CLASSICAL TWIN STUDIES AND BEYOND

Dorret Boomsma, Andreas Busjahn‡ and Leena Peltonen§*

Twin studies have been a valuable source of information about the genetic basis of complex traits. To maximize the potential of twin studies, large, worldwide registers of data on twins and their relatives have been established. Here, we provide an overview of the current resources for twin research. These can be used to obtain insights into the genetic epidemiology of complex traits and diseases, to study the interaction of genotype with sex, age and lifestyle factors, and to study the causes of co-morbidity between traits and diseases. Because of their design, these registers offer unique opportunities for selected sampling for quantitative trait loci linkage and association studies.

ASCERTAINMENT BIAS
A systematic distortion in measuring the true frequency of a phenomenon, such as a trait or a disease, owing to the way in which the data are obtained.

**Department of Biological Psychology, Vrije Universiteit, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands.*
‡HealthTwiSt and Franz-Volhard-Clinic of the Charité, Wiltbergstrasse 50, 13125 Berlin, Germany.
§Departments of Medical Genetics and Molecular Medicine, University of Helsinki and National Public Health Institute, Biomedicum, Haartmaninkatu 8, 00290 Helsinki, Finland.
Correspondence to D.B.
e-mail: di.boomsma@psy.vu.nl
doi:10.1038/nrg932

Recent advances in genetics, such as the completion of the human genome sequence, an increased understanding of sequence variants, the availability of low-cost genome-wide tools to monitor these variants and the development of powerful multivariate statistical tools, have opened new avenues of investigation in human genetics. There is a danger, however, that the usefulness of these tools is limited when applied to human complex traits. ASCERTAINMENT BIAS, problems with phenotypic assessment, lack of follow-up of the phenotypes over time and environmental noise that can arise, for example, from developmental variation, might all contribute to making the genes that underlie complex, multifactorial traits and diseases in humans difficult to identify. During the past decade, considerable effort has been devoted to whole-genome screens to detect QUANTITATIVE TRAIT LOCI (QTL) for complex traits and diseases. The success of linkage studies to map genes of unknown function, and LINKAGE-DISEQUILIBRIUM studies that are aimed at establishing the association between a particular genetic polymorphism and a disease, are considered by some researchers to have been limited (for example, see REFS 1,2). But others, such as Korstanje and Paigen[3], find this pessimistic view premature. Instead, they offer a more optimistic view by summarizing the acceleration in gene identification from mammalian QTL in the past ten years — for example, Crohn disease in humans, blood pressure in rats and diabetes in mice. Although animal research might provide essential information, it cannot encompass the complex interplay of genes with the human environment. For complex human traits, there is an increasing recognition of the need to understand population genetics and the biometrical properties of human traits, so that phenotypes can be defined in a way that maximizes the chances of successful gene mapping[4]. In this review, we discuss how twin studies can help in this endeavour.

We aim to draw the attention of the reader to the enormous potential for research on the genetics of complex traits that lies in the established twin registers. Some of them have existed for decades and have carefully collected LONGITUDINAL DATA on behavioural traits, diseases and environmental risk factors in large samples of twins and their families. By facilitating comparisons between monozygotic (MZ) and dizygotic (DZ) twins, twin registers represent some of the best resources for evaluating the importance of genetic variation in susceptibility to disease. They are also an excellent resource for studying the significance of the GENOTYPE × ENVIRONMENT INTERACTION (for example, lifestyle; see the section on 'Improvements in analysis' for further discussion) and of the contribution of specific polymorphisms to the total genetic variance. Nevertheless, these facts, and the existence of twin registers, remain to be appreciated fully by the wider community of genetic researchers.

© 2002 Nature Publishing Group

QUANTITATIVE TRAIT LOCUS (QTL). Genetic locus or chromosomal region that contributes to the variability in complex quantitative traits, as identified by statistical analysis. Quantitative traits are typically affected by several genes and the environment.

LINKAGE DISEQUILIBRIUM (LD). The condition in which the frequency of a particular haplotype for two loci is significantly greater than that expected from the product of the observed allelic frequencies at each locus.

LONGITUDINAL STUDY
A study in which repeated measurements are taken from the same subjects at different time points.

GENOTYPE × ENVIRONMENT INTERACTION
(G × E). The influence of specific combinations of genetic and environmental factors on a trait that goes beyond the additive action of these factors. It refers to genes that control sensitivity to the environment, or the environment that controls gene expression.

MULTIVARIATE ANALYSIS
The simultaneous inclusion of two or more (dependent) variables in one analysis, for example, in estimating the genetic correlation of birth weight with blood pressure.

COVARIATE
In a multivariate analysis, a variable with known effects that is used to test the effect of the main variables that are independent of those known effects. The inclusion of age in studies of age-dependent traits is a simple example.

HERITABILITY
The proportion of the total phenotypic variation in a given characteristic that can be attributed to additive genetic effects. In the broad sense, heritability involves all additive and non-additive genetic variance, whereas in the narrow sense, it involves only additive genetic variance.

CONCORDANCE
The occurrence of the same trait in both members of a pair of twins. Concordance might occur for diseases as well as for behaviours, such as smoking.

Figure 1 | **Velvet twins.** Oil painting on linen (145 cm × 135 cm) by Mary Waters (1997/1998). From a private collection, courtesy of Flatland Off the Record (publishers), Utrecht, The Netherlands.

To disentangle and to quantify the contributions that genes, the shared environment, the individual-specific environment and their interactions make to human complex traits, we need data from relatives who are genetically related, but who grow up in unrelated environments (the so-called 'twin adoption design'), or relatives who grow up in similar environments but are of differing genetic relatedness (the so-called 'twin design'). If the exposure to environmental risk factors can be assessed, these designs also make it possible to quantify the effect that the genotype × environment interaction has on shaping a particular trait. Recent advances in statistical modelling allow simultaneous analysis of many variables in relatives such as MZ and DZ twins. These advances also make it possible to carry out new types of analysis — such as the MULTIVARIATE ANALYSIS of causes of co-morbidity between disorders, the analysis of the development of childhood psychopathology over time, the inclusion of COVARIATES in linkage analyses, and the estimation of HERITABILITY and linkage that are conditional on the exposure to environmental risk factors. All these improvements in data analysis and the possibilities of new types of analysis have, in turn, led to the establishment of large registers of twins that no longer focus on the assessment of a single phenotype, but collect a wide range of traits and environmental factors in twins, as well as in their family members.

Here, we focus on the use of data from twins and their families in epidemiological and molecular-genetic studies. We draw attention to a large resource for such studies and provide an overview of the most important features of twin registers worldwide. But first, we briefly introduce the classical twin designs and contrast them with some of their recent extensions.

Types of twin study design

Twins have always captured our curiosity, researchers and artists alike (FIG. 1), and there have been proposals to use them as a 'natural experiment' in empirical studies as early as 415 AD (REF. 5). Galton's classic paper on twins, published in the nineteenth century[6] is often cited as the first on the classical twin method, although it is uncertain if Galton knew of the distinction between MZ and DZ twins[7,8]. The systematic analysis of similarity between MZ and DZ twins was introduced by Siemens, a dermatologist, who formulated the twin rule of pathology: any heritable disease will be more CONCORDANT in identical twins than in non-identical twins, and concordance will be even lower in non-siblings[9]. When studying skin moles, Siemens came up with the clever idea of combining correlation analysis with twin data: he correlated mole counts in one twin with mole counts in the other twin, and compared this correlation in MZ and DZ pairs of twins. The correlation for mole count in MZ twins, who share all, or nearly all, of their genetic material was 0.4. In DZ twins, who are genetically 50% identical on average, the correlation was only 0.2. The results indicated the importance of genetic factors in variation in mole count — the greater genetic resemblance in MZ twins is associated with their greater resemblance for the phenotype under study.

Classical twin studies. As exemplified above, the classical twin study compares phenotypic resemblances of MZ and DZ twins. MZ twins derive from a single fertilized egg and therefore inherit identical genetic material. Comparing the resemblance of MZ twins for a trait or disease with the resemblance of DZ twins offers the first estimate of the extent to which genetic variation determines phenotypic variation of that trait. If MZ twins resemble each other more than do DZ twins, then the heritability (h^2) of the phenotype can be estimated from twice the difference between MZ and DZ correlations. For example, typical MZ and DZ correlations for depression are about 0.4 and 0.2 (REF. 10), and therefore heritability is estimated at ~40%. A different pattern of correlations is usually observed for lifestyle factors, indicating the importance of the shared family environment. For taking up smoking during adolescence, typical MZ and DZ correlations are 0.9 and 0.7, respectively (REF. 11), leading to a heritability estimate of 40%, but also indicating the importance of a shared environment. The proportion of the variance that is due to a shared environment is the difference between the total twin correlation and the part that is explained by heritability. That is, $r_{MZ} - h^2$ in MZ or $r_{DZ} - h^2/2$ in DZ twins, where r_{MZ} is the correlation between MZ twins and r_{DZ} is the correlation between DZ twins. For the trait of taking up smoking, this estimate is ~50% (0.9–0.4 in MZ or 0.7–0.2 in DZ twins). The application of this type of analysis has led to substantial changes in the way we think about the determinants of health and disease, and the causes of individual differences in normal and abnormal behaviour. During the past decade, a shift has taken place from strict environmental explanations to a more balanced view that recognizes the importance of genes[12] — for example, in autism and in attention-deficit hyperactivity disorder (ADHD) in children[13,14], or in the development of dependence on alcohol and other drugs[15,16] (FIG. 2).

© 2002 Nature Publishing Group

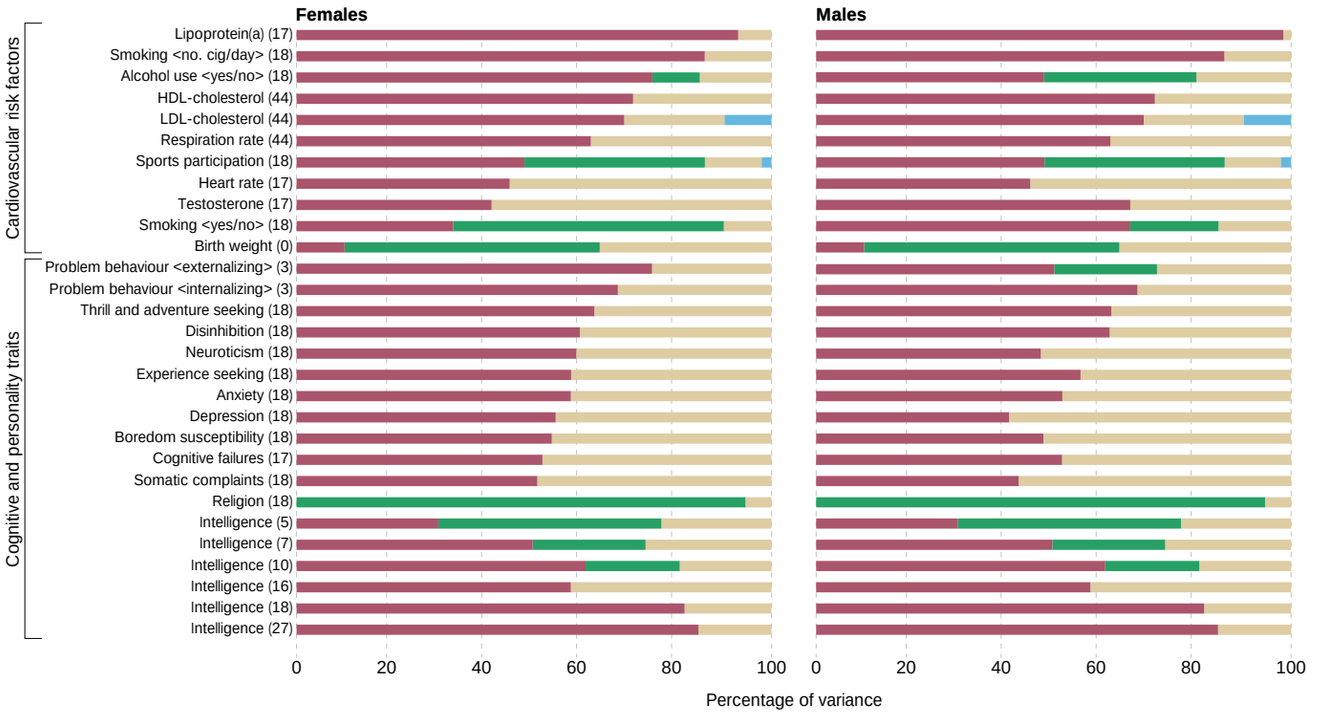

Figure 2 | **Examples of results from classical twin analysis.** Percentage of variances explained by genetic factors (purple), by shared environmental factors (green), by unique environmental influences (beige) and by differences in age (blue). The phenotypes were assessed in Dutch female and male twins (and in some cases also in their parents and siblings). The number in the brackets gives the modal age of the sample in years. Phenotypes include cardiovascular risk factors, and personality and cognitive traits. The heritability for a behavioural as well as cardiovascular risk factor, such as the number of cigarettes smoked per day, is nearly as high as for lipoprotein(a) levels. For personality traits and indices of psychopathology, heritability is ~50%, with a higher estimate for internalizing and externalizing problems in young children. Heritability of intelligence is age dependent and increases rapidly between 5 and 18 years of age. There are few differences in heritability between females and males. HDL, high-density lipoprotein; LDL, low-density lipoprotein. Modified with permission from REF. 23 © (1999) Australian Academic Press.

ANALYSIS OF VARIANCE (ANOVA). A statistical method to test the null hypothesis that the mean values of two or more groups are equal. The variance around the means in groups is compared with the variance of the group means. In genetic applications, the variance between families is compared with the variance within families. A significant *F*-ratio implies that variance between families is larger than within families.

INTRA-CLASS CORRELATION A statistical measure for the strength and direction of resemblance between two variables (or two family members). It can vary between –1 and +1. Intra-class correlation refers to the correlation in defined subgroups — for example, in monozygotic or dizygotic pairs — and can be derived from ANOVA as $t = (F - 1)/(F + 1)$.

Improvements in analysis. The quantitative traits that have been assessed in MZ and DZ twins have traditionally been analysed using ANALYSIS OF VARIANCE and INTRA-CLASS CORRELATIONS to summarize twin resemblance (for example, see REF. 17). This approach, however, cannot accommodate the effect of gender on variances and covariances of twins of opposite sexes. STRUCTURAL EQUATION MODELLING (SEM), also known as covariance modelling, is a more general alternative approach, in which genotypic and environmental effects are modelled as the contribution of unmeasured (latent) variables to the potentially multivariate phenotypic differences between individuals[18]. The latent factors represent the effects of many unidentified influences. For a genetic factor, these effects are due to a potentially large, but unknown, number of POLYGENES. The contributions of the latent variables are estimated as REGRESSION COEFFICIENTS in the LINEAR REGRESSION of the observed variables on the latent variables. Several widely available software programs, such as LISREL[19] or Mx[20] allow the estimation of parameters by using NORMAL THEORY MAXIMUM LIKELIHOOD and WEIGHTED LEAST SQUARES. A useful estimator in the Mx program is the raw data likelihood estimator, which handles data from selected samples and from studies in which part of the sample might have missing data. The latter often arises in longitudinal studies. SEM can accommodate the analysis of gender differences in heritability estimates through the simultaneous analysis of data from male and female MZ and DZ twins. It is possible to test whether the same genes are expressed in males and females by including DZ twins of opposite sexes. If the resemblance between twins of opposite sexes is less than expected on the basis of the heritability in males and females, then this indicates that different genes might influence the same trait in the two sexes. Similarly, heritability that is conditional on environmental exposure can indicate the presence of a genotype × environment interaction[21]. This interaction can be detected by including environmental measurements on the basis of which the twin sample can be stratified. For example, Heath *et al.*[22] found that heritability for depression in married women was lower than in unmarried women.

Evidence for the effect of the genotype × environment interaction on personality comes from a study in Dutch adolescent twins, in which it was shown that a religious upbringing greatly reduces the influence of genetic factors on DISINHIBITION, a trait that closely resembles 'novelty seeking' and that is associated with substance use and abuse[23].

© 2002 Nature Publishing Group

STRUCTURAL EQUATION MODELLING
(SEM). Also known as covariance modelling. A method that estimates regression coefficients ('parameters') between latent (unobserved) and observed variables. These estimates minimize the difference between the covariance structure of the observed data and that predicted by the model. Alternative models (such as family resemblance being due to shared genes versus shared environment) can be compared by how well they fit the data and by the number of parameters estimated.

POLYGENES
A group of genes that influence a complex trait. In contrast to monogenic traits, most traits and diseases are influenced by several genes, only a sum of which is sufficient to cause the effect.

Beyond the classical designs. TABLE 1 summarizes the possibilities for research that go beyond the classical twin design. Extending the MZ–DZ design to include the testing of parents, siblings, spouses and offspring of both MZ and DZ twins, offers the opportunity to assess the presence of cultural transmission, genotype × environment covariance, non-random mating, and social interactions within and between generations[24]. Simpler versions of the fuller, extended twin studies — such as a study of young-adult twins, their middle-aged parents and a second group of middle-aged twins (that is, twins of the same age as the parents of the first group of pairs of twins) — makes it possible to assess the effect of age differences on heritability and to assess differential gene expression as a function of age. Using this design, which can be considered a shortcut for a true longitudinal study, Snieder *et al.*[25] obtained evidence that partially different genes influence lipid levels in plasma at different ages. This might be important information for gene-finding studies, as there might only be a limited time period during which genes that vary over the course of an individual's life can be detected. Other extended twin studies look at the offspring of MZ twins who are genetically half-sibs but socially cousins[26]. The MZ-offspring design also allows for the testing of maternal effects and imprinting by comparing the offspring of male and female MZ pairs.

Table 1 | Types of twin study and their applications

Types of twin study	Application
Classical MZ–DZ comparison	These studies estimate the contributions of genetic and environmental effects to phenotypic variance, and test, for example, for age, cohort and sex differences in gene expression
Multivariate analyses: simultaneous analysis of correlated traits	This type of analysis involves: • direction of phenotypic causality • causes of co-morbidity of two or more traits: multivariate modelling of environmental and genetic correlations between traits • multivariate modelling to obtain genotypic (or environmental) values for individuals • analysis of longitudinal data to study causes of phenotypic stability and tracking over time • testing of G × E using measured environmental indices
Co-twin control study	Case–control studies of MZ twins who are perfectly matched for genes and family background; such studies can also be used to study gene expression in discordant twins
Extended twin study: studies of twins and their families	• Parents of twins can be included to study cultural transmission and G × E covariance • Parents of twins can be studied in a quasi-longitudinal design to determine genetic and environmental stability • Assortative mating can be studied if spouses of twins are included; social interactions and special twin effects, such as prenatal hormone transition, the 'private language' of twins and shared prenatal environment, can be studied if siblings of twins are included • Maternal effects, G × E correlation and imprinting can be studied if offspring of MZ twins are included
Genotyping at candidate loci	These studies include: • genotyping of MZ twins to detect variability genes and to estimate penetrance • genotyping of DZ twins to estimate associations within and between families
Genotyping at marker loci	These studies include: • genotyping of DZ twins (and parents) to detect linkage with QTL • selecting informative families from large twin registers to find QTL of small effect

DZ, dizygotic; G × E, genotype × environment interaction; MZ, monozygotic; QTL, quantitative trait loci.

Multivariate designs. By generalizing the univariate twin study to multivariate designs, in which more than one phenotype per person is analysed, the causes of association and co-morbidity between traits can be investigated. Multivariate twin studies make it possible to ask questions such as: Does variation in exercise behaviour cause variation in depression, or do the traits cluster because they are influenced by a common set of genes? Or, to give another example, do symptoms of anxiety and those of depression cluster because one disorder increases the risk for the other, or is there a common genetic vulnerability?[27,28] The answers to these questions lie in the cross-twin, cross-trait correlations in MZ versus DZ pairs of twins, such as the correlation of anxiety in one twin with depression in the other twin. Using this design, Neale and Kendler[28] and Roy *et al.*[29] found that co-morbidity between major depression (MD) and generalized anxiety disorder (GAD) is not due to chance or to a third independent disorder, but is probably caused by correlated susceptibilities to the disorders. The model that best explained the association of MD and GAD included a strong genetic correlation (of one) between MD and GAD, and a weak correlation with the individual-specific environment. GAD and MD therefore share genetic factors, but their environmental determinants are mostly distinct.

Co-morbidity, the occurrence of two or more disorders together, is often seen as a problem in molecular-genetic psychiatric studies. To overcome this problem, valid and objective diagnostic categories need to be established. It is important to remember, however, that a diagnostic label can be valid, without necessarily being biologically meaningful[30]. ADHD is a good example of the importance of diagnostic categories. In clinical settings, ~50% of children who meet the criteria for ADHD also meet the criteria for oppositional defiant disorder (ODD) and conduct disorder (CD), and another 20–25% of children with ADHD also meet the criteria for an anxiety disorder[31]. Strategies for linkage studies of ADHD could focus on 'pure' cases (for example, children who have ADHD without any co-morbid problems of ODD or anxiety) or on distinguishing in the patient sample between ADHD cases with different co-morbid conditions. Instead of agreeing on a particular scheme before carrying out a linkage study, it might be more powerful to analyse the causes of co-morbidity in a multivariate twin design and to establish the extent to which the phenotypes that cluster share a common genetic basis. Twin registers can be extremely helpful in this respect by coming up with reliable phenotypes that can be useful in molecular-genetic studies. Again, twin studies of ADHD are an example. Todd *et al.*[32] found that different ADHD subtypes are influenced by different genetic factors. Moreover, the genetically distinct severe subtype also accounted for most of the co-morbidity in the sample[33]. It therefore seems that

© 2002 Nature Publishing Group

separate genetic vulnerabilities account for ADHD alone and for ADHD with co-morbid conditions. Pending replication, these results provide a clear indication that linkage studies without phenotypic subtyping would be struggling to deal with different genotypes that underlie a seemingly similar phenotype.

Multivariate analyses are also needed for the simultaneous modelling of phenotypes (such as depression) and ENDOPHENOTYPES or INTERMEDIATE PHENOTYPES (such as neuroticism or cortisol levels) to determine their common genetic aetiology. The power to detect linkage will only be increased through the use of endophenotypes if their association is due to pleiotropic genetic effects[34].

Case–control studies. Twins are also useful in case–control studies. MZ twins offer the possibility of carrying out the ideal case–control study, as they are perfectly matched for genotype and family background. Martin *et al.*[35] studied vitamin C administration in one twin and a placebo in the other twin, and found that, contrary to popular belief, there was no effect of vitamin C on the common cold. Studies of the effects of fetal and infant growth on later health (the Barker hypothesis; for example, see REF. 36) in twins have looked at the differences in birth weight in MZ and DZ twins and at their association with differences in cardiovascular and metabolic parameters. Twin studies can resolve whether these associations are causal, or due to shared genetic factors. For blood pressure, it turned out that the association between low birth weight and high blood pressure in later life is mediated by common genes[37].

In a more general version of this design, MZ twins, only one of whom has a disease, are used in gene-expression studies to distinguish between genes that are related to the causes of the disease and genes that are expressed as a consequence of the disease. Alternatively, such differential expression, congruent with disease discordance, might indicate causal genes that are differentially activated by epigenetic factors[38]. Conversely, the detection of somatic mutation of the same gene in tumours from MZ twins, both of whom have the tumour, might be a powerful way to detect predisposition genes[39,40].

The use of twins in molecular-genetic studies

The concordance between MZ twins sets the upper limit on predictions of individual risk that can be made on the basis of the human genome sequence. Discordant MZ schizophrenic twins, for example, show that disease outcome can be different for two individuals with an identical genetic make-up. The same is also illustrated by MZ twins with cleft lip and palate (FIG. 3). MZ-twin concordance therefore gives important information about disease PENETRANCE and, if MZ twins are genotyped at candidate loci, they provide information about locus-specific penetrances. For quantitative traits, an association between a particular genotype and MZ-trait differences might not only reveal the relationship between the gene and the trait, but also indicate that the gene might be a 'variability', as opposed to a 'level', gene. The idea of variability and level genes was introduced by Berg *et al.*[41] who were inspired by the study of intra-pair variance for cholesterol in MZ twins of M– or M+ blood group[42]. Level genes are considered to affect the mean expression of a trait, or prevalence in the case of a disease, and are the usual target of association studies. By contrast, variability genes need not influence trait levels; instead, they determine the extent of the influence that the environment has on the intra-individual variability.

In rare cases, MZ twins might inherit nearly, but not completely, identical genotypes. Skewed X-inactivation in female MZ twins, asymmetric transmission of mitochondria, somatic mutations and rearrangements of epigenetic signals during gametogenesis might all account for differences between MZ pairs of twins[34,38]. Such MZ twins might be extremely informative in

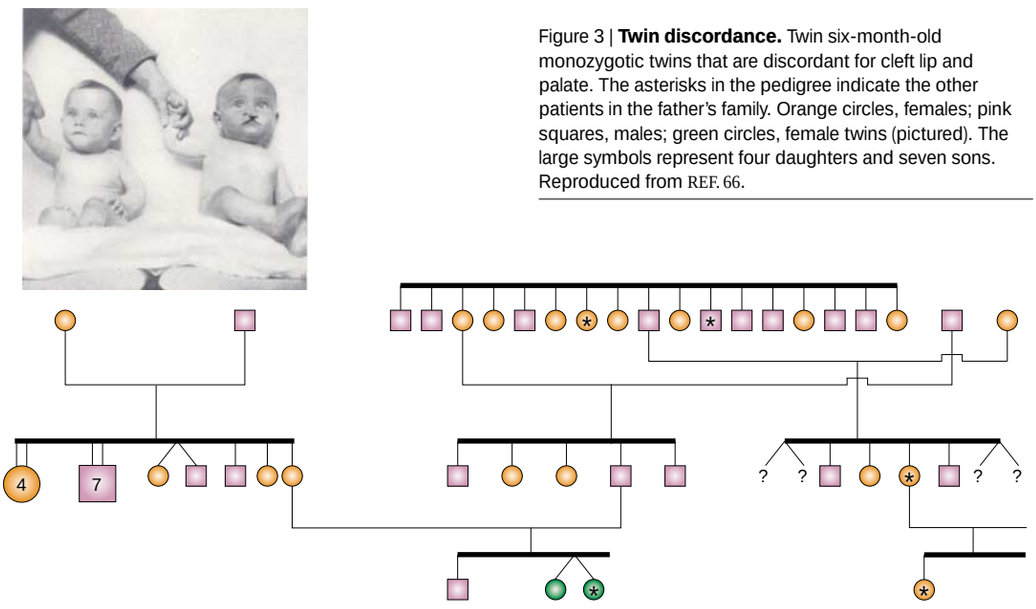

Figure 3 | **Twin discordance.** Twin six-month-old monozygotic twins that are discordant for cleft lip and palate. The asterisks in the pedigree indicate the other patients in the father's family. Orange circles, females; pink squares, males; green circles, female twins (pictured). The large symbols represent four daughters and seven sons. Reproduced from REF. 66.

REGRESSION COEFFICIENTS/ LINEAR REGRESSION
Linear regression is a statistical method to test and to describe the linear relationship between two or more variables. The regression coefficient describes the angle of the regression line and reflects the amount of variance of the dependent variable that is explained by variation of the independent variable.

NORMAL THEORY MAXIMUM LIKELIHOOD
(ML). A statistical method that works by varying the estimates for parameters of a model, so that the likelihood of the observed data points is maximized. Under normal theory, the likelihood corresponds to the height of the normal curve (one variable) and to the height of the multivariate normal probability density function for two or more variables.

WEIGHTED LEAST SQUARES
(WLS). An alternative method for estimating parameters during model fitting. The square of the difference between the observed statistic (for example, mean or covariance) and the statistic that is predicted by the theoretical model is weighted and minimized. Weights are usually chosen to correspond to the accuracy of the observed statistics.

DISINHIBITION
A sub-scale of the psychological trait 'sensation seeking', including items that describe experiences or attitudes that relate to sensation seeking through other exciting people, disinhibited or 'wild' parties, social drinking and sexual variety.

ENDOPHENOTYPES/ INTERMEDIATE PHENOTYPES
The physiological traits that are related to a disease trait; for example, for hypertension this could include blood pressure, angiotensin levels or salt sensitivity.

PENETRANCE
The proportion of affected individuals among the carriers of a particular genotype. If all individuals with a disease genotype show the disease phenotype, then the disease is said to be completely penetrant.

© 2002 Nature Publishing Group

molecular-genetic studies, as was recently shown. A pair of MZ twins that were discordant for Van der Woude syndrome (VWS) helped to identify the interferon regulatory factor 6 gene (*IRF6*) as the locus that is responsible for the syndrome. Subsequently, the same mutation was identified in 45 unrelated affected families[43].

Twins might also offer specific advantages in genome-wide genotyping, such as linkage or association studies, to map QTL. DZ twins are siblings of the same age, so for traits that change with age, this fact will decrease variance and therefore increase the power for linkage studies. Twins are also matched for a broad range of pre- and postnatal factors and are more likely than other siblings to have the same father. The value of combining linkage analyses for Mendelian traits in large pedigrees with twin-based QTL linkage was shown in mapping a cholesterol-lowering gene, known as *CLG* (or *MMP1*)[44]. The mapping of this gene to chromosome 13 was based initially on a single Arab pedigree. The subsequent linkage study in German DZ twins not only confirmed the locus, but also added information on the relevance of the as-yet-unknown gene by showing its influence on lipid levels in the general population.

Of course, MZ-twin data do not contribute towards detecting linkage, as MZ twins share all their genetic material, being identical by descent. However, analysing MZ phenotypic data simultaneously with the linkage data in DZ twins and sibling pairs makes it possible to distinguish between the effects that background genes and a shared family environment have on the amount of familial variance that is not accounted for by the QTL.

Risch and Zhang[45] have proposed that stringent selection of extreme discordant and concordant (EDAC) pairs might be the only reliable strategy for QTL mapping in humans. Large registers of twins (see next section) contain the phenotypes of thousands, sometimes tens of thousands, of twins and their family members, and are a suitable source of families that are informative for linkage studies. Several such QTL-mapping projects are now under way. For example, the genes that influence neuroticism and depression are sought in selected samples of twins and their siblings[46–48]. DZ twins are used to ensure that extremely discordant pairs are not selected because of age differences between them. For ordinary sibs, age differences might be responsible for large phenotypic differences.

Twin registers worldwide

Getting off the ground. Many of the current twin registers (TABLES 2,3) are spin-offs from specific research projects, usually in psychology or medicine. For example, the East Flanders Prospective Twin Study, now among the most comprehensive collections of perinatal twin data, started after the observation of lower degrees of intra-uterine hypoxia in second-born twins. The Berlin Twin Register started as a study on the genetics of blood-pressure regulation during mental stress. Once scientists gain experience with twin research, they usually realize the great potential of such studies that extend beyond the original question and the necessity to 'nurse' this source of data. Resources allowing, a twin collection develops into a register. To define a register, we need to take into account the necessity to maintain and extend the original sample and data collection, the sample size and range, and the sophistication of the data. For this overview, we have chosen large, population-based registers with up to 150,000 pairs of twins, as well as small, but growing, collections of less than a hundred pairs.

The existing or emerging population-based registers, such as those in Scandinavia, Italy, Korea and Sri Lanka, are of extreme value for epidemiological research owing to the lack of ascertainment bias. In some countries, the sampling of pairs of twins is based on computerized population registers, either by using direct information on multiple births or by applying complex filters including sharing of date of birth, family name at birth, place of birth or partly sharing identification numbers. Norway, which has one of the largest twin collections, established a computerized twin register in 1990, but a traditional nationwide archive of birth register was established by the 'Act of Castbergian' as early as 1916. In the absence of government address records, substitutes such as drivers licence records are being used to trace twins. But unbiased population samples can be obtained using local or regional rather than nationwide approaches, as has been shown in Sri Lanka where hospital records of the main maternity hospitals were used as a source of information on twins. In all cases, 'real' twins have to be distinguished from a larger subset of 'potential' twins, as sharing the name of their mother and their date of birth might occur by chance[49]. Other twin collections are gathered independently of centralized records and therefore depend more on the motivation of the twins or their parents. Recruiting twins through advertising of twin studies has been used in some cases, as well as through mass-media articles on twins and twin research, in which information on major achievements were combined with continuing studies and contact information. These approaches have been surprisingly effective, and the possible effects of bias in such non-randomly ascertained samples can be dealt with efficiently by statistical methods[50,51].

Many researchers have started to realize the value of large data collections for molecular-genetic studies and have begun to include siblings, parents and offspring of twins in their research projects.

Running the registers. Recruiting twins is only the first step in running a twin register. There are many processes that are involved in breathing life into such a data collection. Keeping addresses up-to-date, maintaining contact with and motivating twins are continuous tasks. In this respect, twin registers face the same problems as other large epidemiological collections, except that the chances of re-establishing contact are almost doubled owing to the family-based sample and the close relationship of most twins. For all the main registers, rules for starting new data collections have been developed. All the registers included in this overview are open to collaboration, provided that scientific review confirms the merits of a research

© 2002 Nature Publishing Group

proposal, that there are no ethical concerns and that funding is available. Basic funding is typically provided either nationally by the National Institutes of Health in the United States (or equivalent funding agencies elsewhere), or internationally, such as by the European Community. Specific studies on twins are funded by the same grants as are other medical and psychology studies.

Data collection. Administrators of most registers maintain contact with the twins and their family members through Web sites (TABLES 2,3), by sending out newsletters and through mailed surveys. Even DNA samples have been collected by post — for example, DNA from buccal swabs has been collected in this way[52]. The data collection by mailed questionnaires has led to some very large data sets. For example, Lake *et al.*[53] analysed data

Table 2 | Twin registers in Europe

Twin studies	URL or contact e-mail address	Number of twins	Primary interest	Origin of twins	References
East Flanders Prospective Twin Survey (EFPTS)	c.derom@pi.de	6,050	Epidemiology, placentation*, congenital anomalies, perinatal factors and the Barker hypothesis	Belgium	56,67
The Danish Twin Registry	askytthe@health.sdu.dk	65,000	Ageing and age-related health, metabolic and cardiovascular disease, and specific diseases	Denmark	56,68
The Finnish Twin Cohort	jaakko.kaprio@helsinki.fi	15,000	Health, personality and substance abuse	Finland	56,69
Berlin Twin Register (HealthTwiSt)	http://www.healthtwist.de	>900	Complex diseases, health-related QTs and pharmacogenetics	Germany	56,70
German Observational Study of Adult Twins (GOSAT) and the Bielefeld Longitudinal Study of Adult Twins (BiLSAT)	angleitner@uni-bielefeld.de	2,509	Longitudinal assessment of temperament and personality; generalizability of behavioural genetic findings across methods of personality assessment	Germany	56,71
Italian Twin Registry	http://www.gemelli.iss.it	120,000	Ageing, dementia, cardiovascular diseases, MS, celiac disease, diabetes, asthma, allergies, thyroid diseases and behavioural disorders	Italy	56
Register of Italian Twin Athletes (RITA)	casini@iusm.it	4,719	Human biology and development, sport and high-level performance	Italy	56
Twin Register of Rome (TERRY)	casini@iusm.it	13,228	Lifestyle, development and ageing	Italy	56
Norway Twin Registries	mina.bergem@psykiatri.uio.no	>40,000	Mental health, obesity, asthma and allergies, health behaviours and perceptions, perinatal influence on health	Norway	56,72
The NIPH Twin Panel	jennifer.harris@folkehelsa.no	7,668	Physical and mental health, asthma, allergies, obesity and health-related behaviours	Norway	56,73
The Swedish Twin Registry	http://www.mep.ki.se/twin	57,405	Cancer, cardiovascular diseases, dementia, depression, substance use/abuse, cognition, personality, ageing and common complex diseases	Sweden	56,74
The Swedish Young Male Twins Study	finn.rasmussen@imm.ki.se	1,783	Risk factors for metabolic and cardiovascular diseases; obesity and behavioural risk factors	Sweden	56
Netherlands Twin Register (NTR)	http://www.psy.vu.nl/ntr	30,335	Development, behaviour and emotional problems; cognition, depression, addiction and cardiovascular risk factors	The Netherlands	23,56
St Thomas' UK Adult Twin Registry	http://www.twin-research.ac.uk	10,000	Cardiovascular, metabolic, musculoskeletal, dermatological and ophthalmological diseases	UK	56,75
Study of growth before birth and adult health	g.mcneill@abdn.ac.uk	123	Risk factors for coronary heart disease	UK	56
Twins' Early Development Study (TEDS)	a.trouton@iop.kcl.ac.uk	16,810	Longitudinal assessment of verbal and non-verbal cognitive development and delay; language development and delay; childhood behaviour problems	UK	56
Northern Region Multiple Pregnancy Register	christopher.wright@ncl.ac.uk	1,216	Effects of multiple pregnancy, obstetric and paediatric management, and outcomes of pregnancy	UK (North East)	56

*Chorion type. MS, multiple sclerosis; NIPH, National Institute of Public Health; QT, quantitative trait.

© 2002 Nature Publishing Group

from more than 45,000 twins and their relatives on neuroticism, which is a strong risk factor for the development of depression, and concluded that familial resemblance for this trait has a simple genetic basis. Because of the large data sets that are available, it was possible to reject alternative models for familial resemblance, such as cultural transmission.

Although sample size is clearly important in genetic epidemiology, there is sometimes a pay-off in the amount of data that can be collected through questionnaires and large-scale survey studies, and the scope or depth of phenotypes that need to be collected in laboratory settings. To study the genetic basis of disease, intermediate phenotypes often need to be determined in more extensive and costly studies. The selection of subsamples can be based on random selection and also on previous phenotyping from the larger registers (as, for example, in the Australian (TABLE 3) or Scandinavian (TABLE 2) registers); alternatively, smaller twin cohorts can be collected for specific studies. The Register of Italian Twin Athletes, or the Study of Growth Before Birth and Adult Health (TABLE 2) are good examples of successful in-depth studies on a small scale.

Centralized health databases that exist in some countries greatly facilitate efficient data collection. Finland and other Scandinavian countries have centralized registers for hospital discharge data and for fully reimbursable medications. This information is accessible with a unique personal identifier that is given to each individual at birth. This allows record linking for all twins without any self-selection as potential bias. Personal identifiers of twins are used as a query filter for the health databases, and matching records can then be transferred into the twin database. This strategy was successfully used by Lichtenstein *et al.*[54] to obtain heritability estimates for 28 anatomically distinct types of cancer. The authors combined data on 44,788 pairs of Swedish, Danish and Finnish twins and found that genetic factors only make a minor contribution to susceptibility to most types of neoplasm. (However, Risch[55] argues that, for some cancers, the data are compatible with a much larger influence of genetic factors.)

Unique features of individual registers. The behavioural measurements collected for twin registers range from questionnaire-based personality traits that are catalogued in many registers, to standardized psychiatric interviews obtained by telephone and behavioural responses recorded on video during real-life situations (such as in the Bielefeld twin studies, TABLE 2). In studies of young and adolescent twins, substantial resources are devoted to developmental studies of psychopathology and substance use and abuse. Clinical measurements can be collected and include diagnoses, as well as specific measures such as bone density (as in the St Thomas twin register in London) and magnetic resonance imaging scans for cardiac function and morphology (as in the Berlin register) (TABLE 2). Many registers focus on cardiovascular disease and on metabolic syndromes and their intermediate phenotypes (for example, the Netherlands Twin Register (TABLE 2) and the 'Project Grow-2-Gether' in New York (TABLE 3)). The determination of various phenotypes across various regions of interest in the same twin cohort greatly adds to the value of the data collection by allowing multivariate study designs. Traits such as eating habits, addiction or coping with stress are related to various health measures, based either on underlying common genetic or environmental influences or on direct effects. Twin methodology is suitable to disentangle this complex relationship, given that the relevant measures are available.

The value of a twin register is not so much determined by the existing database as by the ability to go back to the twins to add information about the phenotypes in a hypothesis-driven manner. Any finding leads to new questions, and genetic research is no exception to this rule. Finding a disease-relevant gene leads to the question of the physiological pathway in which it acts and the subsequent need for related physiological measures, known as (endo)phenotyping. Most twin registers operate in a longitudinal way, establishing a continuing relationship with study participants, which allows the researcher to go back to interesting families to collect additional data, including DNA samples for molecular-genetic studies. For several registers, longitudinal data collection — for example, to study the development of childhood psychopathology — is the main focus of the endeavour, and twins are recruited at birth or shortly afterwards. The East Flanders Prospective Twin Survey registers twins at birth and is unique in that it determines zygosity and placental information on all twins. Data collected throughout the lifetime of the twins offer unique opportunities, provided we can develop new statistical strategies that take advantage of such data.

The range of phenotypes that are available in the existing registers covers many kinds of behavioural and clinical trait. A more-detailed description of the registers, the samples of twins and twin family members, and the phenotypic data collection, are available in a special issue of *Twin Research*[56] published in October 2002, which is devoted to twin registers worldwide and from which TABLES 2,3 are abstracted.

Conclusion and prospects

For multifactorial traits (such as body height and weight, neuroticism and blood lipid levels) and complex diseases (such as obesity, depression and cardiovascular disease), twin studies have shown that genetics contributes significantly to the variation that is seen at the population level. Many of these traits and diseases are now on the increase in large areas of the world and are influenced by risk factors that include diet, smoking and lack of exercise (see, for example, REF. 57). Although these 'lifestyle' risk factors that are important for the development of complex diseases are often considered to be 'environmental', they might themselves be influenced by genes. Twin studies have been useful in assessing the extent to which variation in lifestyle and healthy behaviour might itself be heritable. In fact, recent twin studies provided considerable evidence that 'lifestyle' risk factors aggregate in families owing to shared genes, in

© 2002 Nature Publishing Group

addition to the shared environment. For example, twin studies have indicated that differences in eating patterns[58], alcohol use[16], smoking initiation and persistence[59,60], sports participation[61] and even religious beliefs[62,63] might all be influenced by genetic variation. Heritability for a particular disease might therefore reflect the direct influence of disease genes, the influence of genes that are responsible for variation in lifestyle factors or the influence of genes that modify the influence of lifestyle on disease risk.

The value of large, unbiased study samples that are needed to verify the role of the genetic variation that underlies common traits is well recognized. As pointed out by Thompson[64], we have now moved into an era in which genotyping is relatively cheap and fast, and the main cost of a study of a complex trait involves family data collection and trait phenotyping. This view is reflected in a recent decision by the European Community to fund a large integrated project called GENOMEUTWIN. The six participating twin cohorts,

Table 3 | Twin registers outside Europe

Twin studies	URL or contact e-mail address	Number of twins	Primary interest	Origin of twins	References
Australian Twin ADHD Project (ATAP)	http://psych.curtin.edu.au/people/hayd.htm	1,959	ADHD and childhood behavioural disorders	Australia	56
Australian Twin Registry	http://www.twins.org.au	27,582	General resource for medical and scientific research	Australia	56,76
Western Australian Twin Register	http://www.ichr.uwa.edu.au	4,729	Asthma and allergy; ADHD; early speech and behaviour	Australia	56
University of British Columbia Twin Project	http://www.psychiatry.ubc.ca	816	Personality and the personality disorder	Canada	56,77
Chinese National Twin Programme (CNTP)	cchp@public3.bta.net.cn	4,576	Establishment of population-based national twin registry; aetiologies of common diseases and health-related behaviour	China	56
Osaka University Aged Twin Registry	hayakawa@sahs.med.osaka-u.ac.jp	12,000	Ageing, dementia, physical diseases, lipids, cognition, lifestyle, life satisfaction and quality of life	Japan	56
Korean Twin Registry	sungjohn@kangwon.ac.kr	154,783	Complex human diseases and traits	S. Korea	56
Seoul Twin Family Study	http://www.ktrc.org	>4,615	Cognitive abilities	S. Korea	56
National Twin Registry of Sri Lanka	http://www.infolanka.com/org/twin-registry	20,294	Establishment of nationwide population-based twin register for multidisciplinary research and international collaborations	Sri Lanka	56
Mid-Atlantic Twin Registry	http://www.matr.vcu.edu	23,000	Behavioural and psychiatric	USA	56
NAS–NRC Twin Registry	http://www.iom.edu/twins	15,924	Somatic and psychiatric disease; ageing; social, psychological and demographic variables	USA	56
Southern Illinois Twins	http://www.siumed.edu/playlab	126	Peer interaction behaviours and pre-school cognitive development	USA	56
Vietnam Era Twin (VET) Registry	birute.curran@med.va.gov	7,500	Veterans health, effects of combat, psychiatric disorders and substance abuse	USA	56
International Twin Study	tmack@usc.edu	17,229	Aetiology of disease and genetic markers	USA; Canada	56
California Twin Program	http://twins.usc.edu	13,096	Aetiology of disease and genetic markers	CA, USA	56
San Diego Twin Blood Pressure Study at UCSD	http://elcapitan.ucsd.edu/hyper	200	Blood pressure; 'autonomic intermediate phenotypes' for high blood pressure	CA, USA	56
Southern California Twin Register	http://www.rcf.usc.edu/~lbaker	2,600	Social and moral development; childhood behaviour problems; cognitive abilities	CA, USA	56
Georgia Cardiovascular Twin Study	http://www.mcg.edu/institutes/gpi	534	Longitudinal development of bio-behavioural antecedents of cardiovascular disease in youth	GA, USA	56
Minnesota Twin Family Study (MTFS)	http://www.tc.umn.edu/~mctfr	4,723	Substance use, related child and adult disorders	MN, USA	56,78
Minnesota Twin Registry	http://www.psych.umn.edu/psylabs/mtfs	5,599	Individual differences	MN, USA	56
NY Obesity Research Center Child Twin Registry (Project 'Grow-2-Gether')	http://cpmcnet.columbia.edu/dept/obesectr/NYORC/twins.html	50	Food intake; body composition	NY, USA	56
Wisconsin Twin Project	http://psych.wisc.edu/wtp	NA*	Childhood behavioural disorders	WI, USA	56

*Began in 1994 by recruiting twins born in 1989 onwards. ADHD, attention-deficit hyperactivity disorder; NA, not applicable; NAS–NRC, National Academy of Sciences–National Research Council; UCSD, University of California at San Diego.

© 2002 Nature Publishing Group

from Scandinavia, The Netherlands and Italy, form an amazing collection of more than 0.6 million pairs of twins. More than 30,000 DNA samples, accompanied by informed consent for genetic studies of common diseases, have been collected from these population-based twin cohorts. Combining the data from the main twin registers of Europe will integrate the efforts of the leading genetic and epidemiological researchers in the field of twin research. In this research project, epidemiological and phenotypic data collection will be integrated, and special emphasis will be placed on quality control of the data to be entered in the databases and on the level and reliability of the collected clinical phenotypes, as well as lifestyle factors and specific life events. Initial 'proof-of-principle' genome-wide genotyping efforts will be targeted at 10,000 twins and will look at stature, body mass index, coronary disease and migraine.

Twins and their family members are often enthusiastic participants in research studies. The increase in the twinning rate that is now seen in The Netherlands and other countries[65] ensures the viability of the application of the classical twin design in genetic epidemiology and in medical, behavioural and psychiatric genetics.

1. Altmuller, J., Palmer, L. J., Fischer, G., Scherb, H. & Wjst, M. Genomewide scans of complex human diseases: true linkage is hard to find. *Am. J. Hum. Genet.* **69**, 936–950 (2001).
2. Ionnidis, J. P. A., Ntzani, E. E., Trikalinos, T. A. & Contopoulos-Ioannidis, D. G. Replication validity of genetic association studies. *Nature Genet*. **29**, 306–309 (2001).
3. Korstanje, R. & Paigen, B. From QTL to gene: the harvest begins. *Nature Genet*. **31**, 235–236 (2002).
4. Sham, P. Shifting paradigms in gene-mapping methodology for complex traits. *Pharmacogenomics* **2**, 195–202 (2001).
5. St Augustine of Hippo. *De Civitate Dei* (*The City of God*), New edn (Penguin, 2001).
6. Galton, F. The history of twins as a criterion of the relative powers of nature and nurture. *J. R. Anthropol. Inst. Gt Br. Ireland* **5**, 391–406 (1875).
7. Bulmer, M. G. *The Biology of Twinning in Man* (Clarendon, Oxford, 1970).
8. Bouchard, T. J. & Propping, P. *Twins as a Tool of Behavioral Genetics* (John Wiley & Sons, Chichester, UK, 1993).
9. Siemens, H. W. *Die Zwillingspathologie: Ihre Bedeutung, ihre Methodik, ihre bisherigen Ergebnisse* (*Twin Pathology: Its Importance, Its Methodology, Its Previous Results*) (Springer, Berlin, 1924).
10. Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C. & Eaves, L. J. A population-based twin study of major depression in women. The impact of varying definitions of illness. *Arch. Gen. Psychiatry* **49**, 257–266 (1992).
11. Boomsma, D. I., Koopmans, J. R., van Doornen, L. J. P. & Orlebeke, J. F. Genetic and social influences on starting to smoke: a study of Dutch adolescent twins and their parents. *Addiction* **89**, 219–226 (1994).
12. Plomin, R., DeFries, J. C., Craig, I. W. & McGuffin, P. *Behavioral Genetics in the Postgenomic Era* (APA Books, Washington, DC, 2002).
13. Folstein, S. & Rutter, M. Genetic influences and infantile autism. *Nature* **265**, 726–728 (1977).
14. Faraone, S. V. & Doyle, A. E. The nature and heritability of attention-deficit/hyperactivity disorder. *Child Adolesc. Psychiatry Clin. N. Am*. **10**, 299–316, viii–ix (2001).
15. Tsuang, M. T., Bar, J. L., Harley, R. M. & Lyons, M. J. The Harvard Twin Study of Substance Abuse: what we have learned. *Harv. Rev. Psychiatry* **9**, 267–279 (2001).
16. Heath, A. C. *et al.* in *Behavioral Genetics in the Postgenomic Era* (eds Plomin, R., DeFries, J. C., Craig, I. W. & McGuffin, P.) 309–334 (APA Books, Washington, DC, 2002).
17. Kempthorne, O. & Osborne, R. H. The interpretation of twin data. *Am. J. Hum. Genet*. **13**, 320–339 (1961).
18. Neale, M. C. & Cardon, L. R. *Methodology for Genetic Studies of Twins and Families*, NATO ASI Series D: Behavioural and Social Sciences, Vol. 67 (Kluwer Academic, Dordrecht, The Netherlands, 1992).
This book describes the theoretical foundations of twin analysis, as well as its practical application.
19. Jöreskog, K. G. & Sörbom, D. LISREL 8.5, Scientific Software International, Lincolnwood, Illinois, USA <http://ssicentral.com/lisrel/mainlis.htm> (2001).
20. Neale, M. C., Boker, S. M., Xie, G. & Maes, H. H. *Mx: Statistical Modeling* 6th edn, Department of Psychiatry, VCU Box 900126, Richmond, Virginia 23298, USA. Program, documentation and sample scripts available at <http://www.vcu.edu/mx> (2002).
A useful source of information on the computer program Mx that is widely used in the twin research community.
21. Eaves, L. J., Last, K. A., Martin, N. G. & Jinks, J. L. A progressive approach to non-additivity and genotype–environmental covariance in the analysis of human differences. *Br. J. Math. Statist. Psychol*. **30**, 1–42 (1997).
22. Heath, A. C., Eaves, L. J. & Martin, N. G. Interaction of marital status and genetic risk for symptoms of depression. *Twin Res*. **1**, 119–122 (1998).
23. Boomsma, D. I., de Geus, E. J., van Baal, G. C. & Koopmans, J. R. A religious upbringing reduces the influence of genetic factors on disinhibition: evidence for interaction between genotype and environment on personality. *Twin Res*. **2**, 115–125 (1999).
24. Truett, K. R. *et al*. A model system for analysis of family resemblance in extended kinships of twins. *Behav. Genet*. **24**, 35–49 (1994).
25. Snieder, H., van Doornen, L. J. P. & Boomsma, D. I. Age-dependency of gene expression for plasma lipids, lipoproteins and apolipoproteins. *Am. J. Hum. Genet*. **60**, 638–650 (1997).
26. Nance, W. E., Kramer, A. A., Corey, L. A., Winter, P. M. & Eaves, L. J. A causal analysis of birth weight in the offspring of monozygotic twins. *Am. J. Hum. Genet*. **35**, 1211–1223 (1983).
27. Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C. & Eaves, L. J. Major depression and generalized anxiety disorder. Same genes, (partly) different environments? *Arch. Gen. Psychiatry* **49**, 716–722 (1992).
28. Neale, M. C. & Kendler, K. S. Models of comorbidity for multifactorial disorders. *Am. J. Hum. Genet*. **57**, 935–953 (1995).
29. Roy, M. A., Neale, M. C., Pedersen, N. L., Mathe, A. A. & Kendler, K. S. A twin study of generalized anxiety disorder and major depression. *Psychol. Med*. **25**, 1037–1049 (1995).
30. Strachan, T. & Read, A. P. *Human Molecular Genetics* Ch. 12 (BIOS, Oxford, UK, 1999).
31. Faraone, S. V. & Biederman, J. Do attention deficit hyperactivity disorder and major depression share familial risk factors? *J. Nerv. Ment. Dis*. **185**, 533–541 (1997).
32. Todd, R. D. *et al.* Familiality and heritability of subtypes of attention deficit hyperactivity disorder in a population sample of adolescent female twins. *Am. J. Psychiatry* **158**, 1891–1898 (2001).
33. Neuman, R. J. *et al.* Latent class analysis of ADHD and comorbid symptoms in a population sample of adolescent female twins. *J. Child Psychol. Psychiatry* **42**, 933–942 (2001).
34. Martin, N., Boomsma, D. & Machin, G. A twin-pronged attack on complex traits. *Nature Genet*. **17**, 387–392 (1997).
35. Martin, N. G., Carr, A. B., Oakeshott, J. G. & Clark, P. Co-twin control studies: vitamin C and the common cold. *Prog. Clin. Biol. Res*. A **103**, 365–373 (1982).
36. Hales, C. N. & Barker, D. J. The thrifty phenotype hypothesis. *Br. Med. Bull*. **60**, 5–20 (2001).
37. Ijzerman, R. G., Stehouwer, C. D. & Boomsma, D. I. Evidence for genetic factors explaining the birth weight–blood pressure relation. Analysis in twins. *Hypertension* **36**, 1008–1012 (2000).
This paper shows the value of twin studies in distinguishing between genetic and environmental sources for correlation between traits.
38. Petronis, A. Human morbid genetics revisited: relevance of epigenetics. *Trends Genet*. **17**, 142–146 (2001).
39. Forsti, A., Jin, Q., Sundqvist, L., Soderberg, M. & Hemminki, K. Use of monozygotic twins in search for breast cancer susceptibility loci. *Twin Res*. **4**, 251–259 (2001).
40. Mack, T. M., Hamilton, A. S., Press, M. F., Diep, A. & Rappaport, E. B. Heritable breast cancer in twins. *Br. J. Cancer* **87**, 294–300 (2002).
41. Berg, K., Kondo, I., Drayna, D. & lawn, R. "Variability gene" effect of cholesteryl ester transfer protein (CETP) genes. *Clin. Genet*. **35**, 437–445 (1989).
42. Magnus, P., Berg, K., Borreson, A.-L. & Nance, W. E. Apparent influence of marker genotypes on variation in serum cholesterol in monozygotic twins. *Clin. Genet*. **19**, 1, 67–70 (1981).
43. Kondo, I. *et al.* Mutations in *IRF6* cause Van der Woude and popliteal pterygium syndromes. *Nature Genet.* **32**, 285–289 (2002).
44. Knoblauch, H. *et al.* A cholesterol-lowering gene maps to chromosome 13q. *Am. J. Hum. Genet*. **66**, 157–166 (2000).
This article combines classical family-based linkage analysis with mapping of a QTL in twins.
45. Risch, N. & Zhang, H. Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* **268**, 1584–1589 (1995).
46. Boomsma, D. I. *et al.* Netherlands Twin Family Study of Anxious Depression (NETSAD). *Twin Res*. **3**, 323–334 (2000).
47. Kirk, K. Anxiety and depression in twin and sib pairs extremely discordant and concordant for neuroticism: prodromus to a linkage study. *Twin Res*. **3**, 299–309 (2000).
48. Martin, N. *et al.* A population-based study of personality in 34000 sib-pairs. *Twin Res*. **3**, 310–315 (2000).
49. Goldberg, J. *et al.* Identification of a cohort of male and female twins aged 65 years or more in the United States. *Am. J. Epidemiol.* **145**, 175–183 (1997).
50. Neale, M. C. & Eaves, L. J. Estimating and controlling for the effects of volunteer bias with pairs of relatives. *Behav. Genet*. **23**, 271–277 (1993).
51. Bechger, T. M., Boomsma, D. I. & Koning, H. A limited dependent variable model for heritability estimation with non-random ascertained samples. *Behav. Genet*. **32**, 145–151 (2002).
52. Meulenbelt, I., Droog, S., Trommelen, G. J., Boomsma, D. I. & Slagboom, P. E. High-yield noninvasive human genomic DNA isolation method for genetic studies in geographically dispersed families and populations. *Am. J. Hum. Genet*. **57**, 1252–1254 (1995).
53. Lake, R. I., Eaves, L. J., Maes, H. H., Heath, A. C. & Martin, N. G. Further evidence against the environmental transmission of individual differences in neuroticism from a collaborative study of 45,850 twins and relatives on two continents. *Behav. Genet*. **30**, 223–233 (2000).
54. Lichtenstein, P. *et al.* Environmental and heritable factors in the causation of cancer — analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med*. **343**, 78–85 (2000).
55. Risch, N. The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiol. Biomarkers Prev*. **10**, 733–741 (2001).
56. *Twin Research* **5** (October) [epub ahead of print] <http://www.australianacademicpress.com.au/Publications/TR/TR.html> (2002).
***Twin Research* is the official journal of the International Society for Twin Studies. This issue gives details on most of the existing twin registers.**
57. Willett, W. C. Balancing life-style and genomics research for disease prevention. *Science* **296**, 695–698 (2002).
58. Van den Bree, M. B., Eaves, L. J. & Dwyer, J. T. Genetic and environmental influences on eating patterns of twins aged ≥50 y. *Am. J. Clin. Nutr*. **70**, 456–465 (1999).

© 2002 Nature Publishing Group

59. Koopmans, J. R., Slutske, W. S., Heath, A. C., Neale, M. C. & Boomsma, D. I. The genetics of smoking initiation and quantity smoked in Dutch adolescent and young adult twins. *Behav. Genet*. **29**, 383–393 (1999).
60. Madden, P. A. *et al.* The genetics of smoking persistence in men and women: a multicultural study. *Behav. Genet*. **29**, 423–431 (1999).
61. Beunen, G. & Thomis, M. Genetic determinants of sports participation and daily physical activity. *Int. J. Obes. Relat. Metab. Disord*. **23** (Suppl.3), S55–S63 (1999).
62. Eaves, L. *et al.* Comparing the biological and cultural inheritance of personality and social attitudes in the Virginia 30,000 study of twins and their relatives. *Twin Res*. **2**, 62–80 (1999).
63. Bouchard, T. J. Jr, McGue, M., Lykken, D. & Tellegen, A. Intrinsic and extrinsic religiousness: genetic and environmental influences and personality correlates. *Twin Res*. **2**, 88–98 (1999).
64. Thompson, E. A. in *Handbook of Statistical Genetics* (eds Balding, D. J., Bishop, M. & Cannins, C.) 541–563 (John Wiley & Sons, Chichester, UK, 2001).
65. Derom, R., Orlebeke, J., Eriksson, A. & Thiery, M. in *Multiple Pregnancy. Epidemiology, Gestation & Perinatal Outcome* (eds Keith, L. G., Papiernik, E., Keith, D. M. & Luke, B.) 145–162 (The Parthenon Publishing Group, New York, 1995).
This book gives a comprehensive overview on all aspects of multiple births.
66. Voute, P. A. *De Differentieele Diagnostiek van Tweelingen*. Ph.D. thesis, Kemink en Zoon NV, Utrecht (1935).
67. Derom, C., Vlietinck, R., Derom, R. & Van den Berghe, H. Increased monozygotic twinning rate after ovulation induction. *Lancet* **1**, 1236–1238 (1987).
68. Kjeldsen, M. J., Kyvik, K. O., Christensen, K. & Friis, M. L. Genetic and environmental factors in epilepsy: a population-based study of 11.900 Danish twin pairs. *Epilepsy Res.* **44**, 167–178 (2001).
69. Cannon, T. D. *et al.* The inheritance of neuropsychological dysfunction in twins discordant for schizophrenia. *Am. J. Hum. Genet.* **67**, 369–382 (2000).
70. Busjahn, A. *et al.* QT interval is linked to 2 long-QT syndrome loci in normal subjects. *Circulation* **99**, 3161–3164 (1999).
71. Borkenau, P., Riemann, R., Angleitner, A. & Spinath, F. M. Genetic and environmental influences on observed personality: evidence from the German Observational Study of Adult Twins. *J. Pers. Soc. Psychol.* **80**, 655–668 (2001).
72. Bergem, A. L. M. & Lannfelt, L. Apolipoprotein E type e4 allele, heritability and age at onset in twins with Alzheimer disease and vascular dementia. *Clin. Genet.* **52**, 408–413 (1997).
73. Harris, J. R., Tambs, K. & Magnus, P. Sex-specific effects for body mass index in the new Norwegian twin panel. *Genet. Epidemiol.* **12**, 251–265 (1995).
74. Lichtenstein, P. *et al.* The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. *J. Intern. Med.* **252**, 184–205 (2002).
75. Spector, T. D. Influence of vitamin D receptor genotype on bone density in postmenopausal women: a British twin study. *Br. Med. J.* **310**, 1357–1360 (1995).
76. Hopper, J. L. *et al.* Genetic, common environment and individual specific components of variance for age- and lean mass-adjusted bone mineral density in 10- to 26-year-old females: a twin study. *Am. J. Epidemiol.* **147**, 17–29 (1998).
77. Jang, K. L. *et al.* The covariance structure of neuroticism and agreeableness: a twin and molecular genetic analysis of the role of the serotonin transporter gene. *J. Pers. Soc. Psychol.* **81**, 295–304 (2001).
78. McGue, M., Iacono, W. G., Legrand, L., Malone, S. & Elkins, I. Origins and consequences of age at first drink. I. Associations with substance-use disorders, disinhibitory behavior and psychopathology, and P3 amplitude. *Alcoholism: Clin. Exp. Res.* **25**, 1156–1165 (2001).

Acknowledgements

We thank J. F. Orlebeke, A. L. Beem, J. M. Vink, J. J. Hudziak and N. G. Martin for their contributions to this paper.

Online links

DATABASES

The following terms in this article are linked online to:

LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink

CLG | *IRF6*

OMIM: http://www.ncbi.nlm.nih.gov/Omim

attention-deficit hyperactivity disorder | autism | Crohn disease | Van der Woude syndrome

FURTHER INFORMATION

International Society of Twin Studies:

http://www.ists.qimr.edu.au

Access to this interactive links box is free online.

© 2002 Nature Publishing Group