00011/7). The Economic and Social Research Council supported NMD via a 'Future research leaders' grant (ES/N000757/1). No funding body has influenced the preparation of the paper. This publication is the work of the authors, who serve as the guarantors for its content.

Neil Davies and Marcus Munafò report a grant for unrelated research from the 'Global research awards for nicotine dependence' which is an 'Independent competitive grants program' supported by Pfizer. Marcus Munafò is Co-Director of Jericoe Ltd, which produces software for the assessment and modification of emotion recognition, and a consultant to Cambridge Cognition Ltd. He has received grant funding from Pfizer, and research support in kind from GlaxoSmithKline, unrelated to the work that is presented in this paper. George Davey Smith

has received research funding from Sanofi, Biogen and GlaxoSmithKline for projects that are unrelated to the work that is presented in this paper.

References

Albouy, V. and Lequien, L. (2009) Does compulsory education lower mortality? *J. Hlth Econ.*, **28**, 155–168.
Becker, G. S. (1975) Front matter. In *Human Capital: a Theoretical and Empirical Analysis, with Special Reference to Education*, 2nd edn. Chicago: University of Chicago Press.
Blanden, J., Gregg, P. and Macmillan, L. (2007) Accounting for intergenerational income persistence: noncognitive skills, ability and education. *Econ. J.*, **117**, C43–C60.
Bowden, J., Del Greco, M. F., Minelli, C., Davey Smith, G., Sheehan, N. A. and Thompson, J. R. (2016) Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I^2 statistic. *Int. J. Epidem.*, **45**, 1961–1974.
Burgess, S. M. (2016) Human capital and education: the state of the art in the economics of education. *Discussion Paper 9885*. Institute for the Study of Labor, Bonn.
Card, D. (2001) Estimating the return to schooling: progress on some persistent econometric problems. *Econometrica*, **69**, 1127–1160.
Clark, D. and Royer, H. (2013) The effect of education on adult mortality and health: evidence from Britain. *Am. Econ. Rev.*, **103**, 2087–2120.
Davey Smith, G. (2003) *Health Inequalities: Lifecourse Approaches*. Bristol: Policy Press.
Davey Smith, G. (2008) Assessing intrauterine influences on offspring health outcomes: can epidemiological studies yield robust findings? *Basic Clin. Pharmcol. Toxicol.*, **102**, 245–256.
Davey Smith, G. and Ebrahim, S. (2003) "Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidem.*, **32**, 1–22.
Davies, N. M., Dickson, M., Davey Smith, G., Windmeijer, F. and van den Berg, G. J. (2018a) The effect of education on adult mortality, health, and income: triangulating across genetic and policy reforms. *Preprint bioRxiv 250068*. University of Bristol, Bristol.
Davies, N. M., Holmes, M. V. and Davey Smith, G. (2018b) Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *Br. Med. J.*, **362**, article k601.
DiPrete, T. A., Burik, C. A. P. and Koellinger, P. D. (2018) Genetic instrumental variable regression: explaining socioeconomic and health outcomes in nonexperimental data. *Proc. Natn. Acad. Sci. USA*, **115**, E4970–E4979.
Eccles, J. S. (2005) Influences of parents' education on their children's educational attainments: the role of parent and child perceptions. *Lond. Rev. Educ.*, **3**, 191–204.
Feinstein, L. (2003) Inequality in the early cognitive development of British children in the 1970 cohort. *Economica*, **70**, 73–97.
Galama, T. J., Lleras-Muney, A. and van Kippersluis, H. (2018) The effect of education on health and mortality: a review of experimental and quasi-experimental evidence. *Working Paper 24225*. National Bureau of Economic Research, Cambridge.
Grossman, M. (1972) On the concept of health capital and the demand for health. *J. Polit. Econ.*, **80**, 223–255.
Grossman, M. (2006) Education and nonmarket outcomes. In *Handbook of the Economics of Education*, vol. 1 (eds A. E. Hanushek and F. Welch), pp. 577–633. Amsterdam: North-Holland.
Hartwig, F. P., Davey Smith, G. and Bowden, J. (2017) Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidem.*, **46**, 1985–1998.
Hartwig, F. P., Davies, N. M. and Davey Smith, G. (2018) Bias in Mendelian randomization due to assortative mating. *Genet. Epidem.*, **42**, 608–620.
He, M., Xiang, F., Zeng, Y., Mai, J., Chen, Q., Zhang, J., Smith, W., Rose, K. and Morgan, I. G. (2015) Effect of time spent outdoors at school on the development of myopia among children in China: a randomized clinical trial. *J. Am. Med. Ass.*, **314**, 1142–1148.
Heckman, J. J., Lochner, L. J. and Todd, P. E. (2006) Earnings functions, rates of return and treatment effects: the Mincer equation and beyond. In *Handbook of the Economics of Education*, vol. 1 (eds A. E. Hanushek and F. Welch), pp. 307–458. Amsterdam: North-Holland.
Hemani, G., Bowden, J. and Davey Smith, G. (2018) Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Molec. Genet.*, **27**, R195–R208.
von Hinke Kessler Scholder, S., Wehby, G. L., Lewis, S. and Zuccolo, L. (2014) Alcohol exposure in utero and child academic achievement. *Econ. J.*, **124**, 634–667.
Jerrim, J. and Vignoles, A. (2013) Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes. *J. R. Statist. Soc.* A, **176**, 887–906.
Karlsson Linnér, R., Marioni, R. E., Rietveld, C. A., Simpkin, A. J., Davies, N. M., Watanabe, K., Armstrong, N. J., Auro, K., Baumbach, C., Bonder, M. J., Buchwald, J., Fiorito, G., Ismail, K., Iurato, S., Joensuu, A., Karell, P., Kasela, S., Lahti, J., McRae, A. F., Mandaviya, P. R., Seppälä, I., Wang, Y., Baglietto, L., Binder, E. B., Harris, S. E., Hodge, A. M., Horvath, S., Hurme, M., Johannesson, M., Latvala, A., Mather, K. A., Medland, S. E., Metspalu, A., Milani, L., Milne, R. L., Pattie, A., Pedersen, N. L., Peters, A., Polidoro, S., Räikkönen,

K., Severi, G., Starr, J. M., Stolk, L., Waldenberger, M., Eriksson, J. G., Esko, T., Franke, L., Gieger, C., Giles, G. G., Hägg, S., Jousilahti, P., Kaprio, J., Kähönen, M., Lehtimäki, T., Martin, N. G., van Meurs, J. B. C., Ollikainen, M., Perola, M., Posthuma, D., Raitakari, O. T., Sachdev, P. S., Taskesen, E., Uitterlinden, A. G., Vineis, P., Wijmenga, C., Wright, M. J., Relton, C., Davey Smith, G., Deary, I. J., Koellinger, P. D. and Benjamin, D. J. (2017) An epigenome-wide association study meta-analysis of educational attainment. *Molec. Psychiat.*, **22**, 1680–1690.

Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B. and Borghans, L. (2014) Fostering and measuring skills: improving cognitive and non-cognitive skills to promote lifetime success. National Bureau of Economic Research, Cambridge.

van Kippersluis, H., O'Donnell, O. and van Doorslaer, E. (2011) Long-run returns to education: does schooling lead to an extended old age? *J. Hum. Res.*, **46**, 695–721.

Krapohl, E. and Plomin, R. (2016) Genetic link between socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Molec. Psychiat.*, **21**, 437–443.

Lawlor, D., Richmond, R., Warrington, N., McMahon, G., Davey Smith, G., Bowden, J. and Evans, D. M. (2017) Using Mendelian randomization to determine causal effects of maternal pregnancy (intrauterine) exposures on offspring outcomes: sources of bias and methods for assessing them. *Wellcm. Open Res.*, **2**, article 11.

Lawlor, D. A., Tilling, K. and Davey Smith, G. (2016) Triangulation in aetiological epidemiology. *Int. J. Epidem.*, **45**, 1866–1886.

Lee, J. J., 23andMe Research Team, Cognitive Genomics Consortium, Social Science Genetic Association Consortium, Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., Fontana, M. A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P. N., Walters, R. K., Willoughby, E. A., Yengo, L., Alver, M., Bao, Y., Clark, D. W., Day, F. R., Furlotte, N. A., Joshi, P. K., Kemper, K. E., Kleinman, A., Langenberg, C., Mägi, R., Trampush, J. W., Verma, S. S., Wu, Y., Lam, M., Zhao, J. H., Zheng, Z., Boardman, J. D., Campbell, H., Freese, J., Harris, K. M., Hayward, C., Herd, P., Kumari, M., Lencz, T., Luan, J., Malhotra, A. K., Metspalu, A., Milani, L., Ong, K. K., Perry, J. R. B., Porteous, D. J., Ritchie, M. D., Smart, M. C., Smith, B. H., Tung, J. Y., Wareham, N. J., Wilson, J. F., Beauchamp, J. P., Conley, D. C., Esko, T., Lehrer, S. F., Magnusson, P. K. E., Oskarsson, S., Pers, T. H., Robinson, M. R., Thom, K., Watson, C., Chabris, C. F., Meyer, M. N., Laibson, D. I., Yang, J., Johannesson, M., Koellinger, P. D., Turley, P., Visscher, P. M., Benjamin, D. J. and Cesarini, D. (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.*, **50**, 1112–1121.

Little, C. W., Hart, S. A., Phillips, B. M., Schatschneider, C. and Taylor, J. E. (2019) Exploring neighborhood environmental influences on reading comprehension. *J. Appl. Devlp. Psychol.*, **62**, 173–184.

Lleras-Muney, A. (2005) The relationship between education and adult mortality in the United States. *Rev. Econ. Stud.*, **72**, 189–221.

Mincer, J. (1974) Schooling, experience, and earnings. In *Human Behavior and Social Institutions*, vol. 2. New York: National Bureau of Economic Research.

Mountjoy, E., Davies, N. M., Plotnikov, D., Smith, G. D., Rodriguez, S., Williams, C. E., Guggenheim, J. A. and Atan, D. (2018) Education and myopia: assessing the direction of causality by Mendelian randomisation. *Br. Med. J.*, **361**, article k2022.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., Turley, P., Chen, G. B., Emilsson, V., Meddens, S. F., Oskarsson, S., Pickrell, J. K., Thom, K., Timshel, P., de Vlaming, R., Abdellaoui, A., Ahluwalia, T. S., Bacelis, J., Baumbach, C., Bjornsdottir, G., Brandsma, J. H., Pina Concas, M., Derringer, J., Furlotte, N. A., Galesloot, T. E., Girotto, G., Gupta, R., Hall, L. M., Harris, S. E., Hofer, E., Horikoshi, M., Huffman, J. E., Kaasik, K., Kalafati, I. P., Karlsson, R., Kong, A., Lahti, J., van der Lee, S. J., deLeeuw, C., Lind, C., Lind, P. A., Lindgren, K. O., Liu, T., Mangino, M., Marten, J., Mihailov, E., Miller, M. B., van der Most, P. J., Oldmeadow, C., Payton, A., Pervjakova, N., Peyrot, W. J., Qian, Y., Raitakari, O., Rueedi, R., Salvi, E., Schmidt, B., Schraut, K. E., Shi, J., Smith, A. V., Poot, R. A., St Pourcain, B., Teumer, A., Thorleifsson, G., Verweij, N., Vuckovic, D., Wellmann, J., Westra, H. J., Yang, J., Zhao, W., Zhu, Z., Alizadeh, B. Z., Amin, N., Bakshi, A., Baumeister, S. E., Biino, G., Bønnelykke, K., Boyle, P. A., Campbell, H., Cappuccio, F. P., Davies, G., De Neve, J. E., Deloukas, P., Demuth, I., Ding, J., Eibich, P., Eisele, L., Eklund, N., Evans, D. M., Faul, J. D., Feitosa, M. F., Forstner, A. J., Gandin, I., Gunnarsson, B., Halldórsson, B. V., Harris, T. B., Heath, A. C., Hocking, L. J., Holliday, E. G., Homuth, G., Horan, M. A., Hottenga, J. J., de Jager, P. L., Joshi, P. K., Jugessur, A., Kaakinen, M. A., Kähönen, M., Kanoni, S., Keltigangas-Järvinen, L., Kiemeney, L. A., Kolcic, I., Koskinen, S., Kraja, A. T., Kroh, M., Kutalik, Z., Latvala, A., Launer, L., Lebreton, M. P., Levinson, D. F., Lichtenstein, P., Lichtner, P., Liewald, D. C., LifeLines Cohort Study, Loukola, A., Madden, P. A., Mägi, R., Mäki-Opas, T., Marioni, R. E., Marques-Vidal, P., Meddens, G. A., McMahon, G., Meisinger, C., Meitinger, T., Milaneschi, Y., Milani, L., Montgomery, G. W., Myhre, R., Nelson, C. P., Nyholt, D. R., Ollier, W. E., Palotie, A., Paternoster, L., Pedersen, N. L., Petrovic, K. E., Porteous, D. J., Räikkönen, K., Ring, S. M., Robino, A., Rostapshova, O., Rudan, I., Rustichini, A., Salomaa, V., Sanders, A. R., Sarin, A. P., Schmidt, H., Scott, R. J., Smith, B. H., Smith, J. A., Staessen, J. A., Steinhagen-Thiessen, E., Strauch, K., Terracciano, A., Tobin, M. D., Ulivi, S., Vaccargiu, S., Quaye, L., van Rooij, F. J., Venturini, C., Vinkhuyzen, A. A., Völker, U., Völzke, H., Vonk, J. M., Vozzi, D., Waage, J., Ware, E. B., Willemsen, G., Attia, J. R., Bennett, D. A., Berger,

K., Bertram, L., Bisgaard, H., Boomsma, D. I., Borecki, I. B., Bültmann, U., Chabris, C. F., Cucca, F., Cusi, D., Deary, I. J., Dedoussis, G. V., van Duijn, C. M., Eriksson, J. G., Franke, B., Franke, L., Gasparini, P., Gejman, P. V., Gieger, C., Grabe, H. J., Gratten, J., Groenen, P. J., Gudnason, V., van der Harst, P., Hayward, C., Hinds, D. A., Hoffmann, W., Hyppönen, E., Iacono, W. G., Jacobsson, B., Järvelin, M. R., Jöckel, K. H., Kaprio, J., Kardia, S. L., Lehtimäki, T., Lehrer, S. F., Magnusson, P. K., Martin, N. G., McGue, M., Metspalu, A., Pendleton, N., Penninx, B. W., Perola, M., Pirastu, N., Pirastu, M., Polasek, O., Posthuma, D., Power, C., Province, M. A., Samani, N. J., Schlessinger, D., Schmidt, R., Sørensen, T. I., Spector, T. D., Stefansson, K., Thorsteinsdottir, U., Thurik, A. R., Timpson, N. J., Tiemeier, H., Tung, J. Y., Uitterlinden, A. G., Vitart, V., Vollenweider, P., Weir, D. R., Wilson, J. F., Wright, A. F., Conley, D. C., Krueger, R. F., Davey Smith, G., Hofman, A., Laibson, D. I., Medland, S. E., Meyer, M. N., Yang, J., Johannesson, M., Visscher, P. M., Esko, T., Koellinger, P. D., Cesarini, D. and Benjamin, D. J. (2016) Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, **533**, 539–542.

Oreopoulos, P. and Salvanes, K. G. (2011) Priceless: the nonpecuniary benefits of schooling. *J. Econ. Perspect.*, **25**, 159–184.

Pickrell, J. K., Berisa, T., Liu, J. Z., Segurel, L., Tung, J. Y. and Hinds, D. A. (2016) Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.*, **48**, 709–717.

Pingault, J. B., O'Reilly, P. F., Schoeler, T., Ploubidis, G. B., Rijsdijk, F. and Dudbridge, F. (2018) Using genetic data to strengthen causal inference in observational research. *Nat. Rev. Genet.*, **19**, 566–580.

Sanderson, E., Davey Smith, G. and Munafo, M. (2019a) The effect of education and general cognitive ability on smoking: a Mendelian randomisation study. *Nat. Communs*, **10**, article 2949.

Sanderson, E., Davey Smith, G., Windmeijer, F. and Bowden, J. (2019b) An examination of multivariable Mendelian randomization in the single sample and two-sample summary data settings. *Int. J. Epidem.*, **48**, 713–727.

Slater, H., Davies, N. M. and Burgess, S. (2012) Do teachers matter?: Measuring the variation in teacher effectiveness in England. *Oxf. Bull. Econ. Statist.*, **74**, 629–645.

Smithers, L. G., Sawyer, A. C. P., Chittleborough, C. R., Davies, N. M., Davey Smith, G. and Lynch, J. W. (2018) A systematic review and meta-analysis of effects of early life non-cognitive skills on academic, psychosocial, cognitive and health outcomes. *Nat. Hum. Behav.*, **2**, 867–880.

Subramanian, S. V., Corsi, D. J., Subramanyam, M. A. and Davey Smith, G. (2013) Jumping the gun: the problematic discourse on socioeconomic status and cardiovascular health in India. *Int. J. Epidem.*, **42**, 1410–1426.

Tillmann, T., Vaucher, J., Okbay, A., Pikhart, H., Peasey, A., Kubinova, R., Pajak, A., Tamosiunas, A., Malyutina, S., Hartwig, F. P., Fischer, K., Veronesi, G., Palmer, T., Bowden, J., Davey Smith, G., Bobak, M. and Holmes, M. V. (2017) Education and coronary heart disease: Mendelian randomisation study. *Br. Med. J.*, **358**, article j3542.

Walker, V. M., Davey Smith, G., Davies, N. M. and Martin, R. M. (2017) Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. *Int. J. Epidem.*, **46**, 2078–2089.

Windmeijer, F., Farbmacher, H., Davies, N. and Davey Smith, G. (2019) On the use of the lasso for instrumental variables estimation with some invalid instruments. *J. Am. Statist. Ass.*, **114**, 1339–1350.

Wootton, R. E., Lawn, R. B., Millard, L. A. C., Davies, N. M., Taylor, A. E., Munafò, M. R., Timpson, N. J., Davis, O. S. P., Davey Smith, G. and Haworth, C. M. A. (2018) Evaluation of the causal effects between subjective wellbeing and cardiometabolic health: Mendelian randomisation study. *Br. Med. J.*, **362**, article k3788.

Zuccolo, L., Lewis, S. J., Davey Smith, G., Sayal, K., Draper, E. S., Fraser, R., Barrow, M., Alati, R., Ring, S., Macleod, J., Golding, J., Heron, J. and Gray, R. (2013) Prenatal alcohol exposure and offspring cognition and school performance: a "Mendelian randomization" natural experiment. *Int. J. Epidem.*, **42**, 1358–1370.