1. Madole & Harden, Building Causal Knowledge in Behavior Genetics

02. Word counts

ABSTRACT: 56

MAIN TEXT: 999

REFERENCES: 692

ENTIRE TEXT: 1779

3. TITLE: The providential randomisation of genotypes

4. NAMES: Jean-Baptiste Pingault[1], Pasco Fearon[1], Essi Viding[1], Neil Davies[2,3,4], Marcus R. Munafò[2], George Davey Smith[2]

5. INSTITUTION

1. Department of Clinical, Educational and Health Psychology, University College London, London, UK.

2. Medical Research Council Integrative Epidemiology Unit at the University of Bristol, BS8 2BN, United Kingdom.

3. Population Health Sciences, Bristol Medical School, University of Bristol, Barley House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom.

4. K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Norwegian University of Science and Technology, Norway.

06. FULL INSTITUTIONAL MAILING ADDRESS

Jean-Baptiste Pingault, Psychology, 26 Bedford Way, WC1H0AP, London.

07. INSTITUTIONAL TELEPHONE NUMBER (for correspondence)

JBP: 020 7679 1896, (x 41896)

08. ONE EMAIL ADDRESS EACH

JBP: j.pingault@ucl.ac.uk

PF: p.fearon@ucl.ac.uk

EV: e.viding@ucl.ac.uk

ND: Neil.Davies@bristol.ac.uk

MRM: marcus.munafo@bristol.ac.uk

GDS: KZ.Davey-Smith@bristol.ac.uk

09. ONE HOME PAGE URL EACH

JBP: www.jeanbaptistepingault.com

PF: https://www.cfr.cam.ac.uk/staff/professor-pasco-fearon

EV: mailto:e.viding@ucl.ac.uk https://www.ucl.ac.uk/pals/people/essi-viding

ND: https://research-information.bris.ac.uk/en/persons/neil-m-davies

MRM: https://research-information.bris.ac.uk/en/persons/marcus-r-munafo

GDS: https://www.bristol.ac.uk/people/person/George-Davey%20Smith-285dce3f-4498-4e97-82de-250a865b4483/

10. ABSTRACT

When building causal knowledge in behavioural genetics, the natural randomisation of genotypes at conception (approximately analogous to the artificial randomisation occurring in Randomised Controlled Trials) facilitates the discovery of genetic causes. More importantly, the randomisation of genetic material within families also enables a better identification of (environmental) risk factors and aetiological pathways to diseases and behaviours.

11. TEXT.

Madole and Harden draw parallels between Randomised Controlled Trials (RCTs) and within-family genetic association designs to elaborate on the notion of genetic causation, i.e. whether genes cause behaviours and how to interpret genetic causes. The article is a thoughtful introduction to these topics. Our comment focuses on the core feature shared by both designs, i.e. randomisation. First, we discuss how to best capitalise on natural randomisation to help build causal knowledge. Second, despite randomisation being a core feature of both, we caution against drawing too literal parallels between the two designs.

The parallel drawn by the authors was made explicitly by Fisher who established a direct filiation between the (artificially) randomised design he theorised and the (natural) randomisation of genetic material at conception, in his words: “the factorial method of experimentation, now of lively concern so far afield as the psychologists, or the industrial chemists, derives its structure and its name, from the simultaneous inheritance of Mendelian factors… Genetics is indeed in a peculiarly favoured condition in that Providence has shielded the geneticist from many of the difficulties of a reliably controlled comparison. The different genotypes possible from the same mating have been beautifully randomised by the meiotic process.” (Fisher, 1952) As highlighted by the authors, this randomisation within families can help to establish genetic causation. Importantly, genetic causation is always indirect and happens entirely through (molecular and environmental) phenotypes. As such, rather than focusing on genetic causation, perhaps the greatest opportunity for building causal knowledge in behavioural genetics lies in leveraging the beautifully randomised process

mentioned by Fisher to understand phenotypic causation in general. To that end, a method called Mendelian randomisation uses genetic variants as instrumental variables to assess phenotypic causation and identify (potentially modifiable) risk factors (Richmond & Davey Smith, 2022; Sanderson et al., 2022; Smith & Ebrahim, 2003). Many Mendelian randomisation studies have focused on complex traits, including a within-family Mendelian randomisation study of the impact of educational attainment on physical health and mortality (Howe, Rasheed, et al., 2022). In addition, the authors rightly state that identifying a causal genetic variant often does not, *per se,* provide insights into causal pathways leading to behaviours. However, novel methods are rapidly developing that leverage genetic variants to understand causal pathways and mechanisms underlying complex phenotypes. Many of these methods directly harness the randomisation of genetic material at conception. For example, recent methods extend Mendelian randomisation to systematically investigate the aetiological role played by gene expression or DNA methylation (Hannon et al., 2018; Porcu et al., 2021). Genetically informed methods for causal inference aiming to identify (environmental) risk factors and biological pathways have been reviewed extensively elsewhere (Davey Smith et al., 2021; Pingault et al., 2021).

Critically, the analogy drawn between within-family genetic association studies and RCTs needs to be used with care. In an RCT, the treatment should be well defined (e.g. a given dose of a drug) and can be administered to individuals. By contrast, generally, a genetic variant cannot be administered or modified during the life course. Thus, while RCTs can provide actionable evidence of a specific intervention's efficacy, a within-family genetic association only indicates the effect of inheriting one variant or another. The difference in timing is also essential as inheriting a genetic variant at conception leads to a lifelong exposure as opposed to a time-bound treatment.

Furthermore, the authors argue that genetic causes relevant to behavioural genetics are analogous to causes uncovered by RCTs in that they are shallow –non-unitary (no single isolable cause), non-uniform (people exposed have heterogeneous outcomes), and non-explanatory (not mechanistically informative). However, this analogy is strained for both monogenic and polygenic disorders. A gene implicated in a monogenic disorder such as cystic fibrosis is close to a well-defined and actionable treatment: although the gene is not, in itself, a treatment, it can potentially be targeted by gene editing techniques such as CRISPR-Cas9 to correct deleterious variants (note that even for monogenic disorders, many deleterious variants can be involved and lead to different phenotypic manifestations) (Jinek et al., 2012). Successful human clinical trials using gene editing techniques for monogenic disorders are emerging (Frangoul et al., 2021). Many rare developmental disorders share a similar genetic architecture, some with relevance to behaviour. For example, adrenoleukodystrophy is a monogenic developmental disorder, which may first manifest with behavioural and cognitive difficulties and can be lethal (Zhu et al., 2020). Even for polygenic disorders such as schizophrenia, some rare variants considerably increase the risk of disease, with odds ratios up to 50 (Singh et al., 2022). As such, genetic causes relevant to behavioural genetics need not be shallow. Variants underlying monogenic disorders or high-risk rare variants may be better conceived as deep causes, i.e., close to unitary, uniform, and explanatory.

Conversely, polygenic influences of small effects underlying complex behaviours can indeed be conceived as shallow causes. However, in this case, the 'treatment' is not well defined (in content or timing), cannot be refined or changed to increase mechanistic insight and is not

directly actionable. Even if extensions of techniques such as CRIPSR-CAS9 could theoretically target hundreds of genetic variants at once, this could never be a treatment strategy given unknown and potentially devastating side effects (like a drug RCT consisting of the simultaneous administration of hundreds of compounds). In sum, causes established by RCTs and genetic causes derived from within-family association studies do not necessarily share many features beyond the core concept of randomisation. Further discussion of the notion of cause in genetics and the parallels between RCTs and genetically informed methods such as Mendelian randomisation are available elsewhere (Lynch, 2021; Nitsch et al., 2006).

In conclusion, we agree that behavioural genetics should look to provide causal knowledge. To that end, perhaps the most useful will be exploiting genetic data to understand phenotypic causation and aetiological pathways. Genetically informed designs for causal inference and, in particular, within-family designs, can play a key role in improving aetiological understanding and, ultimately, prevention and treatment (Howe, Nivard, et al., 2022; Hwang et al., 2021).

12. ACKNOWLEDGEMENTS

NA

13. CONFLICT OF INTEREST STATEMENT

NA

14. FUNDING STATEMENT

This research received no specific grant from any funding agency, commercial or not-for-profit sectors

15. REFERENCES

Davey Smith, G., Richmond, R., & Pingault, J.-B. (Eds.). (2021). *Combining Human Genetics and Causal Inference to Understand Human Disease and Development*. Cold Spring Harbor Laboratory Press,U.S.

Fisher, R. (1952). Statistical methods in genetics. *Heredity, 6*, 1-12 (reprinted in 2010, Int J Epidemiol 39: 329-335). https://doi.org/10.1093/ije/dyp379

Frangoul, H., Altshuler, D., Cappellini, M. D., Chen, Y.-S., Domm, J., Eustace, B. K., Foell, J., de la Fuente, J., Grupp, S., Handgretinger, R., Ho, T. W., Kattamis, A., Kernytsky, A., Lekstrom-Himes, J., Li, A. M., Locatelli, F., Mapara, M. Y., de Montalembert,

M., Rondelli, D., … Corbacioglu, S. (2021). CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *New England Journal of Medicine, 384*(3), 252–260. https://doi.org/10.1056/NEJMoa2031054

Hannon, E., Gorrie-Stone, T. J., Smart, M. C., Burrage, J., Hughes, A., Bao, Y., Kumari, M., Schalkwyk, L. C., & Mill, J. (2018). Leveraging DNA-Methylation Quantitative-Trait Loci to Characterize the Relationship between Methylomic Variation, Gene Expression, and Complex Traits. *American Journal of Human Genetics, 103*(5), 654–665. https://doi.org/10.1016/j.ajhg.2018.09.007

Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., Chittoor, G., Ahlskog, R., Lind, P. A., Palviainen, T., van der Zee, M. D., Cheesman, R., Mangino, M., Wang, Y., Li, S., Klaric, L., Ratliff, S. M., Bielak, L. F., Nygaard, M., … Davies, N. M. (2022). Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nature Genetics, 54*(5), 581–592. https://doi.org/10.1038/s41588-022-01062-7

Howe, L. J., Rasheed, H., Jones, P. R., Boomsma, D. I., Evans, D. M., Giannelis, A., Hayward, C., Hopper, J. L., Hughes, A., Lahtinen, H., Li, S., Lind, P. A., Martin, N. G., Martikainen, P., Medland, S. E., Morris, T. T., Nivard, M. G., Pingault, J.-B., Silventoinen, K., … Davies, N. M. (2022). *Educational attainment, health outcomes and mortality: A within-sibship Mendelian randomization study* (p. 2022.01.11.22268884). medRxiv. https://doi.org/10.1101/2022.01.11.22268884

Hwang, L.-D., Davies, N. M., Warrington, N. M., & Evans, D. M. (2021). Integrating Family-Based and Mendelian Randomization Designs. *Cold Spring Harbor Perspectives in Medicine, 11*(3), a039503. https://doi.org/10.1101/cshperspect.a039503

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Lynch, K. E. (2021). The Meaning of "Cause" in Genetics. *Cold Spring Harbor Perspectives in Medicine*, *11*(9), a040519. https://doi.org/10.1101/cshperspect.a040519

Nitsch, D., Molokhia, M., Smeeth, L., DeStavola, B. L., Whittaker, J. C., & Leon, D. A. (2006). Limits to causal inference based on Mendelian randomization: A comparison with randomized controlled trials. *American Journal of Epidemiology*, *163*(5), 397–403. https://doi.org/10.1093/aje/kwj062

Pingault, J.-B., Richmond, R., & Davey Smith, G. (2021). Causal Inference with Genetic Data: Past, Present, and Future. *Cold Spring Harbor Perspectives in Medicine*, a041271. https://doi.org/10.1101/cshperspect.a041271

Porcu, E., Sjaarda, J., Lepik, K., Carmeli, C., Darrous, L., Sulc, J., Mounier, N., & Kutalik, Z. (2021). Causal Inference Methods to Integrate Omics and Complex Traits. *Cold Spring Harbor Perspectives in Medicine*, *11*(5), a040493. https://doi.org/10.1101/cshperspect.a040493

Richmond, R. C., & Davey Smith, G. (2022). Mendelian Randomization: Concepts and Scope. *Cold Spring Harbor Perspectives in Medicine*, *12*(1), a040501. https://doi.org/10.1101/cshperspect.a040501

Sanderson, E., Glymour, M. M., Holmes, M. V., Kang, H., Morrison, J., Munafò, M. R., Palmer, T., Schooling, C. M., Wallace, C., Zhao, Q., & Davey Smith, G. (2022). Mendelian randomization. *Nature Reviews Methods Primers*, *2*(1), 1–21. https://doi.org/10.1038/s43586-021-00092-5

Singh, T., Poterba, T., Curtis, D., Akil, H., Al Eissa, M., Barchas, J. D., Bass, N., Bigdeli, T. B., Breen, G., Bromet, E. J., Buckley, P. F., Bunney, W. E., Bybjerg-Grauholm, J.,

Byerley, W. F., Chapman, S. B., Chen, W. J., Churchhouse, C., Craddock, N., Cusick, C. M., … Daly, M. J. (2022). Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature*, *604*(7906), 509–516. https://doi.org/10.1038/s41586-022-04556-w

Smith, G. D., & Ebrahim, S. (2003). 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, *32*(1), 1–22.

Zhu, J., Eichler, F., Biffi, A., Duncan, C. N., Williams, D. A., & Majzoub, J. A. (2020). The Changing Face of Adrenoleukodystrophy. *Endocrine Reviews*, *41*(4), 577. https://doi.org/10.1210/endrev/bnaa013