Mendelian Randomization: Progressing Towards Understanding Causality

Identifying and understanding factors that cause neurogenerative disorders such as Parkinson's disease (PD) is relevant for a number of reasons. First, the understanding of the pathobiology and etiology of PD is still limited because of the difficulties of conducting laboratory experiments or randomized clinical trials (RCTs), and knowledge of causal influences would give relevant insight into the underlying processes. Second, knowing parameters that lead to the development of the disease could yield plausible candidates for an early prevention of disease progression or therapy. Thus, the search for risk factors that are both modifiable and causal is a relevant endeavor. An interesting candidate for PD in this context is plasma urate levels, which are hypothesized to be a protective factor and were thus investigated in two contributions in this issue of *Annals of Neurology*.[1,2]

In the recent past, inferences about risk or protective factors outside from experimental settings have mostly been based on observed associations. However, it is well known how difficult it is to evaluate causality from mere associations, given that these can be, among others, a result of causation, reverse causation, confounding, and selection bias.

In a disease like PD, controlling for these effects may be especially problematic given that the limited understanding of the underlying pathobiology complicates the selection and thus control of confounders of associations. For example, urate levels have been shown to be negatively associated with PD, but are also known to be influenced by other factors such as sex, dietary habits, alcohol intake, smoking, and physical activity. Given that these parameters might also be associated with PD risk, it is unclear whether the observed negative association between urate levels and PD risk reflects causality or confounding.

To at least approximate causality despite these complications, the classical criteria defined by Bradford Hill are usually applied.[3] One of these criteria that makes an observation more likely to be causal is temporality, meaning that changes of the risk factor need to occur before the disease outcome. In PD, this criterion is difficult to establish, given that the actual onset of the disease is usually vague. Another criterion refers to the gold standard to infer causality, which is the experimental evidence. However, performing RCTs is again difficult for many risk factors because of ethical or practical reasons. For example, to derive causality for urate levels for the development of PD would, in the strict sense, require the experimental modification of urate levels in healthy probands with a follow-up long enough that a substantial number of these probands develop PD. More realistic RCTs are already being performed in which a urate precursor is being investigated as a disease-modifying therapy in PD patients,[4] but even if a therapeutic effect is shown, this does not answer the question about the causal effect of urate on the development of PD. Other criteria by Hill include the plausibility or the strength of the association and are easier to test, but they could be misleading given their subjectivity and vagueness.

A more recently suggested method to approach this problem is the Mendelian randomization (MR) design.[5–7] Basically, one or a number of genetic variants are identified that can act as instrumental variables for the risk factor in that they are associated with the risk factor, but have no direct effect on PD. Given that these are genetic variables, they have two distinct advantages: First, the alleles of the variants are distributed randomly at conception, so that, in effect, the predisposition for the risk factor is distributed randomly, thus approximating an experimental setting. Second, given that genetic variants are stable from conception on, they always precede other possibly confounding factors. This allows any relationship between these genetic variants and PD to be interpreted as evidence for causality of the risk factor on PD, if the following assumptions are fulfilled: (1) The genetic variants are associated with the risk factor; (2) there is no association between the genetic variants on the one hand and confounders of the relationship between risk factor and PD on the other; and (3) conditioning on the risk factor and possible confounders, there is no direct association between the genetic variants and PD.

Starting from well-established observed associations where the direction of the effect is not clear, or from inconclusive results, this principle has already been applied to explore the causality of a number of parameters on PD, thereby showing, for example, that increased iron levels are causally linked to a decreased risk of PD,[8] or that a higher

body mass index is causally linked to a lower risk of PD.[9] If the assumptions underlying MR are tested rigorously, finding evidence for causality through MR improves the understanding of PD and possibly opens new therapeutic avenues that can then be tested in RCTs, considering that the MR causal effect estimate may not correspond to the effect of a proposed intervention modifying the risk factor, given that it refers to the lifetime modification in the general population.

Even though the underlying MR principle is the same, a number of extensions have been developed on the methods that are reviewed in the literature (eg, see previous works[10–12]). The two studies presented in this issue reflect part of the spectrum of MR approaches that are utilized today: In the work by Kobylecki et al,[2] the MR was performed in a single sample using two genetic variants, both singly and in combination, with individual-level data on genetic variants, urate levels, PD, and a number of confounders. In contrast, the work by Kia et al[1] utilized a two-sample approach based on a score of 31 genetic variants with summary-level data. Another important distinction is that PD incidence was investigated in the first and PD prevalence in the second study. Despite these differences, neither study found evidence for a causal link between urate levels and lower PD risk.

Given the known observed negative association of urate levels and PD risk, how can this result from the MR studies be interpreted? How can we understand whether the null MR result is a lack of statistical evidence or a genuine negative finding? First, it always needs to be checked whether the statistical power of the MR study is large enough to identify a relevant causal effect. This clearly is an issue in many MR studies, given that the power vitally depends on the strength of the association between the genetic variants and the risk factor. And given that, in many instances, even a score of many genetic variants may not explain more than 5% to 10% of the variability of the risk factor, power in typical sample sizes is limited. Second, a judgment of the confidence interval of causal effect estimates and consistency between both MR studies could help. Third, in view of no evidence for causality in an MR study, alternative interpretations for the observed association should be sought, such as confounding or reverse causality, that may explain the observation. Further studies are then required to disentangle these relationships and explore whether urate levels might have a causal impact on PD progression instead of PD onset, to investigate also the reserve causation to understand whether PD could cause urate-level changes and, eventually, identify more plausible candidates for causality.

Acknowledgment

This work was supported by the Deutsche Forschungsgemeinschaft (FOR2488).

Potential Conflicts of Interest

Nothing to report.

Inke R. König PhD
Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Lübeck, Germany

Fabiola M. Del Greco PhD
Institute for Biomedicine, Eurac Research, Bolzano, Italy

References

1. Kia DA, Noyce AJ, White J, et al. Mendelian randomization study shows no causal relationship between circulating urate levels and Parkinson's disease. Ann Neurol 2018 Jul 16. doi: 10.1002/ana.25294. [Epub ahead of print]
2. Kobylecki CJ, Nordestgaard BG, Afzal S. Plasma urate and risk of Parkinson's disease: a Mendelian randomization study. Ann Neurol 2018 Jul 16. doi: 10.1002/ana.25292. [Epub ahead of print]
3. Hill AB. The environment and disease: association or causation? Proc R Soc Med 1965;58:295–300.
4. Parkinson Study Group SURE-PD Investigators, Schwarzschild MA, Ascherio A, et al. Inosine to increase serum and cerebrospinal fluid urate in Parkinson disease: a randomized clinical trial. JAMA Neurol 2014;71:141–150.
5. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003;32:1–22.
6. Ziegler A, Mwambi H, König IR. Mendelian randomization versus path models: making causal inferences in genetic epidemiology. Hum Hered 2015;79:194–204.
7. Grover S, Del Greco M F, Stein CM, Ziegler A. Mendelian randomization. Methods Mol Biol 2017;1666:581–628.
8. Pichler I, Del Greco M F, Gögele M, et al. Serum iron levels and the risk of Parkinson disease: a Mendelian randomization study. PLoS Med 2013;10:e1001462.
9. Noyce A, Kia D, Hemani G, et al. Estimating the causal influence of body mass index on risk of Parkinson disease: a Mendelian randomisation study. PLoS Med 2017;14:e1002314.
10. Zheng J, Baird D, Borges MC, et al. Recent developments in Mendelian Randomization studies. Curr Epidemiol Rep 2017;4:330–345.
11. Pingault JB, O'Reilly PF, Schoeler T, et al. Using genetic data to strengthen causal inference in observational research. Nat Rev Genet 2018 Jun 5. doi: 10.1038/s41576-018-0020-3. [Epub ahead of print]
12. Haycock PC, Burgess S, Wade KH, et al. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. Am J Clin Nutr 2016;103:965–978.

DOI: 10.1002/ana.25293