Journal of the Royal Society of Medicine

Capitalizing on Mendelian randomization to assess the effects of treatments

George Davey Smith

The online version of this article can be found at:
http://jrs.sagepub.com/content/100/9/432

Published by:

$SAGE

http://www.sagepublications.com

On behalf of:

The Royal Society of Medicine

Additional services and information for *Journal of the Royal Society of Medicine* can be found at:

Email Alerts: http://jrs.sagepub.com/cgi/alerts

Subscriptions: http://jrs.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Sep 1, 2007

What is This?

FROM THE JAMES LIND LIBRARY

Capitalizing on Mendelian randomization to assess the effects of treatments

George Davey Smith

Additional material for this article is available from the James Lind Library website [http://www.jameslindlibrary.org], where this paper was previously published.

The stage of disease and general fitness of patients usually influences the treatments they receive. Because of this, tests of treatments and treatment comparisons need to ensure that treatment comparison groups are made up of patients who are as similar as possible. Without this assurance, any differences in the progress of patients receiving different treatments cannot confidently be ascribed to differential effects of treatments: they may simply reflect differences in the characteristics of patients receiving the different treatments. This is the rationale underlying random allocation (randomization) to treatment groups in randomized trials—to ensure that whether treatment is given or withheld is unrelated to factors influencing the prognoses of patients.

Sometimes, for one or more of a variety of reasons, the organization of randomized controlled trials (RCTs) presents challenges that are judged likely to be very difficult to overcome. One such example relates to the need to assess the effects of bone marrow transplantation in the treatment of acute myeloid leukaemia in children. Randomized comparisons of treatment with and without bone marrow transplantation in this form of leukaemia have been judged unlikely to be achieved successfully, yet it is clearly important to obtain reliable, unbiased estimates of the effects of transplantation—wanted and unwanted—because the disadvantages may outweigh any advantages of this invasive treatment.

In 1991, Richard Gray and Keith Wheatley reported an ingenious method for obtaining unbiased estimates of the effects of bone marrow transplantation without conducting a traditional randomized trial.[1] They pointed out that unbiased comparisons could be made between child patients who had a genetically compatible sibling, and so in principle could receive a matched sibling bone marrow transplant, with other child patients who had no genetically compatible

sibling, and so were incapable of receiving a matched sibling bone marrow transplant. Because having or not having a genetically compatible sibling is a matter of chance—it is determined by random assortment of genes at the time of gamete formation and conception—this situation produces what is effectively a randomized comparison.

Whether a child with leukaemia belongs to the group with genetically compatible siblings or the group without such siblings will not be related to potential confounding factors such as disease stage and general fitness at the time of diagnosis. In a form of 'intention-to-treat analysis', Gray and Wheatley noted that an unbiased comparison meant comparing all the patients with a genetically compatible sibling with all the patients without such a sibling, regardless of whether or not every patient with a potential donor actually received a transplant (Figure 1).[1]

Gray and Wheatley appear to have been the first authors to refer to this particular way of capitalizing on chance events in nature to create unbiased comparison groups to assess the effects of a treatment. They called it 'Mendelian randomization'. Several further studies have now been carried out using their design,[2–4] including studies to assess the effects of treatment for acute lymphoblastic leukaemia.[5–7] Some of these studies have confirmed that like is being compared with like in comparison groups defined in this way.[3–6] They have also shown that there are differences in prognostic factors between groups defined by the treatment they received; differences that would confound a conventional observational analysis comparing different treatments.[4–6]

The basic design outlined in Figure 1 might be improved by taking into account the number of siblings each patient has.[8] Patients with more siblings have a greater chance of having a genetically compatible donor, and therefore groups defined by having a compatible donor will differ according to average number of siblings and thus will differ by factors that may be related to prognosis. Indeed, a later study found that the number of siblings could itself be related to survival.[9] The study by Gray and Wheatley (Gray, personal communication) and another more recent study[4] applied an analysis restricted to patients with at least one sibling, but exact stratification or matching on number of siblings might be a more robust approach.
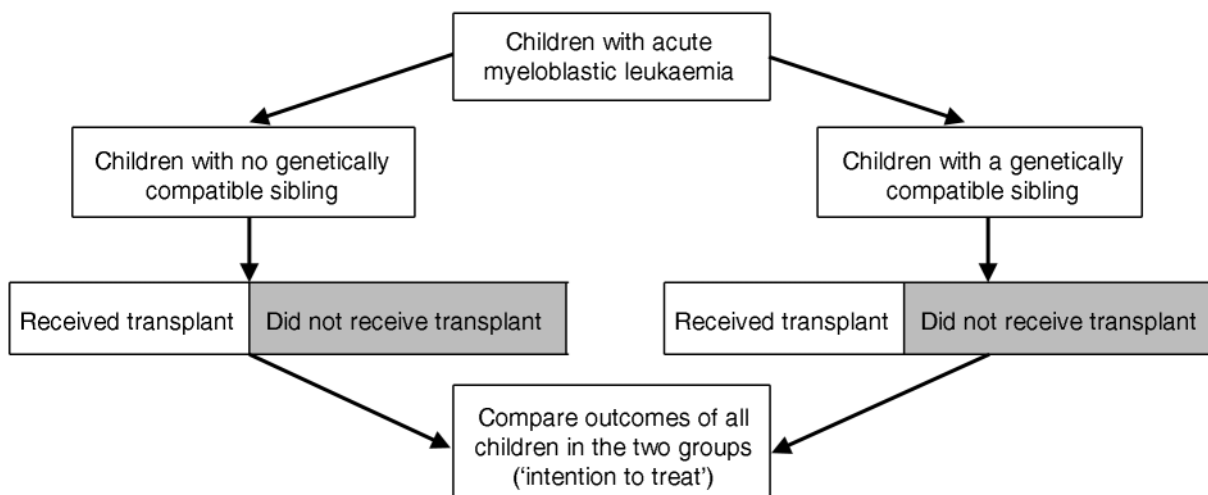
MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol

Email: george.davey-smith@bristol.ac.uk

Figure 1 **Intention to treat analysis of children with leukaemia**

MENDELIAN RANDOMIZATION IN OTHER SETTINGS

Mendelian randomization as a term has also recently been used to describe methods for obtaining unbiased estimates of causal associations in observational studies in aetiological epidemiology. In part, these developments reflect a response to well-publicized cases of conventional observational epidemiology having produced misleading information about supposed causal or protective factors, for example, for β carotene and cancer, vitamins C and E and coronary heart disease, and hormone replacement therapy and cardiovascular disease. Observational epidemiological studies suggested that these factors had important protective effects, but randomized controlled trials failed to confirm this. The probable reason for these discrepancies between analyses of observational data and randomized trials is that there is considerable confounding between, for example, dietary vitamin C or E intake or taking β carotene supplements and various behavioural and socio-economic factors related to increased risks of disease.

Mendelian randomization—the random assortment of genes from parents to offspring which occurs during gamete formation and conception—provides a method of assessing whether certain environmental exposures are causally related to a disease. The association between risk of a disease and a genetic variant that influences the exposure or mimics the biological link between a proposed exposure and disease is not generally susceptible to the 'reverse causation' or confounding that may distort interpretations of conventional observational studies.

THE CONTRIBUTION OF MARTIJN KATAN

A Dutch researcher, Martijn Katan, was an early exponent of what has since become termed Mendelian randomization.

He was concerned that observational studies were suggesting that low serum cholesterol levels were associated with an increased risk of cancer, and thus that treatment to lower cholesterol could have detrimental effects.[10] This association might be explained by early cancer lowering cholesterol levels (reverse causality) or by confounding factors (such as cigarette smoking), both of which are related to future cancer risk and to lower circulating cholesterol levels.[11]

Katan pointed out[12] that polymorphic forms of the apolipoprotein ε (APOE) gene were related to different average levels of serum cholesterol. If low circulating cholesterol levels were indeed a causal factor for cancer, individuals with the genotype associated with lower average cholesterol levels should be expected to have higher cancer risk. If, however, reverse causation or confounding generated the association between low cholesterol levels and cancer, then no association would be expected between the APOE genotype and cancer. Individuals with lower cholesterol because of their genotype, rather than because existing clinically unrecognized cancers had lowered their cholesterol, would not have a higher risk of cancer; nor would there be substantial confounding between genotype-associated differences in cholesterol levels and lifestyle or socio-economic factors. Although Katan did not have any data to investigate these possibilities, he advocated a study design taking these considerations into account.[12]

As far as I am aware, Katan's intriguing suggestion has not yet been explicitly applied to address the important question he posed, although sporadic reports relating APOE to risk of specific cancers continue to appear.[13–17] However, several examples where the phenotypic effects of polymorphisms are well-documented provide encouraging evidence of the explanatory power of Mendelian randomization.[8,18,19]

433

WHY THE TERM 'MENDELIAN RANDOMIZATION'?

Gregor Mendel (1822–1884) concluded from his hybridization studies with pea plants that 'the behaviour of each pair of differentiating characteristics [such as the shape and colour of seeds] in hybrid union is independent of the other differences between the two original plants'.[20] This formulation was actually the only regularity that Mendel referred to as a 'law'. In Carl Correns' 1900 paper (one of a trio appearing that year which are considered to represent the rediscovery of Mendel) he refers to this as Mendel's Law.[21,22]

Morgan[23] discusses independent assortment and refers to this process as being realized 'whenever two pairs of characters freely Mendelize'. Morgan's use of Mendel's surname as a verb did not catch on, but Morgan later christened this Mendel's Second Law,[24] and it has been known by this name, or as 'The Law of Independent Assortment', ever since. The law suggests that inheritance of one trait is independent of—that is, randomized with respect to—the inheritance of other traits.

The analogy with a randomized controlled trial will clearly be most applicable to parent–offspring designs investigating the frequency with which one of two alleles from a heterozygous parent is transmitted to offspring with a particular disease. However, at a population level, traits influenced by genetic variants are generally not associated with the social, behavioural and environmental factors that confound relationships observed in conventional epidemiological studies. Thus while the 'randomization' is approximate and not absolute in genetic association studies, empirical observations suggest that it applies in most circumstances.[25,26]

As discussed above, the term 'Mendelian randomization' itself was introduced by Gray and Wheatley in a somewhat different context, in which advantage is taken of the random assortment of genetic variants at conception to provide an unconfounded study design for estimating treatment effects for childhood malignancies.[1,27] However, the term has recently become widely used with the meaning ascribed to it here.

The notion that genetic variants can serve as an indicator of the action of environmentally modifiable exposures has been expressed in many contexts. For example, since the mid-1960s various investigators have pointed out that the autosomal dominant condition of lactase persistence is associated with milk drinking. Associations of lactase persistence with osteoporosis, bone mineral density or fracture risk thus provide evidence that milk drinking protects against these conditions.[28,29] In a related vein, it was proposed in 1979 that as N-acetyltransferase pathways are involved in the detoxification of arylamine, a potential bladder carcinogen, the observation of increased bladder cancer risk among people with genetically determined slow acetylator phenotype provided evidence that arylamines are involved in the aetiology of the disease.[30]

Since that time various commentators have pointed out that the associations of genetic variants of known function with disease outcomes provides evidence about aetiological factors.[31–35] However, these commentators have not emphasized the key strengths of Mendelian randomization: the avoidance of confounding, bias due to reverse causation or reporting tendency, and the underestimation of risk associations due to variability in behaviours and phenotypes.[18]

These key concepts were present in Martijn Katan's 1986 *Lancet* letter in which he suggested that genetic variants related to cholesterol level could be used to investigate whether the observed association between low cholesterol and increased cancer risk was real,[12] and by Honkanen and colleagues in their understanding of how lactase persistence could better characterize the difficult-to-measure environmental influence of calcium intake than could direct dietary reports.[36] Since 2000 there have been several reports using the term 'Mendelian randomization' in the way it is used here,[8,37–40] and its use is becoming widespread. The fact that Mendelian randomization is one of a family of techniques referred to as 'instrumental variable' approaches for obtaining robust causal inferences from observational data has also been recognized,[41] and statistical approaches to instrumental variables analysis developed within econometrics have been applied to data from Mendelian randomization studies.[25,42]

REFERENCES

1 Gray R, Wheatley K. How to avoid bias when comparing bone marrow transplantation with chemotherapy. *Bone Marrow Transplant* 1991;**7**(Suppl 3):9–12

2 Ljungman P, De Witte T, Verdonck L, *et al.* Bone marrow transplantation for acute myeloblastic leukaemia: an EBMT Leukaemia Working Party prospective analysis from HLA-typing. *Br J Haematol* 1993;**84**:61–66

3 Keating S, de Witte T, Suciu S, *et al.* The influence of HLA-matched sibling donor availability on treatment outcome for patients with AML: an analysis of the AML 8A study of the EORTC Leukaemia Cooperative Group and GIMENA. *Br J Haematol* 1998;**102**:1344–53

434

4   Burnett AK, Wheatley K, Goldstone AH, *et al*. The value of allogenic bone marrow transplant in patients with acute myeloid leukaemia at differing risk of relapse: results of the UK MRC AML 10 trial. *Br J Haematol* 2002;**118**:385–400

5   Harrison G, Richards S, Lawson S, *et al*. on behalf of the MRC Childhood Leukaemia Working Party. Comparison of allogeneic transplant versus chemotherapy for relapsed childhood acute lymphoblastic leukaemia in the MRC UKALL R1 trial. *Ann Oncol* 2000;**11**:999–1006

6   Bleakley M, Shaw PJ, Nielsen JM. Allogenic bone marrow transplantation for childhood relapsed acute lymphoblastic leukaemia: comparison of outcome in patients with and without a matched family donor. *Bone Marrow Transplant* 2002;**31**:1–7

7   Balduzzi A, Valsecchi MG, Uderzo C, *et al*. Chemotherapy versus allogenic transplantation for very-high-risk childhood acute lymphoblastic leukaemia in first complete remission: comparison by genetic randomization in an international prospective study. *Lancet* 2005;**366**:635–42

8   Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32**:1–22

9   Charalampopoulou A, Petridou E, Spyridopoulos T, *et al*. An integrated evaluation of socioeconomic and clinical factors in the survival from childhood acute lymphoblastic leukaemia: a study in Greece. *Eur J Cancer Prev* 2004;**13**:397–401

10 Garcia-Palmiere MR, Sorlie PD, Costas R, Havlik RJ. An apparent inverse relationship between serum cholesterol and cancer mortality in Puerto Rico. *Am J Epidemiol* 1981;**114**:29–40

11 McMichael AJ, Jensen OM, Parkin DM, Zaridze DG. Dietary and endogenous cholesterol and human cancer. *Epidemiol Rev* 1984;**6**:192–216

12 Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet* 1986;**1**:507–8

13 Wessel N, Liestol K, Maehlen J, Brorson SH. The apolipoprotein E epsilon4 allele is no risk factor for prostate cancer in the Norwegian population. *Br J Cancer* 2001;**85**:1418

14 Niemi M, Kervinen K, Kiviniemi H, *et al*. Apolipoprotein E phenotype, cholesterol and breast and prostate cancer. *J Epidemiol Community Health* 2000;**54**:938–9

15 Moysich KB, Freudenheim JL, Baker JA, *et al*. Apolipoprotein E genetic polymorphism, serum lipoproteins, and breast cancer risk. *Mol Carcinog* 2000;**27**:2–9

16 Slattery ML, Sweeney C, Murtaugh M, *et al*. Associations between apoE genotype and colon and rectal cancer. *Carcinogenesis* 2005;**26**:1422–9

17 Watson MA, Gay L, Stebbings WSL, Speakman CTM, Bingham SA, Loktionov A. Apolipoprotein E gene polymorphism and colorectal cancer: gender-specific modulation of risk and prognosis. *Clin Sci* 2003;**104**:537–45

18 Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004;**33**:30–42

19 Lewis S, Davey Smith G. Alcohol, ALDH2 and esophageal cancer: a meta-analysis which illustrates the potentials and limitations of a Mendelian randomization approach. *Cancer Epidemiol Biomark Prev* 2005;**14**:1967–71

20 Mendel G. Experiments in plant hybridization. *Proceedings of the Natural History Society of Brünn* 1866

21 Correns C. G Mendel's Regel über das Verhalten der Nachkommenschaft der Bastarde. *Berichte der Deutschen Botanischen Gesellschaft* 1900; **8**: 158–68. First published in English as: G Mendel's law concerning the behavior of progeny of varietal hybrids. *Genetics* 1950;**35**:33–41

22 Olby RC. *Origins of Mendelism*. London: Constable, 1966

23 Morgan TH. *Heredity and Sex*. New York: Columbia University Press, 1913

24 Morgan TH. *Physical Basis of Heredity*. Philadelphia: JB Lipincott Company, 1919

25 Davey Smith G, Lawlor D, Harbord R, *et al*. Association of C-reactive protein with blood pressure and hypertension: lifecourse confounding and Mendelian randomization tests of causality. *Arterioscler Thromb Vasc Biol* 2005;**25**:1051–6

26 Bhatti P, Sigurdson AL, Wang SS, *et al*. Genetic variation and willingness to participate in epidemiological research: data from three studies. *Cancer Epidemiol Biomark Prev* 2005;**14**:2449–53

27 Wheatley K, Gray R. Commentary: Mendelian randomization—an update on its use to evaluate allogeneic stem cell transplantation in leukaemia. *Int J Epidemiol* 2004;**33**:15–7

28 Birge SJ, Keutmann HT, Cuatrecasas P, Whedon GD. Osteoporosis, intestinal lactase deficiency and low dietary calcium intake. *NEJM* 1967;**276**:445–8

29 Newcomer AD, Hodgson SF, Douglas MD, Thomas PJ. Lactase deficiency: prevalence in osteoporosis. *Ann Intern Med* 1978;**89**:218–20

30 Lower GM, Nilsson T, Nelson CE, Wolf H, Gamsky TE, Bryan GT. N-acetylransferase phenotype and risk in urinary bladder cancer: approaches in molecular epidemiology. *Environ Health Perspect* 1979;**29**:71–9

31 McGrath J. Hypothesis: is low prenatal vitamin D a risk-modifying factor for schizophrenia? *Schizophr Research* 1999;**40**:173–7

32 Ames BN. Cancer prevention and diet: Help from single nucleotide polymorphisms. *Proc Nat Acad Sci* 1999;**96**:12216–8

33 Rothman N, Wacholder S, Caporaso NE, Garcia-Closas M, Buetow K, Fraumeni JF. The use of common genetic polymorphisms to enhance the epidemiologic study of environmental carcinogens. *Biochim Biophys Acta* 2001;**1471**:C1–C10

34 Brennan P. Gene environment interaction and aetiology of cancer: what does it mean and how can we measure it? *Carcinogenesis* 2002;**23**:381–7

35 Kelada SN, Eaton DL, Wang SS, Rothman NR, Khoury MJ. The role of genetic polymorphisms in environmental health. *Environ Health Perspect* 2003;**111**:1055–64

36 Honkanen R, Pulkkinen P, Järvinen R, *et al*. Does lactose intolerance predispose to low bone density? A population-based study of perimenopausal Finnish women. *Bone* 1996;**19**:23–8

37 Youngman LD, Keavney BD, Palmer A, *et al*. Plasma fibrinogen and fibrinogen genotypes in 4685 cases of myocardial infarction and in 6002 controls: test of causality by 'Mendelian randomization'. *Circulation* 2000;**102**(Suppl II):31–2

38 Fallon UB, Ben-Shlomo Y, Elwood P, Ubbink JB, Davey Smith G. Homocysteine and coronary heart disease—Author's reply. *Heart* published online 14 March 2001. Available at http://heart.bmjjournals.com/cgi/eletters/85/2/153

39 Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet* 2001;**358**:1556–60

40 Keavney B. Genetic epidemiological studies of coronary heart disease. *Int J Epidemiol* 2002;**31**:730–6

41 Thomas DC, Conti DV. The concept of 'Mendelian Randomization'. *Int J Epidemiol* 2004;**33**:21–5.

42 Timpson NJ, Lawlor DA, Harbord RM, *et al*. C-reactive protein and its role in metabolic syndrome: mendelian randomization study. *Lancet* 2005;**366**:1954–9

435