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

Oliver Mayo

CSIRO Livestock Industries, Australia

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

The twin method, as generally understood today, 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. As Lottig (1931) put it, in a careful review, 'The essence of the twin method lies in the hereditary equality of the MZ twins accepted by most researchers. DZ twins arise however from the fertilization of two eggs. Hence, DZ twins have no more hereditary similarity than usual brothers and sisters' (pp. 3–4).

Rende et al. (1990) attribute the method to Merriman (1924) and Siemens (1924), and argue that it is very strange that there was little progress in twin studies between Thorndike (1905) and the early twenties. This argument ignores almost the entire German-speaking research world over the relevant period. It is therefore worth exploring this early research further, to clarify the discovery and development of the method.

MZ twins are genetically identical, and if reared together share their environment. DZ twins share half of their genes, like any other pair of siblings, and if reared together they also share their environment. The difference in resemblance (concordance) between the two members of a pair of MZ twins and the difference for a pair of DZ twins can then be compared (combined over many sets of twins), and functions of this difference then give a measure of the degree to which genetics contributes to variation in the trait. The assumptions about the shared or common environment of a pair of twins, viz. that differences between MZ twins would be an indication of the amount of environmental variability, whereas those between DZ twins would include a genetical component, appeared reasonable and modest. Objections and problems were identified early.

Concordance could be either qualitative (e.g., whether both members of a twin pair suffer from the same disease) or quantitative (e.g., difference in stature of members of a twin pair). Very different analytical methods were developed to deal with categorical and continuous variables.

The origins of the method are frequently and correctly traced to Galton (1875). Rende et al. (1990) claimed that he did not really make the fundamental distinction of MZ genetical identity and DZ half-identity. With his understanding of heredity, this was an impossible step, but he was very close; he wrote (Galton, 1883, p. 157): '[unlike-sex twins are] never due to the development of two germinal spots in the same ovum', and '[the MZ/DZ dichotomy explained] ... a curious discontinuity in my results ... Extreme similarity and extreme dissimilarity between twins of the same sex are nearly as common as moderate resemblance'.

It was recognized early that twin analysis might be particularly suited to behavioral traits (e.g., Thorndike 1905), given the difficulty of any other approach.

Rende et al. (1990) conclude that Siemens (1924) and Merriman (1924) jointly discovered it, but they do not consider the fundamental work of Poll and Weinberg, and their predecessors, nor Bonnevie

(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 than 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 genetical 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 $h^2 = 2(r_{MZ} - r_{DZ}) \approx 0.4 \pm 0.4$, though it is not clear quite how good was his twin diagnosis. He explained the difference as indicating a clear influence of 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 $r = (MS_{\text{between twin pairs}} - MS_{\text{within twin pairs}})/(MS_{\text{between twin pairs}} + MS_{\text{within twin pairs}})$ 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 genetic 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 $\pm$ .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

Intraclass Correlation Coefficient for Like- and Unlike-Sex Twins (Merriman, 1924)

Test	Like-sex	Unlike-sex	Difference
Stanford-Binet (intelligence quotients)	0.867	0.504	0.363
Army Beta (intelligence quotients)	0.908	0.732	0.176
National Intelligence Test (intelligence quotients)	0.925	0.867	0.058
Teacher estimates	0.654	0.266	0.388
Averages	0.838	0.592	0.246

'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 presence of one chorion or two and one placenta or two. However, use of Weinberg's differential method to determine the frequency of MZ twinning at a population level gave a higher range 26–36% (von Verschuer 1927) than the frequency of monochorionic births, 14–26% (Komai, 1928). Even if there had been no uncertainty about this criterion, however, it was in many cases not known, since such information was not all recorded, nor provided in a nonperishable form to parents of twins or twins themselves.

Komai (1928) suggested, following Muller (1925), the combination of information on as many traits as possible, especially palm prints. Muller had suggested a probabilistic combination of information, and this

Table 2

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

Sibship type	No. pairs	$r \pm$ standard error
MZ twins	15	0.92 ± 0.04
DZ twins	30	0.54 ± 0.08
Unlike-sexed sibs		0.60 ± 0.12
Unrelated individuals		0.27 ± 0.13

would clearly be the best approach, though not straightforward with a mixture of quantitative and qualitative (discontinuous) data. Qualitative data, such as the ABO blood groups, could often provide certainty of diagnosis of DZ, for example, if an AB mother and an O father had a pair of daughters, one A and the other B, they would certainly be DZ. However, even with many polymorphic systems, diagnosis could not be certain. An example will perhaps give a hint of the difficulties.

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, MNSs, 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 $(V_A + 3/2 V_D)/(V_A + 3/2 V_D + 2 V_E)$ and from the other analysis, h_2^2 estimated $(V_A + V_D)/(V_A + V_D + V_E)$, usually termed broad heritability, but less precisely.

Table 3

Heritabilities (h^2) Estimated in Two Ways (see text for details) for Nine Clotting Factors by Veltkamp *et al.* (1972)

Clotting factor	h_1^2	h_2^2
I	0	0
II	0.555	0.255
V	0.925	0.723
VII	0.976	0.760
VIII	0.839	0.634
IX	0.404	0.306
X	0.600	0.462
XI	1	0
XII	1	0.780

Despite all care, $h^2 < 0$ and $h^2 > 1$ were obtained in some cases, through the properties of the mean 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 naïve application of formulae such as those of Holzinger and Falconer.

The Qualitative Analysis

The qualitative analysis could begin with a simple 2×2 table:

	Number of pairs	
	Concordant	Discordant
MZ	a	b
DZ	c	d

Then a correlation coefficient could be obtained very simply as

$$(ad-bc)/((a+b)(c+d)(a+c)(b+d))^{½}.$$

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, and heritability, relative risk etc. estimated.

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 genotype; 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_D, 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.

Table 4

Genetical Covariances Among Relatives (from Smith, 1975)

Relationship	Additive genetical variance	Dominance genetical variance
	V_A	V_D
Monozygotic twins	1	1
Dizygotic twins	½	¼
Full sibs	½	¼
Parent and offspring	½	0

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 h^2, V_A, V_D, 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, Siemens and others, he

attempted no quantitative genetical analysis of his very extensive data.

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

References

Anonymous. (1940). Value of research on twins. *British Medical Journal, 1*, 261–262.

Aumüller, G., & Grundmann, K. (2002). Anatomy during the Third Reich: The Institute of Anatomy at the University of Marburg, as an example. *Annals of Anatomy, 184,* 295–303.

Bartels, M. (2007). An update on longitudinal twin and family studies. *Twin Research and Human Genetics, 10,* 1–2.

Bleker, J. (2007). Agnes Bluhm, die Wissenschaftlerin unter den Rassenhygieniker, und die Frauenbewegung. *Acta Historica Leopoldina, 48,* 89–111.

Bonnevie, K. (1924). Studies on papillary twins patterns of human fingers. *Journal of Genetics, 15*, 1–23.

Braund, J. & Sutton, D. G. (2008). The case of Heinrich Wilhelm Poll (1877–1939): A German-Jewish geneticist, eugenicist, twin researcher, and victim of the Nazis. *Journal of the History of Biology, 41,* 1–35.

Busjahn, A., & Hur, Y. (2006). Twin registries: an ongoing success story. *Twin Research and Human Genetics, 9,* 705–6.

Crow, J. F. (1999). Hardy, Weinberg and language impediments. *Genetics, 152*, 821–825.

de Vries, H. Sur la loi de disjunction des hybrids. *Comptes rendus de l'Academie des Sciences, 130,* 26 mars.

Diehl, K., & von Verschuer, O. (1933). *Zwillingstuberkulose: Zwillingsforschung und erbliche Tuberkulosedisposition* (*Tuberculosis in Twins: twin research and the inherited predisposition to TB*). Fischer, Jena.

Duffy, D. L. (1993). Twin studies in medical research. *Lancet, 341,* 1418–9.

Dunn, G., Sham, P. C., & Hand, D. J. (1993). Statistics and the nature of depression. *Journal of the Royal Statistical Society A, 156,* 63–87

Ehrenreich, E. (2007). Othmar von Verschuer and the 'scientific' legitimization of Nazi anti-Jewish policy. *Holocaust and Genocide Studies, 21,* 55–72.

Falconer, D. S. (1965). The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics, 29,* 51–76.

Fisher, R. A. (1919). The genesis of twins. *Genetics, 4,* 489–499.

Fisher, R. A. (1925a). *Statistical methods for research workers*. Oliver & Boyd, Edinburgh.

Fisher, R. A. (1925b). The resemblance between twins, a statistical examination of Lauterbach's measurements. *Genetics, 10,* 569–579.

Fisher, R. A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain, 33,* 503–513.

Fisher, R. A. (1928). Triplet children in Great Britain and Ireland. *Proceedings of the Royal Society, Series B, 102,* 286–311.

Fisher, R. A. (1940). Twins research centre. *British Medical Journal, 1,* 363–364.

Fisher, R. A. (1951). Limits to intensive production in animals. *British Agricultural Bulletin*, 4, 217–218.

Fraser, G. R. (1976). *The causes of profound deafness in childhood: A study of 3,535 individuals with severe hearing loss present at birth or of childhood onset.* Johns Hopkins University Press, Baltimore.

Fukuoka, G. (1936). A study on the variation and heredity of finger patters. *Japanese Journal of Genetics, 12,* 81–88.

Füst, G., Arason, G. J., Kramer, J., Szalai, C., Duba, J., Yang, Y., Chung, E. K., Zhou, B., Blanchong, C. A., Lokki, M. L., Bödvarsson, S., Prohászka, Z., Karádi, I., Vatay, A., Kovács, M., Romics, L., Thorgeirsson,

G. and Yu, C. Y. (2004). Genetic basis of tobacco smoking: Strong association of a specific major histocompatibility complex haplotype on chromosome 6 with smoking behavior. *International Immunology, 16*, 1507–1514.

Galton, F. (1875). The history of twins, as a criterion of the relative powers of nature and nurture. *Fraser's Magazine*, November, 566–576.

Goode, E. L., Badzioch, M. D., Kim, H., Gagnon, F., Rozek, L. S., Edwards, K. L., and Jarvik, G. P. (2003). Multiple genome-wide analyses of smoking behavior in the Framingham Heart Study. *BMC Genetics, 4*, S102. Published online 2003 December 31. doi: 10.1186/1471–2156-4-S1-S102.

Hoekstra, C., Zhao Z. Z., Lambalk, C. B., Willemsen, G., Martin, N. G., Boomsma, D. I. & Montgomery, G. W. (2008). Dizygotic twinning. *Human Reproduction Update, 14*, 37–47.

Holzinger, K. J. (1929). The relative effect of nature and nurture influences on twin differences. *Journal of Educational Psychology, 20*, 241–248.

Hopper, J. L. (2002). The Australian Twin Registry. *Twin Research, 5*, 329–336.

Idelberger, K. (1939). Die Ergebnisse der Zwillingsforschung beim angeborenen Klumpfuss (The results of a twin investigation into congenital clubfoot). *Verhandlungen der deutschen orthopädischen Gesellschaft, 31*, 1–20.

Jinks J. L., & Fulker D. W. (1970). Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. *Psychological Bulletin, 73*, 311–349.

Kaprio, J. (2007). Differences in smoking habits of MZ and DZ twins: A commentary on Tishler and Carey. *Twin Research and Human Genetics, 10*, 718–720.

Kendler, K. S., Gatz, M., Gardner, C. O., & Pedersen, N. L. (2006). A Swedish National Twin Study of lifetime major depression. *American Journal of Psychiatry, 163*, 109–114.

Komai, T. (1928). Criteria for distinguishing identical and fraternal twins. *Quarterly Review of Biology, 3*, 408–418.

Lauterbach, C. E. (1925). Studies in twin resemblance. *Genetics, 10*, 525–568.

Laye, C. (1953). *L'enfant noir*. Plon, Paris.

Lewis, C. M., Healey, S. C., & Martin, N. G. (1996). Genetic contribution to DZ twinning. *American Journal of Medical Genetics, 61*, 237–246.

Lichtenstein, P., Sullivan, P. F., Cnattingius, S., Gatz, M., Johansson, S., Carlström, E., Björk, C., Svartengren, M., Wolk, A., Klareskog, L., de Faire, U., Schalling, M., Palmgren, J. and Pedersen, N. L. (2006). The Swedish Twin Registry in the third millennium: An update. *Twin Research and Human Genetics, 9*, 875–882.

Lottig, H. (1931). Hamburger Zwillingsstudien. Anthropologische und charakterologische Unter-

suchungen an ein- und zweieiigen Zwillingen (Hamburg twin studies. Anthropological and phenotypic investigations of MZ- and DZ-twins). *Zeitschrift für angewandte Psychologie. Beiheft 61*.

Luxenburger, H. (1928). Vorläufiger Bericht über psychiatrische Serienuntersuchungen an Zwillingen [Preliminary report on psychiatric investigations of a series of twins]. *Zeitschrift für die gesamte Neurologie und Psychiatrie 116* 297–347.

Martin, N. G., & Eaves, L. J. (1977). The genetical analysis of covariance structure. *Heredity, 38*, 79–95.

Martin, N. G., Eaves, L. J., Kearsey, M. J., & Davies, P. (1978). The power of the classical twin study. *Heredity, 40*, 97–116.

Merriman, C. (1924). The intellectual resemblance of twins. *Psychological Monographs, 33*, 1–58.

Mitchell, K. S., Mazzeo, S. E., Bulik, C. M., Aggen, S. H., Kendler, K. S. and Neale, M. C. 2007 An investigation of a measure of twins' early environment. *Twin Research and Human Genetics, 10*, 840–847.

Montgomery, R. B. (1926). Sole patterns of twins. *Biological Bulletin, 50*, 293–300.

Mueller, B. (1930). Untersuchungen über die Erblichkeit von Fingerbeerenmustern unter besonderer Berücksichtigung rechtlicher Fragestellungen (Investigations on the heritability of fingertip patterns with special consideration of legal questions). *Molecular and General Genetics, 56*, 302–382.

Muller, H. J. (1925). Mental traits and heredity. *Journal of Heredity, 16*, 433–455.

Müller-Hill, B. (1984). *Tödliche Wissenschaft*. Rowolt, Reinbek. (translated GR Fraser 1988 *Murderous Science*. Oxford University Press.)

Neale, M. C., & Cardon, L. R. (1992). *Methodology for Genetic Study of Twins and Families*. Kluwer, Dordrecht.

Newman, H. H. (1929). Studies of human twins. I. Methods of diagnosing monozygotic and dizygotic twins. *Biological Bulletin, 55*, 283–297.

Newman, H. H., Freeman, F. N., & Holzinger, K. J. (1937). *Twins: A study of heredity and environment*. University of Chicago Press.

Phillips, D. I. W. (1993). Twin studies in medical research: Can they tell us whether diseases are genetically determined? *Lancet, 341*, 1008.

Plomin, R., Hill, L., Craig, I. W., McGuffin, P., Purcell, S., Sham, P., Lubinski, D., Thompson, L. A., Fisher, P. J., Turic, D., & Owen, M. J. (2001). A genome-wide scan of 1842 DNA markers for allelic associations with general cognitive ability: A five-stage design using DNA pooling and extreme selected groups. *Behavior Genetics, 31*, 497–509.

Poll, H. (1914a). Über Vererbung beim Menschen. *Die Grenzboten, 73*, 247–259 & 296–311.

Poll, H. (1914b). Über Zwillingsforschung als Hilfsmittel menschlicher Erbkunde. *Zeitschrift für Ethnologie, 46,* 87–105.

Posthuma, D., Luciano, M., de Geus, E. J. C., Wright, M. J., Slagboom, P. E., Montgomery, G. W., Boomsma, D. I., & Martin, N. G. (2005). A genomewide scan for intelligence identifies quantitative trait loci on 2q and 6p. *American Journal of Human Genetics, 77,* 318–326.

Reichle, H. S. (1929). The diagnosis of the type of twinning. 1. Dermatoglyphics. *Biological Bulletin, 56,* 164–176.

Rende, R. D., Plomin, R., & Vandenberg, S. G. (1990). Who discovered the twin method? *Behavior Genetics, 20,* 277–285.

Robertson, A., & Lerner, I. M. (1949). The heritability of all-or-none traits: Viability of poultry. *Genetics, 34,* 395–411.

Schönemann, P. (1997). On models and muddles of heritability. *Genetica, 99,* 97–108.

Seal, H. (1967). Studies in the history of probability and statistics. XV: The historical development of the Gauss linear model. *Biometrika, 54,* 1–24.

Siemens, H. (1924). *Die Zwillingspathologie* [*Pathology of Twinning*]. Springer-Verlag, Berlin.

Skytthe, A., Kyvik, K., Holm, N. V., Vaupel, J. W., & Christensen, K. (2002). The Danish Twin Registry: 127 birth cohorts of twins. *Twin Research, 5*, 352–357.

Slatkin, M. (2008). Exchangeable models of complex inherited diseases. *Genetics, 179,* 2253–2261

Smith, C. (1970). Heritability of liability and concordance in monozygous twins. *Annals of Human Genetics, 35,* 85–91.

Smith, C. (1975). Quantitative inheritance. Pp. 382–441. In *Textbook of Human Genetics* (Eds G. R. Fraser & O. Mayo) Blackwell Scientific, Oxford.

Thorndike, E. L. (1905). Measurements of twins. *Archives of Philosophy, Psychology, and Scientific Methods*, No. I. vi + 64 pp. New York: The Science Press.

Tishler, P. V., & Carey, V. J. (2007). Can comparison of MZ- and DZ-concordance rates be used invariably to estimate heritability? *Twin Research and Human Genetics, 10,* 712–717.

Todd, C. (1913). On the recognition of the individual by haemolytic methods. *Journal of Genetics, 3,* 123–130.

Trunk, A. (2007). Zweihundert Blutproben aus Auschwitz Ein notorisches Forschungsvorhaben und die Frage dem Beitrag Adolf Butenandts (Two hundred blood tests from Auschwitz a notorious research project and the question of Adolf Butenandt's contribution). *Acta Historica Leopoldina, 48,* 9–40.

Veltkamp, J. J., Mayo, O., Motulsky, A.G. and Fraser, G.R. (1972). Blood coagulation factors I, II, V, VII, VIII, IX, X, XI and XII in twins. *Human Heredity, 22,* 102-117.

Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era: Concepts and misconceptions. *Nature Genetics, 9,* 255–266.

von Verschuer, O. (1927). Die vererbungsbiologische Zwillingsforschung. Ihre biologischen Grundlagen. Studien an 102 eineiigen und 45 gleichgeschlechtichen zweieiigen Zwillings- und 2 Drillingspaaren (Genetical twin research. Studies of 102 MZ and 45 same-sex DZ twin pairs and 2 sets of triplets). *Ergebnisse der Inneren Medizin und Kinderheilkunde, 31,* 35–120.

Weinberg, W. (1901). Beiträge zur Physiologie und Pathologie der Mehrlingsgeburten beim Menschen (Contributions to the physiology and pathology of human multiple births). *Pflugers Archiv für gesamte Physiologie des Menschen und der Tiere, 88,* 346–430.

Weinberg, W. (1927). Mathematische Grundlagen der Probandmethode (Mathematical foundations of the proband method). *Zeitschrift für induktive Abstammungs- und Vererbungslehre, 48,* 179–228.

Yang, H., Li, X., Cao, W., Lu, J., Wang, T., Zhan, S., Hu, Y., & Li, L. (2002). Chinese National Twin Registry as a resource for genetic epidemiologic studies of common and complex diseases in China. *Twin Research, 5,* 347–351.

Zhu G., Duffy, D. L., Eldridge A., Grace M., Mayne C., O'Gorman L., Aitken J. F., Neale M. C., Hayward N. K., Green A. C., & Martin N. G. (1999). A major quantitative-trait locus for mole density is linked to the familial melanoma gene CDKN2A: A maximum-likelihood combined linkage and association analysis in twins and their sibs. *American Journal of Human Genetics, 65,* 483–492.

Zhu, G., Evans, D. M.1, Duffy, D. L., Montgomery, G. W., Medland, S. E., Gillespie, N. A., Ewen, K. R., Jewell, M., Liew, Y. W., Hayward, N. K., Sturm, R. A., Trent, J. M., & Martin, N. G. (2004). A genome scan for eye color in 502 twin families: Most variation is due to a QTL on chromosome 15q. *Twin Research, 7,* 197–210.