

Twins

novel uses to study complex traits and genetic diseases

The challenge faced by research into the genetic basis of complex disease is to identify genes of small relative effect against a background of substantial genetic and environmental variation. This has focused interest on a classical epidemiological design: the study of twins. Through their precise matching for age, the common family environment and background environmental variation, studying diseases in non-identical twins provides a means to enhance the power of conventional strategies to detect genetic influence through linkage and association. The unique matching of identical twins provides researchers with ways to isolate the function of individual genes involved in disease together with approaches to understanding how genes and the environment interact.

The classical twin study is established as the definitive study design for investigating the relative importance of genetic and environmental factors to traits and diseases in human populations. Twins have not, however, been regarded traditionally as useful in the search for individual genes. Instead, study methods have focused on family pedigrees and affected sibling pairs to assess linkage and on case-control studies that have been conducted in the general population to assess association. These approaches have achieved success with monogenic diseases that exhibit mendelian patterns of inheritance and for disease associations that involve genes with large effect. When applied to the so-called 'complex' diseases, in which combinations of genetic, constitutional and environmental factors are all presumed to contribute, these methods have encountered problems. Foremost among these have been a lack of power, a tendency towards biases that lead to inconsistent results, and a lack of flexibility to account for the full complexity of disease traits and underlying disease processes.

These limitations have stimulated a search for new analytical approaches, and have prompted a fresh look at the potential contribution of twins. Increasingly, it is now recognized that the twin study offers several features that uniquely enhance our ability to localize genes and understand their function¹. In this review, we consider the value of twins in enhancing existing strategies for gene detection in complex disease (Fig. 1) and consider future use of the twin design in gene discovery.

The classical twin study

Twin studies exploit the unique degree of genetic and environmental sharing among the two types of twin pair: monozygotic (MZ) twins, who share a common set of genes, and dizygotic (DZ) twins, who share on average only 50% of their genes². In addition, both MZ and DZ twins share the same uterus, birth date, age and aspects of their early and later environment. These features allow the population-level variation of traits and diseases to be separated into genetic, shared environmental and random environmental components. Separating shared genes from the shared family environment as causes of family resemblance is difficult to achieve in any other practical family

design. By applying quantitative analytical techniques to twin data it is possible to estimate the size of the contribution of these individual components of variation. This provides an estimate of 'heritability': a measure of the extent to which phenotypic variation in the population can be explained by genetic variation³.

Studies of twins have been used widely to investigate the heritability of common complex diseases. The results have consistently shown an important contribution from genetic variation to disease susceptibility (Table 1). This has changed our perspective of many diseases, which were previously thought to be influenced predominantly by the environment.

The advantages of using twins for association and linkage studies

The presence of a large heritable component underlying most common complex diseases has understandably generated enthusiasm that their genetic basis can be resolved at a molecular level. The chief strategies that have been employed to identify genes that are involved in diseases have entailed studies of linkage, which examines the co-segregation of genes and disease within families, and of association, which examines the joint occurrence of genes and diseases in populations. Despite advances in experimental genetics, progress in identifying individual causal genes for common complex diseases has been slow⁴. For several diseases – asthma⁵ and schizophrenia⁶ are two examples – several genetic linkages have been published, providing a picture that is difficult to interpret. Association studies of candidate genes have also shown inconsistent results for several diseases. For example, in osteoporosis, an association with the gene that encodes the vitamin D receptor⁷, has not been replicated in several other samples, which have variously shown absent, positive and negative associations with the disease^{8,9}. These difficulties have highlighted the formidable problems that are involved in tackling the genetic basis of complex disease and the shortcomings of classical study methods and designs.

The limitations of linkage analysis in complex traits and diseases have been well described¹⁰. Methods that are based on pedigrees require specification of the mode of inheritance, and are inapplicable to most common complex

Alex J. MacGregor
alex.macgregor@kcl.ac.uk

Harold Snieder
harold.snieder@kcl.ac.uk

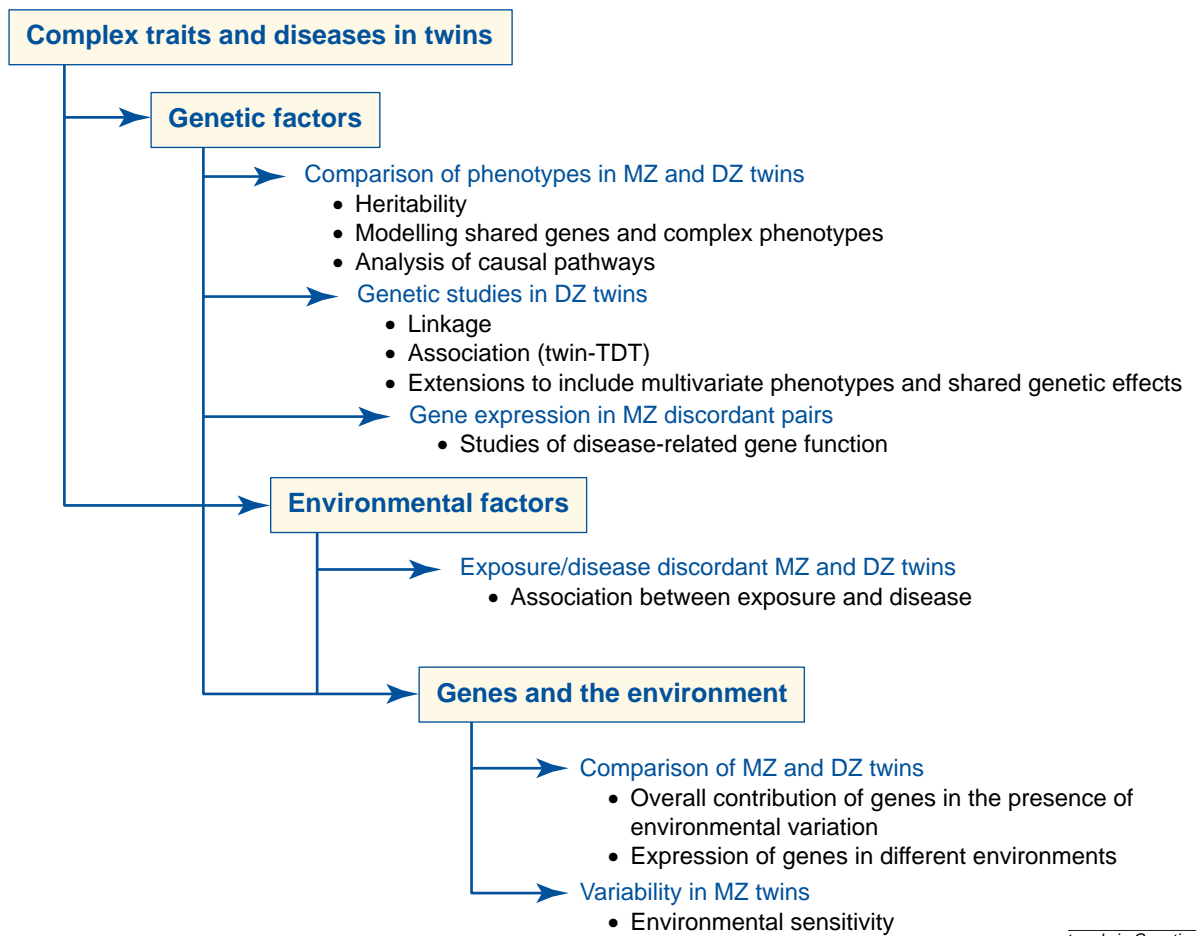
Nicholas J. Schork*
njs2@po.cwru.edu

Tim D. Spector
tim.spector@kcl.ac.uk

Twin Research and Genetic Epidemiology Unit, St Thomas' Hospital, London, UK SE1 7EH.

*Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44109-1998, USA.

FIGURE 1. The contribution of twins to the study of complex traits and diseases



trends in Genetics

In complex diseases, twins provide an estimate of the genetic contribution to disease through the estimation of heritability, enhance strategies to detect genes through linkage and association, provide a matched setting in which to assess the risk that is associated with environmental exposure, and provide insight into the way in which genes and the environment act together to cause disease. Abbreviations: DZ, dizygotic; MZ, monozygotic; TDT, transmission disequilibrium test.

diseases. Allele-sharing methods in affected sibling pairs provide an alternative approach, but are limited in their power and require large sample sizes. Therefore, much emphasis has been placed on ways to increase power of linkage analysis, and several strategies have been suggested¹¹. These include (among others): (1) studying larger sibship sizes; (2) selecting individuals with extreme values from the tails of the distribution (for phenotypes such as blood pressure that are measured on a continuous scale); (3) increasing the marker density; and (4) the use of more informative markers. Power is also influenced by the choice of analytical strategy, for example: (1) by the use of

multi-point as opposed to single-point analysis; (2) by including the full data distribution in the study of continuous traits¹²; (3) by using multivariate methods that consider genetic effects on several phenotypes simultaneously (pleiotropy); and (4) by incorporating repeated measures.

Studying twins also offers the potential to increase power in linkage studies. Their contribution is not, at first sight, obvious. MZ twins are genetically identical and alone are uninformative for linkage. DZ twins are genetically indistinguishable from ordinary siblings. This, however, ignores several important properties of the twin design that might be particularly relevant to complex traits and diseases. The precise matching for age and date of birth is a characteristic of twins that is unique among family studies. For most complex diseases, age is crucially important in disease expression. The shared family environment of twins brings with it a greater degree of matching for a range of environmental variables, both measured and unmeasured, all of which might contribute to the expression of diseases. Therefore, it is much easier to attribute phenotypic differences between twins to genetic rather than to environmental factors. An additional advantage of studying twins is that the probability of non-paternity, an important cause of misclassification in sibling studies, is reduced to almost nil.

TABLE 1. Heritability of common complex traits and diseases from twin studies

	Heritability (%)	Ref.
Asthma	60	20
Blood pressure	40–70	30
Bone mineral density	60–80	31
Cervical and lumbar disc degeneration	60–80	32
Insulin-dependent diabetes mellitus	70	33
Obesity	50–90	34
Osteoarthritis (hand, knee)	50–70	35
Rheumatoid arthritis	60	36
Ulcerative colitis	50	37

As an example, in considering the risk of osteoporosis, we need to take into account the well-characterized decline in bone density with age¹³. Environmental risk factors such as exercise, dietary calcium intake, smoking and alcohol intake all contribute to the risk. The calendar year of birth might also influence disease – hence, temporal changes in, for example, school milk, childhood exercise, sunlight exposure, dietary supplements and availability of drugs, such as the contraceptive pill or hormone replacement, could lead to important environmental differences. All of these factors might detract from the detection of linkages in an unmatched sample; all are accounted for in the design when studying DZ twins.

The matching of DZ twins for the shared environment and for background genetic factors is relevant to studies of disease association with putative candidate genes. Recently, considerable attention has been given to the possible bias that results from population admixture and stratification, in which heterogeneity in the genetic background of cases and controls can lead to detection of spurious genetic effects. The problem can be circumvented using within-family controls in methods that are based on the transmission disequilibrium test (TDT)¹⁴. In its classical form, the TDT involves studying disease associations in parent–offspring units. The method has, however, been extended recently to studying sibling pairs alone in the absence of parental data, both for diseases and for traits that are measured on categorical and continuous scales^{15–17}. Sibling TDT methods are equally applicable to data from DZ twins. However, in studying complex diseases, for the reasons discussed earlier with respect to studying linkage, the matching for age and environmental variance makes the twin-TDT a potentially more powerful tool to detect the influence of genes of small effect against substantial background variation.

Extensions of the twin model

For complex traits and diseases, approaches that consider only the effects of single genes on single measures of disease provide limited insight into the genetic determinants of disease. Multivariate analysis, a statistical method in which several phenotypic variables can be taken into account simultaneously in analysis together with the effects of multiple genes, provides a more appropriate approach. The simplicity of the twin study design has made its extension to multivariate analysis relatively straightforward^{18,19}.

These statistical approaches are applicable to all types of family design. However, the specific value of twin data to multivariate modelling is in its ability to separate shared genetic factors from shared environmental influences. These two components of variation are statistically inseparable in other study designs. Hence, twins allow the existence of shared genes that determine combinations of traits to be identified (pleiotropic effects). Duffy *et al.* provide an example of this approach in a study of the occurrence of asthma, hay fever, dust allergy and eczema in twins²⁰. The analysis shows evidence of a common genetic mechanism underlying all of these traits and suggests that an ‘atopic tendency’ in a range of disease phenotypes has a common genetic basis. By contrast, Nguyen *et al.* have studied the common determinants of bone density and lean body mass in twins and have found that the association between the two is mediated by the shared environment and not by shared genetic factors²¹.

Further extensions of these methods allow twin data to provide insight into the causal pathways that are involved in disease processes^{22,23}. For example, twins allow an assessment of the interaction between groups of related variables, such as the physiological biofeedback relationship between levels of insulin and glucose²⁴. Thus, interactions between variables can mimic the effects of shared genes: twin data allow these possible causal explanations to be compared directly. Twin designs provide a diverse range of models in which biologically plausible models for the action of individual genes can be assessed. Incorporating biological knowledge of disease processes and intermediate phenotypes into mathematical models has the potential to direct the search for genes towards the components of complex disease that have a clear-cut genetic basis. Providing models for a disease that are a better reflection of known disease mechanisms is also likely to increase the usefulness of the approach in detecting genes.

Multivariate approaches not only provide a more plausible approach to modelling complex traits, they also, as discussed earlier, increase statistical power for strategies to detect genes¹. This has been demonstrated in several applications, both to simulated and real datasets. For example, the computation of genetic factor scores based on multivariate data could, in certain circumstances, dramatically increase the power to detect quantitative trait loci (QTLs)²⁵. Multivariate methods have been extended recently to maximize evidence for linkage within datasets. These approaches are particularly powerful when several variables included in the analysis are influenced to some degree by the same locus¹¹.

Gene-expression studies

Advances in molecular biology have highlighted the way in which twins can potentially be used to understand how genes function in health and disease. Here, a focus of interest would be in tissues and other physiological samples that are obtained from MZ twins. The close genetic matching in MZ twin pairs enables an accurate assessment of the pattern of differential gene usage through comparisons of mRNA and protein expression in pairs that are discordant for a disease or trait. A major problem in applying this approach to unrelated individuals, for example in case–control scenarios, has been the wide variation in functioning genes owing to unrelated individuals having greater differences in exposure and genetic background when compared with relatives. By substantially reducing the intra-pair genetic variation through comparing the members of an MZ twin pair, the approach can increase specificity for detecting relevant functioning genes.

A simple mathematical argument can show the usefulness of twins for these kinds of comparison studies. Most statistical tests for the difference between two or more groups (e.g. diseased versus non-diseased, mutant versus non-mutant, etc.) depend on the magnitude of the difference and the variance of that difference. Essentially, one would like a large difference in means across the groups, with a small variance of that difference, as evidenced in the formulation of the standard t-statistic for testing the equality of two groups:

$$t = \frac{\bar{x} - \bar{y}}{\text{var}(x - y)} \quad (1)$$

where \bar{x} and \bar{y} are the means of the two groups and var

$(x - y)$ is the variance of that difference. The variance of the difference between two variables is:

$$\text{var}(x + y) = \text{var}(x) + \text{var}(y) - 2\text{cov}(x,y) \quad (2)$$

where $\text{var}(x)$ and $\text{var}(y)$ are the variances of the variables x and y and $\text{cov}(x,y)$ is the covariance between x and y . Thus, the larger the covariance between x and y , the smaller the variance of the difference $\text{var}(x - y)$, given in Eqn 2. To increase this covariance, and hence decrease the variance of the difference, one would want to choose sampling units that have a positive covariance or correlation. Unrelated individuals or distant relatives are not likely to show much covariation. Twins, especially MZ twins, exhibit strong positive covariation for all of the reasons discussed, and hence they are ideal to use to test differences²⁶.

The increasing availability of gene-array technologies, which can test rapidly for differences in thousands of genes between small numbers of pairs, makes this an attractive and feasible study design. Possible examples include twins that are discordant for diseases in which the tissue is accessible (such as skin or haematological disorders) and can be grown to produce mRNA. With the pace of developments in sequencing, chip technology and protein biochemistry, these methods are likely to be used extensively in the future to examine gene function.

Gene-environment interactions

Twins also have the potential to provide information on the way in which genes and the environment interact. This can be approached through multivariate modelling if data are collected in pairs that are discordant for environmental exposure¹⁹. Intervention studies in twins also provide a means for assessing differential responses to environmental stimuli²⁷.

The phenomenon of 'environmental sensitivity' is also of great interest. This refers to a situation in which a gene's effect is mediated not through an influence on the level of

a measured phenotype, but rather through its variability in response to the environment. This can be approached directly through studying MZ twins. Phenotypic differences within MZ twin pairs are the result of environmental effects alone, as MZ twins uniquely share their entire genetic background. If greater within-pair variance for a disease or trait can be demonstrated in MZ twins of a particular genotype, it can be inferred that this is the direct result of the gene's environmental sensitivity.

The existence of genes that determine variability rather than level of a trait is well recognized in plant and animal literature. Several examples have been described in humans and include the association between blood group M and variability in cholesterol levels²⁸. Martin has pointed out that variability genes might be more important than genes that determine basal levels²⁹. An allele that is associated with an environmentally induced increase in variance of 50%, for example, increases the proportion of cases above the third standard deviation by more than fivefold, and hence might have an important influence on the prevalence of extreme cases of disease. Therefore, such genes might be of great potential importance in human health. For example, they could determine an individual's responsiveness to diet and might influence the likely success of a therapeutic intervention.

Conclusions

Developments in molecular biology and biometric analysis are changing the traditional place of twin research in genetic epidemiological inquiry. The assessment of heritable phenotypes can now be extended to understanding the more complex web of causal and interactive processes, thus aiding the identification of genes. Twins not only provide a useful and powerful tool for identifying genes, by acting as ideally matched sib-pairs, but are also uniquely placed to measure the extent of their action, their expression and the nature of their interaction with the environment.

References

- Martin, N. *et al.* (1997) A twin-pronged attack on complex diseases. *Nat. Genet.* 17, 387-392
- Falconer, D.S. (1989) *Introduction to Quantitative Genetics*, Longman Scientific and Technical
- Hopper, J.L. (1998) Heritability. In *Encyclopaedia of Biostatistics* (Armitage, P. and Colton, T., eds), pp. 1905-1906, Wiley
- Todd, J.A. and Farrell, M. (1996) Panning for gold: genome-wide scanning for linkage in type 1 diabetes. *Hum. Mol. Genet.* 5, 1443-1448
- Abramson, M.J. and Harrap, S.B. (1998) The new asthma genetics and its implications for public health. *Public Health Rev.* 26, 127-144
- Moldin, S.O. (1997) The maddening hunt for madness genes. *Nat. Genet.* 17, 127-129
- Morrison, N.A. *et al.* (1994) Prediction of bone density from vitamin D receptor alleles. *Nature* 367, 284-287
- Gong, G. *et al.* (1999) The association of bone mineral density with vitamin D receptor gene polymorphisms. *Osteoporosis Int.* 9, 55-64
- Cooper, G.S. and Umbach, D.M. (1996) Are vitamin D receptor polymorphisms associated with bone mineral density? A meta-analysis. *J. Bone Miner. Res.* 11, 1841-1849
- Allison, D.B. and Schork, N.J. (1997) Selected methodological issues in meiotic mapping of obesity genes in humans: issues of power and efficiency. *Behav. Genet.* 27, 401-421
- Allison, D.B. *et al.* (1998) Multiple phenotype modeling in gene-mapping studies of quantitative traits: power advantages. *Am. J. Hum. Genet.* 63, 1190-1201
- Fulker, D.W. and Cherny, S.S. (1996) An improved multipoint sib-pair analysis of quantitative traits. *Behav. Genet.* 26, 527-532
- Cooper, C. (1993) Epidemiology and public health impact of osteoporosis. *Baillière's Clin. Rheumatol.* 7, 459-477
- Spielman, R.S. *et al.* (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* 52, 506-516
- Spielman, R.S. and Ewens, W.J. (1998) A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. *Am. J. Hum. Genet.* 62, 450-458
- Allison, D.B. (1997) Transmission disequilibrium test for quantitative traits. *Am. J. Hum. Genet.* 60, 676-690
- Fulker, D.W. *et al.* (1999) Combined linkage and association sib-pair analysis for quantitative traits. *Am. J. Hum. Genet.* 64, 259-267
- Neale, M.C. (1997) *Mx: Statistical Modeling* (4th edn), Department of Psychiatry, Virginia Commonwealth University
- Neale, M.C. and Cardon, L.R. (1992) *Methodology for Genetic Studies in Twins and Families*, Kluwer Academic Publishers
- Duffy, D.L. *et al.* (1990) Genetics of asthma and hay fever in Australian twins. *Am. Rev. Respir. Dis.* 142, 1351-1358
- Nguyen, T.V. *et al.* (1998) Bone mass, lean mass, and fat mass: same genes or same environments? *Am. J. Epidemiol.* 147, 3-16
- Heath, A.C. *et al.* (1993) Testing hypotheses about direction of causation using cross-sectional family data. *Behav. Genet.* 23, 29-50
- Duffy, D.L. and Martin, N.G. (1994) Inferring the direction of causation in cross-sectional twin data: theoretical and empirical considerations. *Genet. Epidemiol.* 11, 483-502
- Snieder, H. *et al.* (1999) Bivariate genetic analysis of fasting insulin and glucose levels. *Genet. Epidemiol.* 16, 426-446
- Boomsma, D.I. (1996) Using multivariate genetic modeling to detect pleiotropic quantitative trait loci. *Behav. Genet.* 26, 161-166
- Schork, N.J. and Schork, M.A. (1993) Some sampling effects of pairwise correlated observations on likelihood ratio tests for the difference between two means. *Commun. Stat.: Theory Methods* 22, 2609-2617
- Bouchard, C. and Tremblay, A. (1997) Genetic influences on the response of body fat and fat distribution to positive and negative energy balances in human identical twins. *J. Nutr.* 127, 943-947
- Birley, A. *et al.* (1997) MN blood group affects response of serum LDL cholesterol to a low fat diet. *Clin. Genet.* 51, 291-295
- Martin, N. (1999) Gene-environment interaction in twin studies. In *Advances in Twin and Sib-pair Analysis* (Spector, T.D. *et al.*, eds), pp. 143-150, Greenwich Medical Media
- Snieder, H. *et al.* (1995) Developmental genetic trends in blood pressure levels and blood pressure reactivity to stress. In *Behavior Genetic Approaches in Behavioral Medicine* (Turner, J.R. *et al.*, eds), pp. 105-130, Plenum Press
- Arden, N.K. *et al.* (1996) The heritability of bone mineral density, ultrasound of the calcaneus and hip axis length: a study of postmenopausal twins. *J. Bone Miner. Res.* 11, 530-534
- Sambrook, P.N. *et al.* (1999) Genetic influences on cervical and lumbar disc degeneration: a magnetic resonance imaging study in twins. *Arthritis Rheum.* 42, 366-372
- Kyvik, K.O. *et al.* (1995) Concordance rates of insulin dependent diabetes mellitus: a population based study of young Danish twins. *Br. Med. J.* 311, 913-917
- Maes, H.H. *et al.* (1997) Genetic and environmental factors in relative body weight and human adiposity. *Behav. Genet.* 27, 325-351
- Spector, T.D. *et al.* (1996) Genetic influences on osteoarthritis in women: a twin study. *Br. Med. J.* 312, 940-943
- MacGregor, A.J. *et al.* (2000) Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum.* 43, 30-37
- Tysk, C. *et al.* (1988) Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut* 29, 990-996