Early Research on Human Genetics Using the Twin Method: Who Really Invented the Method?

Oliver Mayo
CSIRO Livestock Industries, Australia

Received 01 April, 2009; accepted 27 April, 2009.
Address for correspondence: Oliver Mayo, CSIRO Livestock Industries, PO Box 100401 Adelaide BC, SA 5000, Australia. E-mail: Oliver.Mayo@csiro.au

Keywords: concordance, DZ, MZ, history, quantitative inheritance, twin method

The twin method consists of a formal comparison between the resemblance between identical (monozygotic, MZ) twins and the resemblance between fraternal (dizygotic, DZ) twins for some trait of interest. It was developed between 1900 and about 1940, as more accurate tools for diagnosis of zygosity and for statistically analyzing the resemblance between relatives were built. Its early use was in the demonstration that a trait was inherited or that part of the causation of a trait was genetical, but it has now evolved to the point that twin registries constitute an important resource for the identification of specific genes and their interactions both with other genes and with the internal and external environment. Who really invented the method is still an unsettled question, which this article explores.

Keywords: concordance, DZ, MZ, history, quantitative inheritance, twin method
(1924). It is possible that some of the obscurities in the origin of the method arose from the usual language problems of English-speaking scientists (see e.g., Crow 1999), otherwise how could one read in the highly influential book by Newman et al. (1937, p. 7): 'The next study of importance [after Thorndike (1905)] was reported nearly 20 years later by Merriman in 1925?'

How were twins used in the early years of human genetics, following the ‘rediscovery’ of Mendel’s great work in 1900 by Correns, de Vries and von Tschermak-Seysenegg (e.g., de Vries, 1900)?

**The Twin Method as Originally Proposed for Quantitative Traits**

The method could not be developed until three milestones had been passed: a proper understanding of the difference between MZ and DZ twins, which was barely achieved by the end of the 19th century; a clearly understood and correct model for inheritance, which was ‘rediscovered’ around 1900; and a clear method for causal assignment of variability, which Fisher achieved in 1918.

The development also took place in the era when the fundamental Fisherian idea of the necessity for random samples in inference from sample to population was being developed (e.g., Fisher, 1926), so it is not surprising that random samples of twins were rarely discussed; and of course they would rarely be available. Weinberg had introduced methods for analyzing Mendelian data with biased ascertainment, so he was well aware of the problems, of course (see e.g., Weinberg, 1927), but does not appear to have discussed them in connexion with genetical estimation using twins. Dunn et al. (1993), in a thoughtful review of statistical aspects of research on depression, cited Luxenberger (1928) as being the first to address sampling issues soundly. However, Reichle (1929) was aware of these problems, possibly as a result of reading the work of Bonnevie (1924), and did not cite Luxenberger. He noted that Lauterbach (1925) and Montgomery (1926) used unselected samples in their dermatoglyphic work; these samples (like that of Newman (1929), who followed Siemens (1924) in methodology) were not randomly ascertained, however.

Weinberg (1901) was the first with a really clear understanding of twinning, building on the work of Moser and Bertillon. As he wrote, ‘Moser had already [in 1839] decided, from rather large differences of the actual distribution of the different sex combinations of the twins as compared with the probability calculation, that there must be special causes, which favor the predominance of the twins of same sex, and looked for these causes in the disparities in age of parents, which he overrated under the influence of Hofacker-Sadler.’ (Weinberg, 1901, p. 362) Weinberg clearly understood that MZ twins were identical genetically and differed only because of environmental factors, whereas DZ twins were no more alike than any other pair of siblings, apart from being the same age and having shared a very similar uterine environment. Weinberg did not, however, develop the variance-partitioning twin study.

Weinberg’s careful analysis of his own data and others’ results yielded a number of very important findings, not all of which were heeded by his successors. He concluded that: the male to female sex ratio in twins was slightly lower in singles, but the male excess was slightly higher in premature births; the sex ratio divergence from 1:1 has the same causes in single and multiple births; the sexes of successive births were not associated, though sex ratio did change with birth order; parental age influenced sex ratio, older mothers having more male births as well as more twin births; and the variation in the frequency of like-sexed pairs (which were always in substantial excess) was a fact on which his difference method of estimating the frequency of MZ twinning could be built (double the frequency of unlike sexed-pairs and subtract this from 1, p. 369). Using unlike-sexed pairs as a surrogate for DZ data and comparing them with all like-sexed pairs, Weinberg further concluded that; intra-uterine and birth problems (including prematurity and stillbirth) were less frequent in DZ twins, but much more frequent in twin than in single births; twins were smaller than singles at birth; fertility of twins was normal, in contrast both to livestock (freemartins in cattle) and to certain widespread views on the infertility of human twins (‘Perhaps these exist only in the imagination of [English] animal breeders.’ p. 388); the rate of DZ but not MZ twinning varied with season and region, being more frequent in rural than in urban populations; the rate of DZ but not MZ twinning differed between nations and populations; higher population frequencies of twinning were associated with a higher rate of DZ twinning; the rate of DZ but not MZ twinning increased with maternal age and birth order; that there was no evidence that mothers of twins were more or less fertile than other mothers; and that there was some tendency for mothers to bear more than one set of DZ twins more frequently than the chance expectation. That there is a modest genetic contribution to the causation of DZ twinning but almost none to MZ has been substantiated by later research (Lewis et al., 1996; Hoekstra et al., 2008).

Although Weinberg noted that the proximate causes of twinning must be either fertilisation of multiple eggs or splitting of the early embryo, he did not speculate further about the physiology of twinning.

Weinberg noted that Weissmann had had an early insight into the role MZ twins could play in understanding heredity: ‘The small differences between identical twins are [to Weissmann] an indication of how far external influences can change this development course.’ (p. 382). He himself thought that both kinds of twins should be considered, because the rarity of MZ twins meant that their causation might be the result of relevant heritable factors. He further noted,
‘since Darwin, the inheritance of a trait is taken to be the rule, non-inheritance the exception’ (p. 413), but that evidence for the inheritance of multiple births was incomplete and anecdotal; only thorough statistical investigation would suffice. He found little evidence that MZ twinning was inherited, but considerable evidence for DZ twinning, especially in the female line, where the statistical analysis was more direct.

Pearson (1901) introduced the intraclass correlation coefficient, the basis of simple early twin comparisons, but did not conduct any twin studies. Merriman (1924) calculated the intraclass correlation coefficients for groups of twins of like and unlike sex, obtaining the results in Table 1. For skin naevi, Siemens (1924) obtained $r_{MZ} = 0.4 \pm 0.1$ and $r_{DZ} = 0.2 \pm 0.2$, whence $h^2 = 2(r_{DZ} - r_{MZ}) = 0.4 \pm 0.4$, though it is not clear quite how good was his twin diagnosis. He explained the difference as indicating a clear difference in heredity on development of skin naevi. The two coefficients were not significantly different, but the conclusion was not unreasonable. Zhu et al. (1999) obtained $h^2 \geq 0.4$ (significantly $> 0$) in the course of a more complex and deeper analysis of a much larger sample (150 MZ pairs, 200 DZ pairs).

Fisher (1918, 1925a), through his analysis of variance, showed how to obtain the intraclass correlation simply, as $r = (MS_{between twin pairs} - MS_{within twin pairs})/(MS_{between twin pairs} + MS_{within twin pairs})$ from a one-way analysis of variance partitioning variation between and within twin pairs. On simple assumptions, a function of $r_{MZ}$ and $r_{DZ}$ could then allow the estimation of the heritability of a trait, that is, the proportion of total variance in a trait that could be attributed to genetical factors as against environmental factors. Fisher himself, though interested in this question, did not advocate the use of heritability, because as a ratio it could be highly misleading (Fisher 1951). He did analyze the results of Thorndike (1905) and Lauterbach (1925). (The latter had cited Merriman’s (1924) work approvingly.) Holzinger (1929) presented expressions that, while not statistically sound, did provide an early attempt at deriving something like the usual heritability, that is, the proportion of total phenotypic variance attributable to genetic causation. As late as 1936, $r_{MZ}$ and $r_{DZ}$ were being presented without any notion of combination or further estimation (e.g., Fukuoka, 1936).

Fisher (1919) noted that Thorndike’s correlations for ‘six mental traits range from .69 to .90 with a standard error of about ± .05’ and similarly for ‘eight physical traits’ giving ‘a general level of correlation not far from .80. This is an astonishingly high value.’ (p. 489) Fisher’s speculations on the reasons for this similarity include the idea that dizygosis could come about through an ovum splitting before fertilization, rather than after, the two resulting ova being fertilized by two sperms, so that there were three types of twins, MZ, half-identical DZ (such twins would ‘share the hereditary nature of one gamete but not of the other’, p. 496) and fraternal DZ. Thus, Fisher’s (1919) understanding of the origin of twins was defective, especially cf. Weinberg’s; he ceased to advance these cautiously phrased hypotheses over the next few years, though he still regarded diembryony as likely to be more important than dizygosity in 1928, an idea since disproved.

Fisher (1925b) showed that the like-sex twins of each sex were a heterogeneous mixture, divisible into two or more classes, the most probable division being, of course, identical (MZ) and fraternal (DZ). This meant that the like-sex correlations shown in Table 1 would not be validly calculated; such calculations awaited better identification of the MZ or DZ origin of twins.

In principle, the linear model developed by Fisher (1918, 1919) was that of Gauss (see Seal 1967) and the method could have been introduced earlier, but the understanding that continuous variation could arise from Mendelian segregation came slowly. Fisher’s (1919) variance-partitioning of twin variability was the first such application to twins.

Poll (originally Pollak, Aumüller, & Grundmann 2002) was a human geneticist and eugenicist. In that era, it was possible for a Jewish scientist or a feminist scientist (like the experimental geneticist Agnes Bluhm; see Bleker, 2007) to combine eugenics with good science and strong political views. See Braund and Sutton (2008) for an account of Poll’s life and fate.

Poll (1914a; 1914b) understood that MZ twin pairs gave an opportunity to understand the role of the environment, in its effects on two identical genotypes. However, he was hampered by the need to demonstrate unequivocally that each twin pair was MZ or DZ using the same information as would be used for the study of inheritance. He wrote (pp. 87–88): ‘MZ twins and triplets are in fact the sole humans with identical genomes, the sole isozygotic individuals: for the same sperm and the same egg should yield them the same genetical endowment, according to theory. If this idea is correct, the well-planned and critical investigation of each suspected inherited character for its modification in MZ twins must be conducted as an essential first step in all human genetics investigations.’ Thus, MZ twins would form a

---

**Table 1**

<table>
<thead>
<tr>
<th>Test</th>
<th>Like-sex</th>
<th>Unlike-sex</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stanford-Binet (intelligence quotients)</td>
<td>0.867</td>
<td>0.504</td>
<td>0.363</td>
</tr>
<tr>
<td>Army Beta (intelligence quotients)</td>
<td>0.908</td>
<td>0.732</td>
<td>0.176</td>
</tr>
<tr>
<td>National Intelligence Test (intelligence quotients)</td>
<td>0.925</td>
<td>0.867</td>
<td>0.058</td>
</tr>
<tr>
<td>Teacher estimates</td>
<td>0.654</td>
<td>0.266</td>
<td>0.388</td>
</tr>
<tr>
<td>Averages</td>
<td>0.838</td>
<td>0.592</td>
<td>0.246</td>
</tr>
</tbody>
</table>

---

Downloaded from https://www.cambridge.org/core. IP address: 54.191.40.80, on 29 May 2017 at 17:53:11, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1375/twin.12.3.237
‘control’ group whose variability would reveal environmental influences uncontaminated by genetical differences. Poll obtained data on 83 pairs of presumptive MZ twins. Although the resemblance of fingerprints between members of a pair was strong, there was no identity of pattern between members, and in many cases there was a discrepancy in the type of pattern (whorl, loop, arch, and so on). This has been borne out in subsequent research; ridge count is more concordant than pattern.

At the end of a lengthy discussion of the nearness of the fingerprints and other traits of MZ twins as against DZ twins; however, Poll could only conclude (p. 104): ‘Deductive discussions of the questions first of the origin of multiple births as such, secondly on the influences, which possibly before, during or after fertilization modify the twin-germs, finally on the roles of incomplete dominance, of throw-backs [Reversionen] and other disturbances of heritable outcomes are not difficult to separate out. The more important task lies in establishing inductively (from the extent, kind and direction [of variation]) the rules for these irregularities. Here is a wide field of study for both existing and novel techniques. The biophysics of the most similar humans would also be complemented by biochemical observation: for example, if the observations of Todd (1913) prove correct, that individuals differ in haemolytic reactions, then really identical twins must be identifiable as the most similar individual-plasmatic humans by this method.’

Bonnevie (1924) was one of the first to combine a clear understanding of twinning diagnosis, correlation and fingerprints. She realized that a comparison of MZ and DZ twins must only include those identified with complete certainty, to the extent that this was possible. She obtained the results shown in Table 2 for a measure of finger ridge count.

### Diagnosis of Zygosity

All early twin studies contained an element of circularity, in that diagnosis of zygosity often used the same traits as were then analyzed within and between the two defined groups. As Fisher (1925b) had shown, proper statistical analysis of twin data awaited certain diagnosis of zygosity. The certain diagnosis appeared at that time to be the analysis of twin data awaited certain diagnosis of zygosity. As Fisher (1925b) had shown, proper statistical analysis of twin data awaited certain diagnosis of zygosity.

**Table 2**

<table>
<thead>
<tr>
<th>Sibship type</th>
<th>No. pairs</th>
<th>r ± standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZ twins</td>
<td>15</td>
<td>0.92 ± 0.04</td>
</tr>
<tr>
<td>DZ twins</td>
<td>30</td>
<td>0.54 ± 0.08</td>
</tr>
<tr>
<td>Unlike-sexed sibs</td>
<td></td>
<td>0.60 ± 0.12</td>
</tr>
<tr>
<td>Unrelated individuals</td>
<td></td>
<td>0.27 ± 0.13</td>
</tr>
</tbody>
</table>

Table 2: Correlations for a Measure of Finger Ridge Count Obtained by Bonnevie (1924)

Veltkamp et al. (1972) performed a study of nine blood coagulation factors in 26 (14 MZ, 12 DZ) twin pairs. Ten polymorphic systems (ABO, Rh, MNs, and so on) were used to check parental and twins’ own diagnosis. For only nine of 12 DZ pairs was diagnosis certain. For the MZ pairs, the lowest probability of MZ using the 10 polymorphisms was 0.9438. (Today 10 reliable DNA microsatellite markers will yield P(DZ|concordance) < \(10^{-4}\).)

Variance in clotting factors was analyzed using a hierarchical analysis of variance and using a one-way analysis of the individual means, since the replication was not the same for all factors and all individuals, on account of the exigencies of the laboratory at the time. These two different analyses gave the results in Table 3. From the hierarchical analysis, \(h_1^2\) estimated \(\frac{(V_1+3/2V_2)}{(V_1+3/2V_2+2V_3)}\) and from the other analysis, \(h_2^2\) estimated \(\frac{(V_1+V_2)}{(V_1+V_2+V_3)}\), usually termed broad heritability, but less precisely.

**Table 3**

<table>
<thead>
<tr>
<th>Clotting factor</th>
<th>(h_1^2)</th>
<th>(h_2^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>II</td>
<td>0.555</td>
<td>0.255</td>
</tr>
<tr>
<td>V</td>
<td>0.925</td>
<td>0.723</td>
</tr>
<tr>
<td>VII</td>
<td>0.976</td>
<td>0.760</td>
</tr>
<tr>
<td>VIII</td>
<td>0.839</td>
<td>0.634</td>
</tr>
<tr>
<td>IX</td>
<td>0.404</td>
<td>0.306</td>
</tr>
<tr>
<td>X</td>
<td>0.600</td>
<td>0.462</td>
</tr>
<tr>
<td>XI</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>XII</td>
<td>1</td>
<td>0.780</td>
</tr>
</tbody>
</table>

Oliver Mayo

Downloaded from https://www.cambridge.org/core. IP address: 54.191.40.80, on 29 May 2017 at 17:53:11, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1375/twin.12.3.237
Despite all care, \( b^2 < 0 \) and \( h^2 > 1 \) were obtained in some cases, through the properties of the means squares. No non-zero estimate of \( h^2 \) was significantly different from zero or unity because of the large variances of these ratios of estimates of second-order statistics. This work was of course conducted, like hundreds of small twin studies, before Martin et al. (1978) showed unequivocally how imprecise such studies were likely to be. Jinks and Fulker (1970) in a thorough comparison of different approaches to the analysis of quantitative twin data (which showed, for example, the conceptual errors in tools like Holzinger’s (1929) ‘heritability’), advocated an analytic approach which would avoid the most obvious errors of inference arising from naive application of formulae such as those of Holzinger and Falconer.

### The Qualitative Analysis

The qualitative analysis could begin with a simple \( 2 \times 2 \) table:

<table>
<thead>
<tr>
<th></th>
<th>Concordant</th>
<th>Discordant</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZ</td>
<td>(a)</td>
<td>(b)</td>
</tr>
<tr>
<td>DZ</td>
<td>(c)</td>
<td>(d)</td>
</tr>
</tbody>
</table>

Then a correlation coefficient could be obtained very simply as

\[
\frac{(ad-bc)/((a+b)(c+d)(a+c)(b+d))^{1/2}}{.}
\]

In the case of diseases, where an individual is classified as affected or unaffected, or for discontinuous traits such as smoking habit (e.g., ‘ever smoked’ vs. ‘never smoked’), concordance (‘the proportion of co-twins affected for each twin independently ascertained’, in Smith’s (1970) words) could be calculated very simply, as affected or unaffected, or for discontinuous traits such as smoking habit (e.g., ‘ever smoked’ vs. ‘never smoked’), concordance (‘the proportion of co-twins affected for each twin independently ascertained’, in Smith’s (1970) words) could be calculated very simply, or for discontinuous traits such as smoking habit (e.g., ‘ever smoked’ vs. ‘never smoked’), concordance (‘the proportion of co-twins affected for each twin independently ascertained’, in Smith’s (1970) words) could be calculated very simply.

Siemens (1924) introduced the concordance–discordance analysis. In Newman, Freeman & Holzinger’s (1937, p. 18) words:

> Assuming that it is possible to diagnose with accuracy [MZ] and [DZ] twins, it is then reasonable to say that pathological and other characters which are always or nearly always, when they occur at all, present in both members of [MZ] pairs, but rarely or never appear in both members of [DZ] pairs, are hereditary.

Following this line of reasoning, many workers concluded from low concordance in MZ pairs that a trait was not strongly influenced by heredity. As Smith (1970), building on the work of Falconer (1965) and Lerner and Robertson (1949), showed, this is not correct: MZ concordance ‘will not be expected to be high unless the heritability is very high (or the population incidence is very high)’. Smith gave as an example the data on club foot of Idelberger (1939): MZ and DZ concordance rates in 40 MZ and 134 DZ pairs were 33% and 3% respectively, with population incidence 0.12%. Then \( h^2 = 0.8 \pm 0.3 \). For a recent discussion of these issues, see Slatkin (2008).

Twin methods can also yield valuable information on known Mendelian traits. For example, if a trait is highly variable, a study of MZ and DZ twins can yield information on the extent to which the variability is genetically and developmentally influenced, given the precisely defined expectations for Mendelian genes. Fraser (1976, p. 267) noted ‘no gross discrepancy between zygosity and concordance’ in Mendelian deafness, whereas perinatally acquired deafness was frequently discordant, as well as involving much stillbirth of twins.

### Problems With and Objections to the Method

The problem of using the same data to diagnose zygosity and to study the inheritance of one or more traits has already been mentioned. It was not recognized by Newman et al. (1937) in their pioneering text, and this was not unusual. It did not cause problems when genetical influences on a disease were being studied, diagnosis having been determined from a range of normal phenotypic traits (e.g., Diehl and von Verschuer 1933).

It was recognized by Weinberg and others that the prenatal environment would not be identical for dichorionic twin pairs, so that the assumption of identity was an approximation. Similarly, it was recognized that MZ twins might be treated differently from DZ twins by their parents, especially that they might be treated more alike; this is widely recognized across cultures, for example, Laye (1953) wrote: ‘It is our custom [in Upper-Guinea] that twins must agree on everything and that they are entitled to an equality more strict than the other children: anything given to one must immediately be given also to the other.’ (p. 76) However, these were not big differences in an era when data collection and analysis was slow and laborious; twins provided the best means for demonstrating some influence of heredity. Thus, when Fisher (1958a, 1958b) sought evidence of possible genetical influences on smoking habit, during the controversy over the now well-established causal role of smoking in lung cancer, he turned to twin data, which he obtained from noted twin researchers O. von Verschuer and E. Slater. Today one might write ‘notorious’ for von Verschuer (e.g., Ehrenreich, 2007; Trunk, 2007), but Fisher had corresponded with and met him before World War II, and was probably unaware of the extent of his involvement in ‘murderous science’ (Müller-Hill, 1984). In a set of 51 MZ and 31 DZ male twin pairs, MZ concordance was 0.65, DZ 0.35. In a set of 53 MZ and 18 DZ English female twin pairs, MZ concordance was 0.83, DZ 0.50. Twenty-seven of these English MZ pairs had been separated at birth, among whom concordance was 0.85, as against 0.81 in those brought up together. Fisher correctly recognized that these results, taken together, clearly showed that smoking habit was influenced by genotypic; recent studies confirm the role of genetics in the determination of smoking behavior.
(Goode et al., 2003; Füst et al., 2004). (That Fisher was not correct about the causation of lung cancer is not relevant here.)

See Mitchell et al. (2007) for discussion and the conclusion that, with careful attention to study design and analysis, problems arising from the certainly false assumption of identical MZ environment can be overcome.

That twins may not always be representative of the general population was recognized early (Weinberg 1901), but the manifold advantages of twin studies clearly outweighed this potential problem. It occasionally surfaces (e.g., Phillips, 1993; Tishler & Carey, 2007) but modern investigations have shown it not to be of major concern (e.g., Duffy, 1993; Kaprio 2007).

Twin Registers
Weinberg (1901) had called attention to the importance of good civic records for the use of twins in research. In many countries, informal registers were set up, but their progress was impeded by the disastrous debasement of science in the Nazi period. It is perhaps noteworthy that the British Medical Journal called for the establishment of a national twin research center, which implied a registry, at the Galton Laboratory (Anon., 1940) and R. A. Fisher, director of the Laboratory, responded enthusiastically (Fisher 1940). Nothing came of the proposal, and the first national register to be set up was that in Denmark, in 1954 (Skytthe et al., 2002). Others have followed, for example, Sweden a few years later (Lichtenstein et al., 2006), Australia in 1981 (Hopper 2002) and China in 2001 (Yang et al., 2002). For a general review, see Busjahn and Hur (2006).

As well as allowing precise genetical studies, these registries can permit study of change over time (Bartels, 2007). For example, Kendler et al. (2006) used the Swedish National Registry to confirm earlier estimates of substantial heritability for severe depression and the difference between males and females, and were also able to show that there was no evidence ‘for differences in the roles of genetic and environmental risk factors … in cohorts spanning six decades’ (p. 109).

Developments
Smith (1975), who clarified many of the statistical uncertainties in the analysis of twin data, summarized most of the useful methods. Tables 3 and 4 are examples. Knowing the relationships among variances from different relationships and the underlying variance components $V_a$, $V_c$, and so on, it is straightforward to estimate the latter.

The Method Today
Two major changes revolutionized twin studies in the second half of the 20th century.

First came computers. Their use meant that tractable normal theory approximations could largely be abandoned; more realistic, larger, more complex models could be fitted to much larger sets of data. These larger sets of data could be collected, edited, stored, shipped and combined more readily. There were many important advances; for example, in use of multivariate methods and in analysis of testing methods (e.g., Martin and Eaves, 1977; Martin et al., 1978). The new statistical approach is well summarized by Neale and Cardon (1992).

The second change was in many ways even more profound: the introduction of DNA-level mapping. This meant that the entire genome could be simultaneously marked and the association of trait values with enormous numbers of individual DNA markers analyzed.

Twin studies are no longer needed to demonstrate the influence of heredity on a trait or to estimate $b^2$, $V_s$, $V_a$, and so on. Genes of individual biological significance can now be sought, and sets of twins remain vital as a resource. Two studies illustrate their use.

Posthuma et al. (2005) conducted a genome-wide scan for chromosomal locations influencing measured intelligence. They used over 400 families from the Australian and Dutch twin registers and typed over 750 microsatellite markers in each subject, giving approximate 5cM coverage across the genome. Information from twin pairs, from non-twin sibs and (in the Australian sample) from some parents was used. Several different tests of cognitive performance were used to obtain the IQ score that was used as the measure of intelligence. Direct linkage analysis was used. Two chromosomal regions (2q24.1–2q31.1 and 6p25.3–6p22.3) were identified that were strongly (LOD score > 3 after correction for multiple testing etc.) linked to IQ score. These regions accounted for about 10% of total genetical variance in IQ, hence 5–7% of total phenotypic variance. This study did not recover the same regions found by Plomin et al. (2001) using a non-twin sample and many more DNA markers of the same kind. These studies stand in stark contrast to the thorough study, typical of the early era, by Lottig (1931) of 10 MZ and 10 DZ pairs, from which he hoped to draw deep understanding of human psychosocial variation; although his conclusions hold, for example that the MZ pairs are much more alike than the DZ in traits like musicality, no useful lessons for the future could be drawn. Furthermore, although Lottig was thoroughly conversant with the work of Poll, Siemans and others, he...
Zhu et al. (2005) conducted a genome-wide scan for chromosomal locations influencing eye color. They used about 500 families from the Australian twin registers and typed about 750 microsatellite markers in each subject, giving approximate 5cM coverage across the genome. Information from twin pairs, from non-twin sibs and from some parents was used. Eye color was rated discontinuously on a scale related to the well-understood blue–brown phenotypic polymorphism. Direct linkage analysis was used. One chromosomal region (15q) was identified that was extremely strongly (LOD score > 19 after correction for multiple testing, and so forth) linked to eye color score. This region accounted for about 74% of total genetical variance in eye color, hence over 70% of total phenotypic variance.

**Conclusions**

Bonnevie (1924) and Siemens (1924) seem to have recognized simultaneously the need to diagnose twins carefully, only to compare certainly known DZ pairs with certainly known MZ pairs, and how the correlations within pairs could be used. Weinberg (1901) and Poll (1914) had recognized the first two requirements. Correct estimation of genetical components of variance, following Fisher (1918), seems to have been fully realized only by Jinks & Fulker (1970). Although modern investigators avoid the inference problem (using the same traits to diagnose zygosity and estimate genetical parameters) through the use of very large numbers of DNA markers or separate sets of such markers, it is not clear when and by whom the problem was solved.

**Acknowledgments**

I thank J. Braund, D. Duffy, G. R. Fraser, D. Früh, C. R. Leach and N. G. Martin for assistance of various kinds with this paper. D. Duffy in particular corrected a number of errors in early drafts of the article.

**References**


