Published in final edited form as:

JAMA. 2009 June 10; 301(22): 2386–2388. doi:10.1001/jama.2009.812.

Mendelian Randomization: Nature's Randomized Trial in the Post-Genome Era?

George Thanassoulis, MD^{1,3} and Christopher J. O'Donnell, MD MPH^{1,2,3}

¹National Heart, Lung and Blood Institute's Framingham Heart Study

²National Heart, Lung and Blood Institute Division of Intramural Research

³Boston University Medical Center

⁴Cardiology Division, Department of Medicine, Massachusetts General Hospital

Despite several observational studies showing that lipoprotein(a), or Lp(a), is associated with myocardial infarction (MI)^{1, 2}, only circumstantial evidence exists regarding the causal nature of this association. Observational epidemiological studies, even with a sound prospective design, can provide hints to disease pathogenesis when the effect size is modest but cannot provide definitive evidence for causal relationships. Much of the current understanding of the causal factors in cardiovascular disease, such as the role of LDL, has been confirmed by randomized clinical trials (RCT)^{3, 4}. However, RCTs are not always feasible. In the case of Lp(a), the modest effect size and the lack of specific Lp(a) lowering therapy are major obstacles to obtaining causal evidence for its role in cardiovascular disease. In this issue of JAMA, Kamstrup and colleagues⁵ provide insights using a "Mendelian randomization" approach and provide evidence for the causal role of Lp(a) in MI. This study elegantly demonstrates how Mendelian randomization can be used to improve the evidence for causality from observational studies and highlights the advantages and limitations of such an approach.

The study by Kamstrup et al., can be viewed as a form of natural "randomized trial" of plasma Lp(a) level on myocardial infarction, with a few important differences from traditional clinical trials. Instead of using drug therapy to decrease plasma Lp(a) levels, genetic variation in the lipoprotein apo(a) gene, *LPA*, that controls plasma levels of Lp(a) are used. "Randomization", which minimizes the effect of confounding variables, is achieved through the random assortment of these *LPA* gene variants from parents to offspring that occurs during gamete formation and conception. In this way, because *LPA* variants are randomly assigned, they are unlikely to associate with other nongenetic factors, such as lifestyle factors, that could act as confounders, thereby providing an unbiased assessment of the role of Lp(a) in MI.

Lp(a) represents an excellent example of a genetic exposure that can be evaluated using the Mendelian randomization approach. Such an approach requires highly accurate quantitation of the triangulation of three associations: *LPA* gene variants and serum Lp(a) levels, serum Lp(a) levels and myocardial infarction, and *LPA* gene variants and myocardial infarction. Much of the population variation (30–60%) in Lp(a) levels is explained by the number of copy number variants, known as KIV-2 repeats, a form of structural variation, in the *LPA* gene^{5, 6}. With increasing number of KIV-2 repeats in the *LPA* gene, plasma levels of Lp(a) are reduced. In the Copenhagen City Heart Study (CCHS), the authors demonstrate that this

important relationship holds from early adulthood to old age, confirming that individuals with increased KIV-2 repeats have lifelong lower plasma levels of Lp(a). Furthermore, the previously established association between plasma Lp(a) levels and myocardial infarction is also confirmed in the same population^{1, 2}. The authors also demonstrate that *LPA* gene variants are unequivocally associated with MI, an association that has been reported by other groups⁷⁻¹².

The major strengths of this study include the estimation of the three key associations required for Mendelian randomization within the same population, the replication of these associations in several independent cohorts, and the use of instrumental variable analysis, frequently used in econometrics, to demonstrate that a genetically determined doubling of Lp(a) plasma levels leads to a 22% increase in the risk of MI. Since Lp(a) levels are stable with increasing age, this represents the best estimate for the largely unconfounded effect of lifelong Lp(a) plasma levels on myocardial infarction.

Despite the many advantages of this approach, a number of considerations can invalidate the results of such studies¹³. For example, as in a RCT, the randomization process can fail leading to biased estimates of effect. In this case, genetic variation did not appear to be related to the major cardiovascular covariates given the well balanced distribution of covariates across genotypes. In addition, population stratification, a form of confounding by ethnicity, can also lead to biased estimates. Simply put, if both the frequency of *LPA* alleles were higher in certain ethnic subgroups and the incidence of myocardial infarction was also higher in these same subgroups, then this would lead to spurious associations between the *LPA* gene variants and MI. However, as the authors point out, this type of confounding is unlikely in the relatively homogeneous populations studied. Also, if the *LPA* gene has pleiotropic effects on myocardial infarction not mediated by Lp(a) plasma levels, the instrumental variable analysis for the effect estimates of plasma Lp(a) lowering on MI will be biased. In fact, such pleiotropic effects may exist for the *LPA* gene as the KIV-2 variants affect both Lp(a) plasma levels and Lp(a) isoform size. If smaller Lp(a) isoforms, independent of plasma Lp(a) levels, are true risk factors for MI as has been suggested,^{9, 14} the increase in MI risk attributed to higher plasma Lp(a) by Kamstrup et al, would be overestimated, due to the combined effects of smaller Lp(a) isoforms and higher plasma Lp(a) concentrations in individuals with fewer KIV-2 repeats.

Taken in the context of previous observational studies of Lp(a), the study by Kamstrup *et al*, adds the strongest evidence to date that lifelong increased Lp(a) plasma levels are causally related to MI. This is an important biological finding that elevates the status of Lp(a), as a biomarker for MI, from putative risk marker to confirmed causal factor and that stimulate renewed interest in the biology of Lp(a) and its role in cardiovascular disease. While this study certainly provides interesting mechanistic insights into the biology of Lp(a) in the context of MI and suggestive evidence regarding the potential benefit of decreasing Lp(a) early in life, clinicians may ask: "How will these results affect current approaches for prediction, prevention and treatment of my patients?" At present, the clinical implications remain quite limited. These results do not provide the necessary evidence that genetic testing of the *LPA* locus or measurements of plasma Lp(a) have a role in cardiovascular risk stratification or decisions regarding lipid-lowering therapy. Ultimately, despite nature's best efforts to provide causal evidence for Lp(a), only a true RCT demonstrating reductions in MI with targeted Lp(a) lowering therapy can provide the evidence for benefits and risks of a Lp(a) lowering strategy.

There is currently no well tolerated drug that can specifically reduce Lp(a) levels. Even if such a drug were to be developed, the modest 22% increase in risk with a lifelong doubling of plasma Lp(a) raises the question whether a shorter duration of Lp(a) lowering achieved

by pharmacotherapy in adulthood would be effective in decreasing the risk of MI. By comparison, lifelong LDL reductions of only 15% have been reported to produce a 47% reduction in risk of cardiovascular disease¹⁵ illustrating that, as a life-long cardiovascular risk factor and as a pharmacologic target, LDL is likely much more potent than Lp(a). These results only provide evidence that lower levels of Lp(a) levels throughout life reduce the risk of MI. It is not known whether reductions later in life when MI risk is highest and when most individuals would presumably be treated, would lead to similar reductions in the risk of MI.

In the short number of years following the complete elucidation of the sequence of all 3 billion basepairs in the human genome, a remarkable explosion of genomewide association studies has provided evidence for common genetic variants numbering in the thousands—single nucleotide polymorphisms (SNPs) as well as copy number polymorphisms (CNPs)—that underlie hundreds of diseases as well as quantitative disease traits that are diagnosed and treated daily by readers of JAMA¹⁶. To this catalogue of robust and well-replicated genetic associations, a huge number of lower frequency and very rare genetic variants identified by whole genome sequencing will soon be added. As the number of genetic associations continues to increase, clinicians will be increasingly confronted with Mendelian randomization designs and should be aware that successful Mendelian randomization studies, such as the Lp(a) association with MI reported here, may be particularly uncommon and many such studies may fail to identify true causal relationships, even when they exist, for a number of reasons. First, because the magnitude of effect is small to modest for most associations between SNPs and diseases, the sample size required for adequately powered Mendelian randomization studies can be prohibitively large (>10,000 subjects) and many studies will be grossly underpowered. Second, the specific causal SNP and the mechanism of most SNP associations are often not obvious at the discovery stage and require further experimental work to identify the causal SNPs and the intermediate phenotypes with which to conduct these studies. Weak correlations between SNPs and intermediate phenotype, or between discovery SNPs and truly causal variants, will significantly attenuate estimated effect sizes. Third, most GWAS to date have been conducted in middle age and older adults, and the absence of an association between a SNP and an intermediate phenotype or a disease outcome may result from the influence of multiple cumulative environmental effects in older age and other gene-environment or gene-gene interactions that dilute a modest but real genetic effect that may be more apparent in early life.

Despite these important limitations, large-scale Mendelian randomization will be increasingly used given the success of GWAS and the ease of genotyping in large, prospective cohorts such as CCHS and many other cohorts around the world. Thus far, the Mendelian randomization approach has successfully been used in cardiovascular disease to raise questions about any causal role for CRP¹⁷, to confirm the causal role of LDL¹⁵, and now, with the study by Kamstrup et al., to demonstrate the causal role of Lp(a). Given the unique contributions of Mendelian randomization to the understanding of biology, this approach will continue to provide one avenue for the evaluation of causal associations. Such studies will continue to demand careful interpretation, particularly when findings are negative, and any positive results will need to be placed in the appropriate biological and clinical context. While nature's randomized trials may provide a window to evaluate causality, confirmatory evidence from human-made RCTs will continue to be required to inform clinical practice.

Acknowledgments

Dr. Thanassoulis holds a research fellowship from the Canadian Institute of Health Research and the Fonds de Recherche en Santé Québec. The authors gratefully acknowledge Dr. Jacques Genest and Dr. Sekar Kathiresan for their helpful comments during the preparation of this manuscript.

References

1. Danesh J, Collins R, Peto R. Lipoprotein(a) and Coronary Heart Disease: Meta-Analysis of Prospective Studies. *Circulation*. 2000; 102(10):1082–1085. [PubMed: 10973834]
2. Bennet A, Di Angelantonio E, Erqou S, et al. Lipoprotein(a) Levels and Risk of Future Coronary Heart Disease: Large-Scale Prospective Data. *Arch Intern Med*. 2008; 168(6):598–608. [PubMed: 18362252]
3. Downs JR, Clearfield M, Weis S, et al. Primary Prevention of Acute Coronary Events With Lovastatin in Men and Women With Average Cholesterol Levels: Results of AFCAPS/TexCAPS. *JAMA*. 1998; 279(20):1615–1622. [PubMed: 9613910]
4. Shepherd J, Cobbe SM, Ford I, et al. Prevention of Coronary Heart Disease with Pravastatin in Men with Hypercholesterolemia. *N Engl J Med*. 1995; 333(20):1301–1308. [PubMed: 7566020]
5. Kamstrup PR, Tybjaerg-Hansen A, Steffensen R, Nordestgaard G. Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. *JAMA*. 2009
6. Boerwinkle E, Leffert CC, Lin J, Lackner C, Chiesa G, Hobbs HH. Apolipoprotein(a) gene accounts for greater than 90% of the variation in plasma lipoprotein(a) concentrations. *J Clin Invest*. 1992; 90(1):52–60. [PubMed: 1386087]
7. Chasman DI, Shiffman D, Zee RYL, et al. Polymorphism in the apolipoprotein(a) gene, plasma lipoprotein(a), cardiovascular disease, and low-dose aspirin therapy. *Atherosclerosis*. 2009; 203(2): 371–376. [PubMed: 18775538]
8. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet*. 2009; 41(1):56–65. [PubMed: 19060906]
9. Kronenberg F, Kronenberg MF, Kiechl S, et al. Role of Lipoprotein(a) and Apolipoprotein(a) Phenotype in Atherogenesis: Prospective Results From the Bruneck Study. *Circulation*. Sep 14; 1999 100(11):1154–1160. [PubMed: 10484534]
10. Luke MM, Kane JP, Liu DM, et al. A Polymorphism in the Protease-Like Domain of Apolipoprotein(a) Is Associated With Severe Coronary Artery Disease. *Arterioscler Thromb Vasc Biol*. 2007; 27(9):2030–2036. [PubMed: 17569884]
11. Shiffman D, O'Meara ES, Bare LA, et al. Association of Gene Variants With Incident Myocardial Infarction in the Cardiovascular Health Study. *Arterioscler Thromb Vasc Biol*. 2008; 28(1):173–179. [PubMed: 17975119]
12. Tregouet D-A, König IR, Erdmann J, et al. Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. *Nat Genet*. 2009; 41(3):283–285. [PubMed: 19198611]
13. Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum Genet*. 2008; 123(1):15–33. [PubMed: 18038153]
14. Wu HD, Berglund L, Dimayuga C, et al. High lipoprotein(a) levels and small apolipoprotein(a) sizes are associated with endothelial dysfunction in a multiethnic cohort. *Journal of the American College of Cardiology*. 2004; 43(10):1828–1833. [PubMed: 15145108]
15. Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence Variations in PCSK9, Low LDL, and Protection against Coronary Heart Disease. *N Engl J Med*. 2006; 354(12):1264–1272. [PubMed: 16554528]
16. Hindorff, L.; Junkins, H.; Mehta, J.; Manolio, T. [Accessed May 5th 2009] A Catalog of Published Genome-Wide Association Studies. www.genome.gov/26525384
17. Zacho J, Tybjaerg-Hansen A, Jensen JS, Grande P, Silleesen H, Nordestgaard BG. Genetically Elevated C-Reactive Protein and Ischemic Vascular Disease. *N Engl J Med*. 2008; 359(18):1897–1908. [PubMed: 18971492]