Twin Research (2000) 3, 294–298

www.nature.com/tr

Estimation of penetrance from twin data

Leide A Praxedes and PA Otto

Departamento de Biologia, Universidade de São Paulo, Brazil

A simple method for estimating the gene frequency p and the penetrance value K from data on polymorphic monogenic characteristics on monozygotic twin pairs is presented. In spite of the method here presented having limited value because the results it yields cannot be evaluated on their own, the estimates of p and K it provides can be indirectly tested by comparing them to the ones obtained in familial aggregates through classical segregation analysis or by using the latter to calculate the expected proportions of dominant–dominant, dominant–recessive and recessive–recessive monozygotic twin pairs. When the method is applied to data on tongue-rolling ability published in the literature, a good agreement is observed between twin and familial estimates, thus indicating that the method is reliable and that it can be used as an ancillary way of corroborating or otherwise evidence of monogenic autosomal dominant mechanism inferred from the analysis of familial data. Twin Research (2000) 3, 294–298.

Keywords: genetics, statistics, methodology, incomplete penetrance, tongue-rolling ability

Introduction

The issue of incomplete penetrance has received in the past a considerable amount of attention in the literature. Its concept, after being correctly introduced by Vogt,[1] was successively modified or generalised by several authors.[2–20] In the present paper, we shall adopt the concept of Rogatko[21] and Rogatko et al,[22] who used for the definition of penetrance the following transitional matrix p_{ij}:

		f	
	f_1	f_2	f_3
a_1a_1	p_{11}	p_{12}	p_{13}
a_1a_2	p_{21}	p_{22}	p_{23}
a_2a_2	p_{31}	p_{32}	p_{33} ,

where the penetrance (value or coefficient) is the conditional probability p_{ij} that associates the phenotype f_k with the genotype a_ma_n; for example, the conditional probability $p(f_1| a_1a_1)$ of an individual with genotype a_1a_1 presenting the phenotype f_1 is the penetrance coefficient p_{11}.

Many of the above-mentioned papers proposed segregation analysis methods to cope with penetrance estimation from family data; at present, the parameter can be estimated on a routine basis using complicated computer-assisted complex segregation analysis methods (for a discussion on the theme, see Morton[20,23]). A previous publication,[24] however, has shown that some situations can be dealt satisfactorily with relatively simple segregation analysis models using family data, from which the penetrance coefficient can be estimated with no difficulties or complications. The aim of the present communication is to show that the same is true in relation to the estimation of the parameter from random samples of monozygotic twin pairs.

There exists, in the literature, a number of papers dealing with the estimation of penetrance values K from data on monozygotic and dizygotic twin pairs. Some authors have proposed algebraically equivalent methods for obtaining the penetrance value from the concordance rate observed among pairs of identical twins (Schinz,[25] Lasker,[26] Allen[27] and Pfändler[28]). In all these papers, using different symbols, the authors obtained the penetrance parameter K directly or indirectly from $K = 2p_1/(1 + p_1)$ or $K = 2n_1/(2n_1 + n_2)$, where p_1 is the concordance rate and n_1 and n_2 are respectively the observed numbers of affected–affected and affected–normal monozygotic twin pairs. The method is simple, correct and effective but uses truncated data with exclusion of pairs where both twins are normal and can be applied to pathological or monomorphic genetic traits only. Rife[29] developed simple monogenic models using data on monozygotic and dizygotic twin pairs for estimating the gene frequency of monogenic polymorphic traits, assuming fixed penetrance values; following this author, several others have developed twin methods to deal specifically with the complicated issue of handedness, a trait strongly influenced by environmental factors (the

Correspondence: PA Otto, Departamento de Biologia, Universidade de São Paulo, Caixa Postal 11461, 05422-970, São Paulo SP, Brazil. Tel: 55 11 3818 7591; Fax: 55 11 3818 7553;
E-mail: otto@usp.br
Received 31 May 2000; revised 19 August 2000; accepted 22 August 2000

https://doi.org/10.1375/twin.3.4.294 Published online by Cambridge University Press

important paper by Laland et al[30] lists several recent references to this particular subject).

Recently, Otto et al[24] modified the familial segregation method originally proposed by Snyder,[31,32] making room in it for estimating, besides the gene frequency p, the penetrance value K. Three penetrance models (I, II and IV) proposed in the above-mentioned paper fitted well-observed familial data on tongue-rolling ability and are summarised below, where K_1, K_2 and K_3 are the respective probabilities of AA, Aa and aa individuals presenting the dominant phenotype:

	K_1	K_2	K_3
I	$1-(1-K)^2$	K	0
II	1	K	0
IV	1	1	K

These models reduce to the penetrance concept using conditional transition matrices introduced before. For instance, in the case of model I the matrix p_{ij} is

	dom	rec
AA	$1-(1-K)^2$	$(1-K)^2$
Aa	K	1-K
aa	0	1

where dom (dominant) and rec (recessive) are the two phenotypes admitted in the models.

Penetrance model II is a standard one used in human genetics. Model I assumes that in the dominant homozygote AA the effects of the two A genes are independent. Since it might be as reasonable to assume that the recessive has reduced penetrance as has the dominant, in model IV the recessive genotype has reduced penetrance. In the present paper we develop a method for estimating the parameters p and K from data on monozygotic twin pairs using the same three models, which are very similar. In fact, since the penetrance value is generally high, the penetrance value of AA homozygotes in model I, $1-(1-K)^2$, is always near unity, which makes the estimates obtained by models I and II (where the penetrance value of the homozygous dominant genotype is assumed to be 1) very similar. As to models I and IV, they are completely equivalent, the estimates of p and K obtained in each of them leading to same expected proportions of recessive individuals in the population and in the offspring of dom × dom, dom × rec and rec × rec couples.

Models for estimating gene frequency and the penetrance value

Assuming[15] that the non-penetrance of a genetic trait is the lack of phenotypic manifestation due exclusively or predominantly to environmental factors, we can obtain the expected proportions of dominant–dominant [P_1 = P(dom–dom)], dominant–recessive [P_2 = P(dom–rec)] and recessive–recessive [P_3 = P(rec–rec)] monozygotic twin pairs in panmitic populations. The final expressions, as functions of p (frequency of the dominant allele) and K (penetrance value) are shown in Table 1.

Let us now suppose that, in a sample of N pairs of monozygotic twin pairs, n_1 are dom–dom, n_2 dom–rec and n_3 rec–rec; the likelihood function, in logarithmic form, is given by: $L = \Sigma\, n_i \log P_i = n_1 \log P_1 + n_2 \log P_2 + n_3 \log P_3$. The maximum likelihood estimates p and K are the solutions of the set of equations by putting $\partial L/\partial K = 0$ and $\partial L/\partial p = 0$. Because generally it is not possible to obtain explicit solutions for this set of equations, iterative numerical methods (such as the generalised Newton-Raphson method) are used instead. It is possible, however, using other simple algebraic argument, to obtain explicit solutions for the models under the assumption of panmixia. For instance, in model I the explicit solutions taken directly from the algebraic manipulation of the expressions for P_1, P_2 and P_3 shown in Table 1 are

$$K = 2 - [1 - \sqrt{(n_3/N)}]/\{1 - \sqrt{[(n_2 + 2n_3)/2N]}\}$$

and

$$p = \{1 - \sqrt{[(n_2 + 2n_3)/2N]}\}/K$$

These explicit solutions (as well as the corresponding ones to models II and IV), however, do not take into account random sample deviations from panmixia; furthermore, they do not permit the exact calculation of the standard errors of p and K, that, on the contrary, is directly provided by the inspection

Table 1 Expected frequencies of dom–dom, dom–rec and rec–rec monozygotic twin pairs in random-mating populations in models I, II and IV

Model	P_1=P (dom–dom)	P_2=P (dom–rec)	P_3=P(rec–rec)
I	$pK^2[2(1+p)-pK(4-K)]$	$2pK(1-K)\,[2-pK\,(3-K)]$	$[1-pK\,(2-K)]^2$
II	$p\,[p+2\,(1-p)\,K^2]$	$4p\,(1-p)\,K(1-K)$	$(1-p)\,[1-p+2p\,(1-K)^2]$
IV	$1-(1-p)^2\,(1-K^2)$	$2(1-p)^2K\,(1-K)$	$(1-p)^2\,(1-K)^2$

p: frequency of the dominant allele; K: penetrance value.

https://doi.org/10.1375/twin.3.4.294 Published online by Cambridge University Press

of the variance–covariance matrix of the Newton-Raphson method, evaluated at the estimated points for p and K. Since the explicit solutions shown above must coincide, for samples with exact Hardy-Weinberg proportions, with those obtained by the maximum likelihood method, they can be used as starting numerical values in the Newton-Raphson procedure.

The maximum likelihood estimates of p and K cannot be tested directly in the samples from which they were drawn because there exist in each model three different classes (corresponding to P_1, P_2 and P_3) and two different parameters (gene frequency p and penetrance value K), besides the sample size N, should be extracted from the sample to calculate the expected numbers of pairs. In spite of the method having apparently limited value because the results it yields cannot be evaluated on their own, the models can, however, be indirectly tested (a) by comparing the confidence intervals of the twin estimates of p and K to those obtained from familial data drawn from similar populations, a good match between them indicating that the estimates are appropriate. Besides that, (b) the estimates of p and K obtained from the analysis of independent family data can be used for calculating the expected numbers of monozygotic twin pairs in each model, a good fitting obtained in χ^2 statistics indicating that the parameters fit well the data in the model being tested.

The models described above allow estimation of the values of p and K from random samples of twin pairs analysed in relation to polymorphic characteristics exhibiting incomplete penetrance. To test the models we used published material on tongue-rolling, a trait that can be satisfactorily explained by an autosomal dominant mechanism with incomplete penetrance.[24] In relation to the distribution of the trait among twin pairs, we were able to locate in the literature four samples of monozygotic dizygotic twin pairs[33–36] listed in Table 2. Data on familial distribution of the trait, which we used for comparing the estimates with those obtained using the twin pairs method, were taken from the combined samples of Sturtevant,[37] Vogel[34] and Otto et al.[24] The last authors obtained also the familial estimates of p and K for the three models shown in Table 3.

Results

Applying to the individual samples above described the methods just presented, we obtained in all instances consistent estimates of p and K. Since the data regarding the distribution of the trait among monozygotic twin pairs were homogeneous between samples, as shown by heterogeneity χ^2 tests performed on contingency tables, we present in Table 3 the estimates of p and K and of their respective 95% confidence intervals obtained by agglutinating the data of all four samples.

Discussion

Inspection of our results reveals an almost perfect match for each of models I, II and IV of the confidence intervals of both p and K obtained from family and monozygotic twin data.

The twin estimates cannot be tested directly, as discussed before. However, the familial estimates can be applied to each of the corresponding twin models to calculate the expected numbers of dom–dom, dom–rec and rec–rec monozygotic twin pairs. Then, since the parameters p and K were extracted from similar but statistically independent samples, the observed and expected numbers of types of twin

Table 2 Data on distribution of tongue rolling ability among pairs of monozygotic twin pairs

Sample	n_1 (dom–dom)	n_2 (dom–rec)	n_3 (rec–rec)	Total
Matlock[33]	18	7	8	33
Vogel[34]	62	16	14	92
Reedy et al[35]	43	7	11	61
Martin[36]	15	8	5	28
Total	138	38	38	214

dom=roller; rec=non-roller.

Table 3 Estimates of p and K obtained from pooled data on tongue rolling among monozygotic twin pairs (mz) and families (f)

Mod.	sample	p	se (p)	K	se (K)	95% confidence intervals p	95% confidence intervals K
	mz	0.602	0.035	0.804	0.032	0.532–0.672	0.741–0.868
I	f	0.573	0.030	0.798	0.035	0.513–0.633	0.728–0.868
	mz	0.615	0.045	0.750	0.066	0.526–0.704	0.618–0.882
II	f	0.618	0.061	0.684	0.114	0.496–0.740	0.456–0.912
	mz	0.368	0.032	0.333	0.051	0.305–0.431	0.231–0.435
IV	f	0.346	0.027	0.311	0.050	0.292–0.400	0.211–0.411

se=standard error; 95% confidence intervals; p±1.96 se(p), K±1.96 se(K). Familial estimates were taken from reference 24.

https://doi.org/10.1375/twin.3.4.294 Published online by Cambridge University Press

Table 4 Results of χ^2 tests (d.f.=2) for fitting expected numbers of dom–dom, dom–rec and rec–rec monozygotic twin pairs (calculated using the estimates p and K obtained from family data) to the observed ones

Model	P	K	Observed numbers			Expected numbers			χ^2
			dom–dom	dom–rec	rec–rec	dom–dom	dom–rec	rec–rec	
I	0.573	0.798				131.33	39.26	43.41	1.05
II	0.618	0.684	138	38	38	129.00	43.68	41.32	1.63
IV	0.346	0.311				131.32	39.23	43.45	1.06

pairs can be compared using the usual χ^2 statistics with two degrees of freedom. The results of such tests are shown in Table 4.

The results of the tests indicate an excellent fit for all three models, thus corroborating the findings obtained by the comparison between confidence intervals of twin and familial estimates.

Estimates of p and K could also be obtained straightforwardly from random samples of dizygotic twin pairs. The details are omitted here, but the corresponding expressions for P_1 = P(dom–dom), P_2 = P(dom–rec) and P_3 = P(rec–rec) in model I, for example, are $P_1 = pK^2(4 + 12p - 8pK + pK^2 - 8p^2K + 2p^2K^2 + p^3K^2)/4$, $P_2 = pK(8 - 4K - 16pK + 8pK^2 - pK^3 + 8p^2K^2 - 2p^2K^3 - p^3K^3)/2$ and $P_3 = (2 - 4pK + pK^2 + p^2K^2)^2/4$. These expressions, those obtained in models II and IV, and the corresponding likelihood expressions are far more complicated than the ones obtained for the case of monozygotic twins. In addition, only two reliable samples describing the distribution of tongue-rolling ability among dizygotic twin pairs[34,36] could be located in the literature; one additional dizygotic sample (described in Reedy et al[35]) had to be discarded because its data were frankly heterogeneous in relation to the other two. The size of the combined dizygotic sample was thus small. This and the complicated likelihood expressions used for estimating p and K in the three models explain the large standard errors associated with the estimated parameters. In any case, these were obtained without difficulty and were (estimate ± 1 s.e.): a) for model I: p = 0.682 ± 0.141 and K = 0.671 ± 0.136; b) for model II: p = 0.657 ± 0.116 and K = 0.606 ± 0.233; c) for model IV: p = 0.250 ± 0.120 and K = 0.478 ± 0.168. The results of χ^2 tests for fitting expected numbers using family estimates were, respectively, for models I, II and IV, 0.96, 0.87 and 0.96, thus indicating, as in the case of monozygotic twins, an excellent fit.

All these facts indicate that the method here presented, in spite of its limitations, can be used as an ancillary way of corroborating or otherwise evidence of monogenic autosomal dominant mechanism inferred from the analysis of familial data through classical segregation analysis.

Acknowledgements

Many useful observations and suggestions from two anonymous referees from the Editorial Board of Twin Research have been incorporated in the final (revised) version of the manuscript. This article was partly supported by a grant from CNPq, Brazil.

References

1 Vogt O. Psychiatrisch wichtige Tatsachen der zoologisch-botanischen Systematik. Zeitschr Ges Neurol Psychiat 1926; 101: 805–832.
2 Strömgren E. Beiträge zur psychiatrischen Erblehre. Munksgaard, Copenhagen, 1938.
3 Kealin A. Statistisches Prüf- und Schätzverfahren für die relative Häufigkeit von Merkmalsträgern in Geschwisterreihen bei einem der Auslese unterworfenen Material mit Anwendung auf das Retinoblastom. Arch Julius-Klaus Stift 1955; 30: 263–285.
4 Berglin CG. On the concepts and calculus of penetrance. Acta Genet Stat Med 1955; 5: 240–262.
5 Berglin CG. Generalized formulae of the penetrance calculus. Acta Genet Stat Med 1957; 7: 66–67.
6 Berglin CG. Some penetrance formulae in recessive proband material. Acta Genet Med Gemellol 1957; 6: 451–458.
7 Trankell A. Penetrance calculus in population genetics. Amer J Hum Genet 1956; 8: 43–48.
8 Furusho T. Estimation of frequency and penetrance of genes responsible for incompletely penetrant, monosomic characters in man. Jap J Hum Genet 1960; 5: 11–19.
9 Sang JH. Penetrance, expressivity and thresholds. J Heredity 1963; 154: 143–151.
10 Murphy EA, Bolling DR. Testing of single locus hypothesis where there is incomplete separation of the phenotypes. Amer J Hum Genet 1967; 19: 322–334.
11 Goto G. Mathematical methods for human gene analysis with special reference to the irregular manifestation of genotypes. Kumamoto J Sci Series B 1967; 8: 91–118.
12 Murphy EA. Incomplete penetrance and the ascertainment of recessive characters in man. I. Introduction and distribution theory. J Chron Dis 1970; 23: 9–27.
13 Rao DC, Gorai JK. Penetrance of the tongue pigmentation allele. Jap J Hum Genet 1970; 15: 186–191.
14 Wilson SR. Fitting of models of incomplete penetrance to family data. Ann Hum Genet 1971; 35: 99–108.
15 Murphy EA, Chase GA. Principles of Genetic Counseling Year Book Medical Publishers, Chicago, 1975.
16 Suarez BK, Reich T, Trost J. Limits of the general two-allele single locus model with incomplete penetrance. Ann Hum Genet 1976; 40: 231–244.

https://doi.org/10.1375/twin.3.4.294 Published online by Cambridge University Press

17 Suarez BK, Fishman PM, Reich T. Estimating the parameters of incompletely penetrant single locus model using multiple populations. Hum Hered 1977; 27: 336–351.
18 Matsunaga E. On estimating penetrance of the retinoblastoma gene. Hum Genet 1980; 56: 127–128.
19 Cannings C, Thompson EA, Skolnick M. In: Mielke JH and Crawford MH. Current Developments in Anthropological Genetics. Plenum Press, New York, 1980.
20 Morton NE. Outline of Genetic Epidemiology. S Karger, Basel, 1982.
21 Rogatko A. Diallelic model combining selection and incomplete penetrance. J Theoret Biol 1983; 102: 1–6.
22 Rogatko A, Pereira CAB, Frota-Pessoa O. A Bayesian method for the estimation of penetrance: application to mandibulofacial and frontonasal dysostoses. Amer J Med Genet 1986; 24: 231–246.
23 Morton NE. The future of genetic epidemiology. Ann Med 1992; 24: 557–562.
24 Otto PA, Frota-Pessoa O, Polcan SP. Snyder's ratios with incomplete penetrance. J Heredity 1994; 85: 331–335.
25 Schinz HR. Konkordanz, Diskordanz und Penetranz bei eineiigen Zwillingen (Versuch einer elementaren Darstellung). Arch Julius-Klaus Stift 1945; 20: 2–25.
26 Lasker GW. Penetrance estimated by frequency of unilateral occurrences and by discordance in monozygotic twins. Hum Biol 1947; 19: 217–230.
27 Allen G. The meaning of concordance and discordance in estimation of penetrance and gene frequency. Amer J Hum Genet 1952; 4: 155–172.
28 Pfändler U. La signification statistique de la concordance chez les jumeaux MZ et DZ. Acta Genet Med Gemellol 1970; 19: 160–164.
29 Rife DC. Heredity and handedness. Sci Monthly 1951; 73: 188–191.
30 Laland KN, Kumm J, Van Horn JD, Feldman MW. A gene-culture model of human handedness. Behav Genet 1995; 25: 433–445.
31 Snyder LH. The inheritance of taste deficiency in man. Ohio J Sci 1932; 32: 436–440.
32 Snyder LH. Studies on human inheritance. X. A table to determine the proportion of recessives to be expected in various matings involving a unit character. Genetics 1934; 19: 1–17.
33 Matlock P. Identical twins discordant in tongue-rolling. J Heredity 1952; 43: 24.
34 Vogel F. Über die Fähigkeit, die Zunge um die Längaschse zu rollen. Acta Genet Med Gemellol 1957; 6: 225–230.
35 Reedy JJ, Szczes T, Downs TD. Tongue rolling among twins. J Heredity 1971; 62: 125–127.
36 Martin MG. No evidence for a genetic basis of tongue-rolling or hand clapping. J Heredity 1975; 66: 179–180.
37 Sturtevant AH. A new inherited character in man. Proc Nat Acad Sci 1940; 26: 100–102.

https://doi.org/10.1375/twin.3.4.294 Published online by Cambridge University Press