Analysis of family history data for evidence of non-Mendelian inheritance resulting from vertical transmission¹

PAUL E. M. FINE

From the Ross Institute, London School of Hygiene and Tropical Medicine, London

SUMMARY A number of infections of man, as well as of other animal and plant species, are heavily dependent upon 'vertical transmission'—that is direct parent-to-progeny transfer—for their maintenance in host populations. Such vertical transmission may be considered as a form of inheritance. It is usually non-Mendelian. This paper discusses the implications of such inheritance for the distribution of disease in families. A method is described for making quantitative predictions of prevalence rates of infection and of disease within different classes of relatives of either infected or uninfected probands. It is pointed out that, whereas a maternal line excess is to be expected among relatives of positive probands, the opposite should be found in families of negative probands. Expected differences between maternal and paternal line prevalence rates of observable disease decline rapidly with distance of relationship from the proband, and are greatly reduced by diagnostic insensitivity (analogous to penetrance). The implications of this analytic method for the design of family history studies are discussed. Published data on familial breast cancer are reviewed, and found to show no evidence that this condition is associated with a non-integrated vertically transmitted agent.

The phrase 'vertical transmission' was coined by Ludwik Gross (1944, 1949) to describe the direct transfer of infection from parent organisms to their immediate progeny. It refers to a highly useful concept, as such transmission is known to play a significant role in the epidemiology of many diseases of medical, veterinary, and agricultural importance (Baker and Smith, 1966; Elliott and Knight, 1973; Fine, 1975). In addition, such a mechanism has been suggested as a possible explanation for the familial clustering of a number of diseases the aetiology or epidemiology of which are still uncertain—human breast cancer (Gross, 1949), leukaemia (Gross, 1954), hepatitis B (Stevens *et al.*, 1975), Leber's optic atrophy (Wallace, 1970), Kuru (Gajdusek, 1963), and Creutzfeldt-Jakob disease (Ferber *et al.*, 1973) being among the better known examples.

Confirmation of the vertical transmission of infection is in some cases straightforward. This is in general true if good diagnostic tools are available for

¹Much of the work reported here was carried out while the author was in the Department of Epidemiology, School of Public Health, University of California, Berkeley, California 94720, USA, and supported by Research Grant AI03028 from the National Institute of Allergy and Infectious Diseases, and General Research Support Grant 5-501-RR-05441 from the National Institutes of Health.

Received for publication 23 March 1977

the recognition of infection in neonates, as by the isolation of an infectious agent or by the detection of IgM antibodies specific to such an agent. But such diagnostic techniques are not always available—in which case the demonstration of vertical transmission may require an epidemiological argument. Among the more promising of these approaches is that of the retrospective study of family history. It is the purpose of this paper to discuss those implications of vertical transmission that bear on such retrospective studies (Fine, 1976).

The basic problem—a qualitative solution

It is widely recognised that the vertical transmission of infection differs fundamentally from classical genetic inheritance. Among its distinctive features is its tendency to occur mainly through the female parent. Because the ovum provides some 99% of the extranuclear material of a zygote and because of the residence of the developing embryo within the female body, and because of the special role of the female parent in nursing and caring for the young, the transfer of infection is almost certain to be more efficient from an infected mother to her progeny than

from an infected father. (We here avoid the Pandora's box opened up by the recent discoveries of reverse transcriptase, and of the resultant Mendelising of some avian and murine oncogenic viruses (Weiss, 1975). This mechanism apparently permits some 'infectious' agents to masquerade as Mendelian alleles which, if autosomal, should be inherited with equal regularity from male or female parent. The ultimate solution of the epidemiology of such agents may well require a synthesis of classical population genetics with the sort of non-Mendelian analysis described here.)

The tendency for vertical transmission to be matroclinal has important implications for the framing and testing of epidemiological hypotheses. In the prospective sense, it has led to the use of reciprocal-cross analysis—comparison of the progeny produced by affected females paired with unaffected males, as against the progeny of unaffected females paired with affected males. A higher prevalence rate of the condition among progeny of the former cross may be an indication of extrachromosomal inheritance, and hence of vertical transmission. Experiments of this kind have played a major role in studies of infectious heredity, providing strong evidence for the vertical transmission of such agents as the mouse mammary carcinoma virus (Staff of Roscoe B. Jackson Memorial Laboratory 1933) and the CO₂ sensitivity virus of *Drosophila melanogaster* (l'Heritier and Teissier, 1938). Several variants of the prospective design have been described by Jinks (1964).

Prospective reciprocal-cross studies are difficult to carry out in man, for obvious social, ethical, and

temporal reasons. This difficulty has led to the formulation of a retrospective analogue of the argument which may be summarised as follows: if a factor or infectious agent is equally distributed among the males and females of a population, and if its hereditary transfer is more efficient from female parents to their progeny than from males, then there must be more mother-to-progeny transfers of infection taking place in the system than there are father-to-progeny transfers. From this it follows that any infected offspring is more likely to have become infected through the mother than through the father. This means that—given vertical transmission of some infectious agent—there should be a higher prevalence rate of infection among the mothers (and maternal relatives) of affected probands than among their fathers (and paternal relatives).

Penrose *et al.* (1948) were apparently the first to formulate this argument, in the context of their important study of hereditary influences and breast cancer. And the identical argument has been applied by several subsequent workers, also with reference to human breast cancer (Macklin, 1959; Henderson *et al.*, 1974). Other publications on this subject have presented data applicable to the hypothesis, but have not explicitly utilised or discussed the relation. Table 1 presents a summary of relevant data from a number of such studies. It will be noted that only in the data of Penrose *et al.* (1948) is there a suggestion of a significant excess of breast cancer in maternal as opposed to paternal relatives of affected probands. It is noteworthy that Penrose *et al.* attributed this apparent trend to ascertainment bias, and not to an actual preponderance in maternal relatives (see

Table 1 Summary of published data on incidence rates of breast cancer among female relatives of breast cancer patients (positive probands) and of controls (negative probands)

Author and relatives compared	Positive probands						Negative probands					
	Maternal relatives			Paternal relatives			Maternal relatives			Paternal relatives		
	No.	Cases	B _a rate	No.	Cases	B _a rate	No.	Cases	B _a rate	No.	Cases	B _a rate
Penrose <i>et al.</i> (1948)												
Grandmothers	np	9		np	2							
Aunts	np	28		np	16							
Macklin (1959)												
Grandmothers	163	10	0.061	134	7	0.052	np	np	np	np	np	np
Aunts	355	21	0.059	327	22	0.067	np	np	np	np	np	np
Henderson <i>et al.</i> (1974)												
Aunts	396	24	0.061	312	24	0.077	369	17	0.046	298	8	0.027
Jacobsen (1946)												
Grandmothers	183	4	0.022	157	4	0.025	172	0	0	162	2	0.012
Aunts	316	17	0.054	224	12	0.054	312	0	0	223	2	0.009
Murphy and Abbey (1959)												
Aunts	254	5	0.020	229	9	0.039	312	6	0.019	247	5	0.020

Maternal and paternal line relatives are shown separately. On the assumption of no recovery from 'infection' with a breast cancer 'agent', incidence rates would reflect (some proportion of the) ultimate prevalence rates (B_a) of 'infection' among the different groups of adult women. np indicates data which were apparently collected, but not published. It should be noted that these incidence figures represent crude rates, and have not been standardized for age.

J Med Genet: first published as 10.1136/jmg.14.6.399 on 1 December 1977. Downloaded from http://jmg.bmj.com/ on April 26, 2023 by guest. Protected by copyright.

Discussion section). More important, it should be stressed that Penrose *et al.* presented only numbers of cases among the aunts and grandmothers, whereas rates are essential for a valid comparison.

The weakness of this argument is its qualitative nature. It provides no indication of the *quantitative* difference to be expected between prevalence rates in different classes of relatives. And, without such quantitative estimates, it is difficult to assess the conclusiveness of the trends indicated in such data as in Table 1. We shall now, therefore, search for quantitative inferences from the basic assumption of vertical transmission.

General quantitative solution

The analytical method to be described here represents the retrospective analogue of a general vertical transmission model discussed in Fine (1975). The model, which was developed to describe the epidemiological implications of vertical transmission in any sexually reproducing species, is based upon the following assumptions and parameters.

ASSUMPTIONS

1. We assume that a factor or infectious agent has been present in the study population at constant prevalence rate for a period of time which spans the several generations included in our family history data.

2. We assume that the prevalence rate of infection is equal in males and females. (It is important to add that this assumption does *not* necessarily mean that males and females *manifest* the infection with equal frequency, but only that they *carry* the agent with equal frequency. Such a distinction may be relevant to arguments bearing upon the epidemiological pattern of such sex-specific conditions as mammary carcinoma.)

3. We assume that the host population is large and that males and females mate at random, at least in relation to the agent in question.

4. We assume that, once infected, either by vertical or by horizontal—that is, not parent-to-progeny—transmission, an individual remains infected for life.

5. We assume that our diagnostic test for this infection is highly *specific*, and that there are no false positives. (The problem of diagnostic sensitivity is discussed below.)

These assumptions resemble closely those which underlie the Hardy-Weinberg principle of population genetics. The analogy is a useful one, as the logical arguments for the two—that is, the Mendelian and the non-Mendelian—conditions are somewhat similar.

PARAMETERS

B_a = *Prevalence rate* of infection among parents in the study population.

d = *Maternal vertical transmission rate*—that is, the proportion infected among the progeny of infected females mated with uninfected males.

v = *Paternal vertical transmission rate*—that is, the proportion infected among the progeny of infected males mated with uninfected females.

β = *Relative survival rate*—that is, the average relative survival rate, from birth to reproductive age, of infected as compared with uninfected individuals.

α = *Relative fertility rate*—that is, the average relative number of progeny born to an infected individual as compared with an uninfected individual.

It may be noted that the two separate selection parameters, β and α , are required because rates of vertical transmission are assessed at birth, or at weaning, rather than at conception. The product $\beta\alpha$ is then roughly analogous to the fitness value used in Mendelian population genetics.

The assumptions and parameters defined above are general enough to cover the epidemiological pattern of many infectious agents in which vertical transmission plays a role. It can be shown that in certain circumstances—for example, $\alpha\beta(d+v) > 1.0$ —a vertical transmission mechanism may be competent in itself to maintain an infection within a host population (Fine, 1975). In many cases, however, vertical transmission alone is not sufficient to maintain the infection at constant prevalence rate over successive generations. Supplementary horizontal transfer must be involved as well. We describe this horizontal transmission contribution by the following incidence rate parameter:

h = *net horizontal incidence rate*—that is, the incidence rate of infection over the period from birth to reproductive age.

On the assumption that horizontally acquired infections cause at most a negligible mortality during the period before reproductive age, it is possible to define the extent of horizontal transmission required, as a supplement to the vertical transmission, in order to maintain the constant equilibrium prevalence rate B_a over successive generations. Using the symbol h_e

¹In an earlier publication (Fine, 1975), the author used the symbol r for this maternal vertical transmission rate. This was an unwise choice, because of the common association of this symbol with recovery rates and intrinsic rates of increase. The symbol d is thus used here for consistency with the notation of l'Heritier (1970). For a discussion of the estimation of these rates, see Fine and Sylvester (1978).

for this equilibrium horizontal-transmission rate, we have (see Fine, 1975);

$$h_e = 1 - \frac{\alpha\beta B_a(1 - B_a)\{d(1 - B_a + B_a\alpha) + v(1 - B_a + B_a\alpha - B_a\alpha d)\}}{(1 - B_a + B_a\alpha - B_a\alpha d)(1 - B_a + B_a\alpha - B_a\alpha v)} \quad (1)$$

In the present context, it is important to note that this equilibrium incidence rate of horizontally acquired infections is dependent upon all of the other five defined parameters.

The problem is now to manipulate these several parameters in order to derive estimates of prevalence rates for use in retrospective studies. Expressions for several different classes of relatives, of both infected and uninfected probands, are required. One is derived here, as an illustration of the underlying argument, and the rest are presented in a summary table (Table 2).

PREVALENCE RATE AMONG MOTHERS OF AFFECTED PROBANDS

We assume that the probands are adults, who have thus at least had an opportunity to become infected by horizontal as well as by vertical transmission. This is appropriate for most human family history studies—for example, on breast cancer, in which adult women are normally chosen as probands. (It would also be straightforward to derive expressions for neonatal probands.)

The prevalence rate of infection among mothers of infected (= 'positive') probands may be defined as follows: it is the probability that a positive proband

(+) was infected vertically ($\downarrow+$) and had an infected mother (m+) or that the positive proband was not infected vertically ($\downarrow-$) yet still had an infected mother (m+). This can be expressed by a probability equation, using conventional format for conditionals. If $B_a\{m+|+\}$ is the probability a proband's mother is positive, given that the proband is positive, and if $B_a\{\downarrow-|+\}$ is the probability that a given positive proband was not infected vertically, then the required equation is:

$$B_a\{m+|+\} = B_a\{\downarrow+|+\} \cdot B_a\{m+|\downarrow+\} + B_a\{\downarrow-|+\} \cdot B_a\{m+|\downarrow-\}. \quad (2)$$

Note that each of these conditionals expresses equally a probability or a proportion (prevalence rate).

We next express the right-hand side of (2) in terms of the defined parameters. We recognise that the product $B_a \cdot B_a\{\downarrow+|+\}$ is the prevalence rate of vertically acquired infections among adults in the population, and that $B_a - B_a \cdot B_a\{\downarrow+|+\}$ is the prevalence rate of 'non-vertical' infections. Hence at equilibrium, h_e , which is the incidence rate of non-vertical infections in the population initially at risk ($1 - B_a \cdot B_a\{\downarrow+|+\}$) is given by:

$$h_e = \frac{B_a - B_a \cdot B_a\{\downarrow+|+\}}{1 - B_a \cdot B_a\{\downarrow+|+\}}.$$

This, on rearrangement, gives:

$$B_a\{\downarrow+|+\} = \frac{B_a - h}{B_a(1 - h)}. \quad (3)$$

Table 2 *Theoretical prevalence rates of infection among relatives of observed infected + or observed uninfected - probands, on basis of general retrospective vertical transmission model. Final expressions are given in first portion of the table. Essential subsidiary equations are given in the second portion.*

Basic equations

Parents

$$\begin{aligned} B_a\{m+|+\} &= \Phi[B_a\{\downarrow+|+\}B_a\{m+|\downarrow+\} + B_a\{\downarrow-|+\}B_a\{m+|\downarrow-\}] \\ B_a\{f+|+\} &= \Phi[B_a\{\downarrow+|+\}B_a\{f+|\downarrow+\} + B_a\{\downarrow-|+\}B_a\{f+|\downarrow-\}] \\ B_a\{m+|-\} &= \Phi[B_a\{\downarrow+|-\}B_a\{m+|\downarrow+\} + B_a\{\downarrow-|-\}B_a\{m+|\downarrow-\}] \\ B_a\{f+|-\} &= \Phi[B_a\{\downarrow+|-\}B_a\{f+|\downarrow+\} + B_a\{\downarrow-|-\}B_a\{f+|\downarrow-\}] \end{aligned}$$

Sibs

$$\begin{aligned} B_a\{s+|s+\} &= \Phi\left[h + (1-h)\left(\frac{\beta B_a\{s\downarrow+|+\}}{\beta B_a\{S\downarrow+|+\} + (1-B_a\{s\downarrow+|+\})s}\right)\right] \\ B_a\{s+|-\} &= \Phi\left[h + (1-h)\left(\frac{\beta B_a\{s\downarrow+|-\}}{\beta B_a\{S\downarrow+|-\} + (1-B_a\{s\downarrow+|-\})s}\right)\right] \end{aligned}$$

Grandparents

$$\begin{aligned} B_a\{mgm+|+\} &= \Phi[B_a\{m+|+\}B_a\{m+|+\} + B_a\{m-|+\}B_a\{m+|-\}] \\ B_a\{pgm+|+\} &= \Phi[B_a\{f+|+\}B_a\{m+|+\} + B_a\{f-|+\}B_a\{m+|-\}] \\ B_a\{mgm+|-\} &= \Phi[B_a\{m+|-\}B_a\{m+|+\} + B_a\{m-|-\}B_a\{m+|-\}] \\ B_a\{pgm+|-\} &= \Phi[B_a\{f+|-\}B_a\{m+|+\} + B_a\{f-|-\}B_a\{m+|-\}] \end{aligned}$$

Aunts and uncles

$$\begin{aligned} B_a\{ms+|+\} &= \Phi[B_a\{m+|+\}B_a\{s+|+\} + B_a\{m-|+\}B_a\{s+|-\}] \\ B_a\{ps+|+\} &= \Phi[B_a\{f+|+\}B_a\{s+|+\} + B_a\{f-|+\}B_a\{s+|-\}] \\ B_a\{ms+|-\} &= \Phi[B_a\{m+|-\}B_a\{s+|+\} + B_a\{m-|-\}B_a\{s+|-\}] \\ B_a\{ps+|-\} &= \Phi[B_a\{f+|-\}B_a\{s+|+\} + B_a\{f-|-\}B_a\{s+|-\}] \end{aligned}$$

Table 2 (continued)

Subsidiary equations

$$\begin{aligned}
B_a\{+|- \} &= \frac{B_a(1-\Phi)}{1-B_a\Phi} & B_a\{-|- \} &= 1-B_a\{+|- \} \\
B_a\{\downarrow+|+ \} &= \frac{B_a-h}{B_a(1-h)} & B_a\{\downarrow-|+ \} &= 1-B_a\{\downarrow+|+ \} \\
B_a\{\downarrow+|- \} &= B_a\{+|- \}B_a\{\downarrow+|+ \} & B_a\{\downarrow-|- \} &= 1-B_a\{\downarrow+|- \} \\
B_a\{m+|+ \} &= \frac{d(1-B_a+B_a\alpha)+v(B_a\alpha-B_a\alpha d)}{d(1-B_a+B_a\alpha)+v(1-B_a+B_a\alpha-B_a\alpha d)} \\
B_a\{f+|+ \} &= \frac{v(1-B_a+B_a\alpha)+d(B_a\alpha-B_a\alpha v)}{d(1-B_a+B_a\alpha)+v(1-B_a+B_a\alpha-B_a\alpha d)} \\
B_a\{m+|\downarrow- \} &= \frac{B_a\alpha(1-d)}{1-B_a+B_a\alpha-B_a\alpha d} & B_a\{f+|\downarrow- \} &= \frac{B_a\alpha(1-v)}{1-B_a+B_a\alpha-B_a\alpha d} \\
B_a\{m+|+ \} &= B_a\{\downarrow+|+ \}B_a\{m+|\downarrow+ \} + B_a\{\downarrow-|+ \}B_a\{m+|\downarrow- \} \\
B_a\{f+|+ \} &= B_a\{\downarrow+|+ \}B_a\{f+|\downarrow+ \} + B_a\{\downarrow-|+ \}B_a\{f+|\downarrow- \} \\
B_a\{m+|- \} &= B_a\{m+|\downarrow- \} & B_a\{f+|- \} &= B_a\{f+|\downarrow- \} \\
B_a\{m+|- \} &= B_a\{\downarrow+|- \}B_a\{m+|\downarrow+ \} + B_a\{\downarrow-|- \}B_a\{m+|\downarrow- \} \\
B_a\{f+|- \} &= B_a\{\downarrow+|- \}B_a\{f+|\downarrow+ \} + B_a\{\downarrow-|- \}B_a\{f+|\downarrow- \} \\
B_a\{m+f+|+ \} &= B_a\{\downarrow+|+ \}B_a\{m+f+|\downarrow+ \} + B_a\{\downarrow-|+ \}B_a\{m+f+|\downarrow- \} \\
B_a\{m+f+|\downarrow- \} &= \frac{B_a\alpha d+v(B_a\alpha-B_a\alpha d)}{d(1-B_a+B_a\alpha)+v(1-B_a+B_a\alpha-B_a\alpha d)} \\
B_a\{m+f+|\downarrow- \} &= \frac{(B_a\alpha-B_a\alpha d)(B_a\alpha-B_a\alpha v)}{(1-B_a+B_a\alpha-B_a\alpha d)(1-B_a+B_a\alpha-B_a\alpha v)} \\
B_a\{m+f-|+ \} &= B_a\{\downarrow+|+ \}B_a\{m+f-|\downarrow+ \} + B_a\{\downarrow-|+ \}B_a\{m+f-|\downarrow- \} \\
B_a\{m+f-|\downarrow+ \} &= \frac{d(1-B_a)}{d(1-B_a+B_a\alpha)+v(1-B_a+B_a\alpha-B_a\alpha d)} \\
B_a\{m+f-|\downarrow- \} &= \frac{B_a\alpha(1-B_a)(1-d)}{(1-B_a+B_a\alpha-B_a\alpha d)(1-B_a+B_a\alpha-B_a\alpha v)} \\
B_a\{m-f+|+ \} &= B_a\{\downarrow+|+ \}B_a\{m-f+|\downarrow+ \} + B_a\{\downarrow-|+ \}B_a\{m-f+|\downarrow- \} \\
B_a\{m-f+|\downarrow+ \} &= \frac{v(1-B_a)}{d(1-B_a+B_a\alpha)+v(1-B_a+B_a\alpha-B_a\alpha d)} \\
B_a\{m-f+|\downarrow- \} &= \frac{B_a\alpha(1-B_a)(1-v)}{(1-B_a+B_a\alpha-B_a\alpha d)(1-B_a+B_a\alpha-B_a\alpha v)} \\
B_a\{m+f+|- \} &= B_a\{\downarrow+|- \}B_a\{m+f+|\downarrow+ \} + B_a\{\downarrow-|- \}B_a\{m+f+|\downarrow- \} \\
B_a\{m+f-|- \} &= B_a\{\downarrow+|- \}B_a\{m+f-|\downarrow+ \} + B_a\{\downarrow-|- \}B_a\{m+f-|\downarrow- \} \\
B_a\{m-f+|- \} &= B_a\{\downarrow+|- \}B_a\{m-f+|\downarrow+ \} + B_a\{\downarrow-|- \}B_a\{m-f+|\downarrow- \} \\
B_a\{o\downarrow+|m+f- \} &= d \\
B_a\{o\downarrow+|m-f+ \} &= v \\
B_a\{o\downarrow+|m+f+ \} &= d+v-dv \\
B_a\{s+|+ \} &= h+(1-h)\left(\frac{\beta B_a\{s\downarrow+|+ \}}{\beta B_a\{s\downarrow+|+ \}+1-B_a\{s\downarrow+|+ \}}\right) \\
B_a\{s+|- \} &= h+(1-h)\left(\frac{\beta B_a\{s\downarrow+|- \}}{\beta B_a\{s\downarrow+|- \}+1-B_a\{s\downarrow+|- \}}\right) \\
B_a\{s\downarrow+|+ \} &= B_a\{m+f+|+ \}B_a\{o\downarrow+|m+f+ \} + B_a\{m+f+|+ \}B_a\{o\downarrow+|m+f- \} + B_a\{m-f+|+ \}B_a\{o\downarrow+|m-f+ \} \\
B_a\{s\downarrow+|- \} &= B_a\{m+f+|- \}B_a\{o\downarrow+|m+f+ \} + B_a\{m+f+|- \}B_a\{o\downarrow+|m+f- \} + B_a\{m-f+|- \}B_a\{o\downarrow+|m-f+ \} \\
B_a\{s\downarrow+|- \} &= B_a\{m+f+|- \}B_a\{o\downarrow+|m+f+ \} + B_a\{m+f+|- \}B_a\{o\downarrow+|m+f- \} + B_a\{m-f+|- \}B_a\{o\downarrow+|m-f+ \}
\end{aligned}$$

Key:

$d, v, \alpha, \beta, B_a, \Phi$ Parameters as defined in text.

$B_a(y|x)$ = Probability of true infection state y , given true state x .

$B_a(y|x)$ = Probability of observing infection state y , given an observed infection state x .

x and y symbols as follows: +, infected; -, not infected; $\downarrow+$, infected vertically (i.e. via parent); $\downarrow-$, not infected vertically; $m+$, mother infected; $m-$, mother not infected; $f+$, father infected; $f-$, father not infected; $m+f+$, mother and father infected; $m+f-$, mother but not father infected; $m-f+$, not mother but father infected; $o\downarrow+$, offspring (son or daughter) vertically infected; $s+$, sib infected; $s\downarrow+$, sib infected vertically (via parent); $mgm+$, maternal grandmother infected; $pgm+$, paternal grandmother infected; $ms+$, maternal sib (aunt or uncle) infected; $ps+$, paternal sib (aunt or uncle) infected.

As $\{\downarrow+|+ \}$ and $\{\downarrow-|+ \}$ are complementary events, we also have:

$$B_a\{\downarrow-|+ \} = 1 - B_a\{\downarrow+|+ \}. \quad (4)$$

Completion of the right-hand side of equation (2) now requires expressions for $B_a\{m+|\downarrow+ \}$ and $B_a\{m+|\downarrow- \}$. These expressions are easily derived by means of a diagrammatic technique used in population genetics, and shown here in Fig. 1 (see also Fine, 1975). The figure comprises a square Venn

diagram, structured to show the distribution of different parent types among the offspring (the central area of the square). The square's vertical edge is measured off to show the relative proportions of progeny from infected and from uninfected females—that is, $B_a\alpha$ and $(1-B_a)$, respectively. As only a proportion d of the progeny of infected females actually receive the infection from their mothers, then a proportion $B_a\alpha d/(1-B_a+B_a\alpha)$ of all progeny receive the infection from their mothers.

Fig. 1 Dimensioned Venn diagram square, illustrating the relation between infection rates among parents and their infant progeny (newborn or, if milk transfer is in question, newly weaned). The prevalence rate of infection among both males and females of the parental generation is B_a . The shaded area in the diagram corresponds to progeny which receive the infection from at least one parent (see Fine (1975)).

This proportion is reflected across the entire square by a dotted horizontal line. Precisely the same argument applies to the fatherhood, as shown along the horizontal side of the square, and with v substituted for d . Of the resultant offspring, represented by the area within the square, the proportions receiving the infection from their mothers and/or fathers are represented by diagonal hatching. This diagram shows that $B_a\{m+|\downarrow+\}$, or the probability that a vertically infected offspring had an infected mother—regardless of which parent actually transmitted the offspring's infection—can be expressed as the proportion of all infected progeny—that is $\{B_a\alpha d(1-B_a+B_a\alpha)+B_a\alpha v(1-B_a+B_a\alpha-B_a\alpha d)\}$, which have positive mothers, or:

$$B_a\{m+|\downarrow+\} = \frac{d(1-B_a+B_a\alpha)+v(B_a\alpha-B_a\alpha d)}{d(1-B_a+B_a\alpha)+v(1-B_a+B_a\alpha-B_a\alpha d)} \quad (5)$$

Equation (2) also requires $B_a\{m+|\downarrow-\}$, i.e. the probability that an uninfected offspring had an infected mother. This too can be read directly from the diagram, and simplifies to:

$$B_a\{m+|\downarrow-\} = \frac{B_a\alpha(1-d)}{1-B_a+B_a\alpha-B_a\alpha d} \quad (6)$$

Expressions (3), (4), (5), and (6) together make up equation (2), thereby providing a calculable estimate

of the prevalence rate of infection among mothers of positive probands.

Finally, we recognise that diagnostic insensitivity will play a role in analyses such as this, and that it will be essential to distinguish between *truly* infected individuals (the biological reality) and *observed* positive individuals (the clinical reality). We thus define:

ϕ = Diagnostic sensitivity—that is, the proportion of true positives (infected individuals) actually recognised as such. (This epidemiological sensitivity concept is almost identical the genetical concept of *penetrance*—defined as the proportion of carrier individuals who actually manifest some trait.)

As far as expression (2) is concerned, we note that only a proportion ϕ of the true positive mothers of infected probands will be recognised as such. Therefore, introducing bold face symbols for observed status, we have the following expression for the proportion observed positive among the mothers of observed positive probands:

$$B_a\{m+|\downarrow+\} = \phi[B_a\{\downarrow+|\downarrow+\} \cdot B_a\{m+|\downarrow+\} + B_a\{\downarrow-|\downarrow+\} \cdot B_a\{m+|\downarrow-\}].$$

Analogous arguments can be formulated for the proportion observed to be positive among any class of relatives of either positive or negative probands. A set of these expressions is presented in Table 2.

Implications of model

Because of the complexity of the expressions presented in Table 2, the model's implications are more easily sought by simulation than by analysis. Though such a method may lack elegance, it does reveal a number of relations of epidemiological significance.

The standard retrospective method for testing a vertical transmission hypothesis has been to compare prevalence rates of infection in maternal versus paternal line relatives of infected probands (Table 1). A set of such differences has been predicted on the basis of the expressions in Table 2, and is illustrated in Fig. 2. We here take for illustrative purposes an overall prevalence rate of infection among adults of $B_a = 0.2$, and assume that there is no selective effect attributable to the infection ($\alpha = \beta = 1.0$). Infected mothers are taken to transmit the agent to 40% of their progeny ($d = 0.4$), and infected fathers to 10% of theirs ($v = 0.1$). The net equilibrium incidence rate of horizontally acquired infections, calculated by equation (1), comes to $h_e = 0.112$. In this initial analysis, we take a diagnostic sensitivity of 100%, or $\phi = 1.0$.

Two striking relations are evident in Fig. 2. The

Fig. 2 Difference between maternal and paternal line prevalence rates, as predicted by the general retrospective vertical transmission model. Differences illustrated here are based upon expressions in Table 2, with $d = 0.4$, $v = 0.1$, $\alpha = \beta = 1.0$, $B_a = 0.2$ and $\Phi = 1.0$.

first is that the predicted difference in prevalence rate between maternal and paternal line relatives decreases rapidly with increasing ‘distance’ from the proband. This should not come as a surprise, but it has important implications for this sort of epidemiological study. Investigation of the closest possible relatives to the proband (parents > grandparents > aunts/uncles) will optimise the chances of disclosing a maternal *versus* paternal line difference. As far as a sex-specific condition such as breast cancer is concerned, it would be more efficient—if difficulties in data collection were equal—to investigate grandmothers than aunts, for example.

The second implication evident in Fig. 2 is even more striking. Whereas a maternal line excess is predicted for relatives of positive probands, a paternal line excess is predicted for relatives of negative probands. This result is fully robust, as long as $d > v$ (the opposite trend would be expected if the paternal were greater than the maternal vertical transmission rate).

This result is intuitively reasonable, and can be deduced directly from the initial qualitative argument presented above. It does not appear to have been explicitly stated in the past. And its implication is almost paradoxical—that, in certain restricted circumstances, it might be reasonable to test a vertical transmission hypothesis by investigating only the relatives of negative, rather than of positive, probands. This might be so if it were possible to collect family history data on negative individuals more easily, and with less bias, than for infected probands.

The implication of diagnostic insensitivity for such studies is illustrated in Fig. 3, which shows the same relations as in Fig. 2, but with the additional variable of penetrance along the horizontal axis. The relative ‘efficiency’ of investigating family histories of negative

Fig. 3 Diagram illustrating the effect of diagnostic insensitivity ($\Phi < 1.0$) upon the observed difference between maternal and paternal line prevalence rates, as predicted by the general retrospective vertical transmission model. The d , v , α , β , and B_a parameters are the same as in Fig. 2.

probands decreases rapidly if the diagnostic test used is insensitive.

Predictions similar to those in Figs. 2 and 3 can be made by substituting any desired set of d , v , α , β , B_a and Φ values into the expressions in Table 2. Furthermore, given an estimate of prevalence rate excess or deficit in some class of relatives, it is then possible to predict the sample sizes that would be required in order to test the vertical transmission hypothesis. A set of such sample size predictions is presented in Table 3. The advantage in collecting data on grandparents—rather than on aunts and uncles—is immediately clear. In addition, it is apparent that the greater the difference between maternal and paternal vertical transmission rates $|d - v|$, the easier it should be to demonstrate the asymmetry in such retrospective studies.

Table 3 Sample sizes required to test vertical transmission hypothesis

Conditions	$d = 0.4$ $\alpha = \beta = 1$		$v = 0.1$ $B_a = 0.2$		$d = 0.8$ $\alpha = \beta = 1$		$v = 0.05$ $B_a = 0.2$	
	0.1	0.5	1.0	0.1	0.5	1.0	0.1	0.5
Positive probands								
Parents	1351	226	84	183	36	9		
Grandmothers	7921	1415	600	373	58	17		
Aunts/uncles	51099	9165	4010	562	91	32		
Negative probands								
Parents	> 10 ⁶	10154	801	226074	1254	83		
Grandmothers	> 10 ⁶	89331	7575	404975	2277	166		
Aunts/uncles	> 10 ⁶	> 10 ⁶	58273	672825	4006	312		

The number of relatives of each class required in order to be 90% sure of detecting a significant ($P < 0.05$) difference between maternal and paternal line prevalence rates. Sample size estimates are based upon using a one-tailed test for the difference between proportions—see equation (B.13.1) in Snedecor and Cochran (1967).

Discussion

The general model described here may prove useful in studies of a wide variety of conditions and diseases whose aetiologies have a non-Mendelian heritable component. Its qualitative conclusions, that maternal minus paternal line prevalence rate differences should decrease with increasing distance from the proband, and that they should be positive among relatives of infected probands, yet negative among relatives of uninfected controls, provide several initial inferences against which to screen family history data. The prior estimation of the magnitude of these differences in prevalence rates, and hence of sample sizes required to show such differences, requires the stipulation of numerical values for the several parameters α , β , d , v , B_a and Φ . This may pose severe difficulties, especially when the concern is with a condition whose epidemiology is not understood. None the less, the substitution of even very rough estimates into the expressions in Table 2 may be helpful in providing at least a range estimate of prevalence rates and prevalence rate differences which are to be expected. Such an exercise may be helpful in the initial planning stages of a family history study.

The theoretical basis of the model deserves close scrutiny. Like any model, its assumptions entail a considerable simplification of reality. Perhaps the most crucial of the assumptions is that of random mating within the population, regardless of infection status. It should be noted that, if an infectious agent readily undergoes horizontal transmission between adults, there may well be transfer between mating pairs, which will lead to a non-random, highly aggregated, distribution of infection among parental pairs. Such a possibility must be recognised, and, if necessary, could be allowed for in a further development of the model.

The model's derivation takes no account of several biases known to plague retrospective family studies—specifically those of proband ascertainment and of recall. It is assumed in the analysis that the probands are a random selection from either infected or uninfected segments of the population. And the sensitivity parameter will not compensate for bias caused by either 'asymmetric ignorance' or the problems introduced by relying on a proband's memory for information about relatives. This latter bias is of critical importance when testing for non-Mendelian inheritance patterns, as it is known that women are typically better acquainted with their family histories, particularly in the maternal line, than are men (Murphy and Abbey, 1959). Such differential recall could lead to a spurious excess of cases among maternal line relatives, thereby mimicking a vertical transmission pattern. Unfortunately,

many of the studies in the literature have relied heavily upon such information in their data collection; and this may well be responsible for some instances of apparent maternal line excess, as in breast cancer studies. Penrose *et al.* (1948) recognised this difficulty in interpreting their data on aunts and grandmothers (see Table 1), and were therefore hesitant to draw conclusions from the apparent maternal line clustering.

It may be noted that this development of the family history model fails to reveal evidence for vertical transmission in the breast cancer data summarised in Table 1. In particular, there is no evidence for paternal line excess among female relatives of negative probands.

The implications of diagnostic sensitivity are clearly of major importance for such studies. The current analysis assumes that the diagnostic sensitivity will be the same for each class of relatives. Of course, this may well be untrue, especially as there have been changes in disease classification and diagnostic tests over time. If such changes in diagnostic ability could be specified, then it would be possible to introduce different Φ parameters into the model for different classes of relatives. Diagnostic insensitivity has a greater effect upon prevalence rate differences in families of observed negative than of observed positive probands. This is because such insensitivity will mean that some of the observed negative probands are, in fact, misclassified. But, as the model presented here supposes a high specificity of diagnosis, there is assumed to be no misclassification among observed positives. It would be possible to introduce a further term for false positives as well.

The relation of this model to Mendelian inheritance patterns should also be mentioned. Autosomal alleles and traits should be equally distributed among maternal and paternal line relatives, and should pose no problem in such analyses. They provide a logical explanation for symmetrical family distribution. But some sex-linked factors are another matter—as their asymmetrical familial distribution might be misinterpreted as evidence for vertical transmission. It should be recognised that the assumptions of the model described here do not fit linkage. Assumption (2) specifies equal distribution of the factor between males and females, which is true of neither X nor the Y chromosome. And the model supposes a single maternal vertical transmission rate, d , whereas the analogous rate for an X-linked allele would be either 0.5 or 1.0, dependent upon whether the female parent were heterozygous or homozygous. For these reasons, the model does not purport to describe the condition of Mendelian sex-linked inheritance. Of course, patterns of sex-linked inheritance are well known, and provide additional predictions against which to test family history data.

Each of the prevalence rate predictions requires either five or six parameters. Therefore, when estimates of the prevalence rates of infection for six different classes of relatives are available, it is theoretically possible—given the appropriateness of the model—to calculate back and derive estimates of the implicit α , β , d , v , B_0 and Φ parameters. An exploration of this approach, and of the model's relation to sex-linked inheritance patterns, is intended for a subsequent publication.

In conclusion, we may recognise the wide scope of this analytical technique. Its relevance to epidemiological studies of familial neoplastic diseases in man is obvious. This is perhaps especially true now that we have become aware of the variety and extent of vertically transmitted viral material in many species including our own (Chandra *et al.*, 1970; Kalter *et al.*, 1973, 1975; Rongey *et al.*, 1973). The literature on the epidemiological implications of vertical transmission is growing rapidly, not only with reference to oncornaviruses, but also to such agents as cytomegalovirus (for example, Namazaki *et al.*, 1970). HB_sAg (Stevens *et al.*, 1975), and several putative slow viruses (Ferber *et al.*, 1973; Wallace, 1970). It is possible that retrospective family history analyses may provide a useful tool for assessing the epidemiological significance of the vertical transfer of such infections.

The encouragement and criticism of Dr James Renwick, on several aspects of this work, is gratefully appreciated.

References

- Baker, K. F., and Smith, S. H. (1966). Dynamics of seed infection of plant pathogens. *Annual Review of Phytopathology*, **4**, 311-334.
- Chandra, S., Liszczak, T., Korol, W., and Jensen, E. M. (1970). Type C particles in human tissues. I. Electron microscopic study of embryonic tissues in vivo and in vitro. *International Journal of Cancer*, **6**, 40-45.
- Elliott, K., and Knight, J. (1973). (Eds.) *Intrauterine Infections. Ciba Foundation Symposium*, 10 (n.s.). Elsevier-North Holland, Amsterdam, London, and New York.
- Ferber, R. A., Wiesenfeld, S. L., Roos, R. P., Bobowick, A. R., Gibbs, C. J., and Gajdusek, D. C. (1973). Familial Creutzfeldt-Jacob disease: transmission of the familial disease to primates. In *Proceedings of the 10th International Congress Neurology, Barcelona*. Excerpta Medica Foundation, Amsterdam.
- Fine, P. E. M. (1975). Vectors and vertical transmission: an epidemiologic perspective. *Annals of the New York Academy of Sciences*, **266**, 173-194.
- Fine, P. E. M. (1976). Non-Mendelian inheritance due to vertical transmission: a method for the analysis of family history data (abstract). *American Journal of Epidemiology*, **104**, 329.
- Fine, P. E. M., and Sylvester, E. S. (1978). Calculation of vertical transmission rates of infection, illustrated with data on an aphid borne virus. *American Naturalist*. In the press.
- Gajdusek, D. C. (1963). Kuru. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **57**, 151-169.
- Gross, L. (1944). Is cancer a communicable disease? *Cancer Research*, **4**, 293-303.
- Gross, L. (1949). The 'vertical epidemic' of mammary carcinoma in mice. Its possible implications for the problem of cancer in general. *Surgery, Gynecology and Obstetrics*, **88**, 295-308.
- Gross, L. (1954). *Transmissible Mouse Leukaemia: Biological Properties of the Mouse Leukaemia Agent. Ciba Foundation Symposium on Leukaemia Research*. J. and A. Churchill, London.
- Henderson, B. F., Powell, D., Rosario, I., Keys, C., Hanisch, R., Young, M., Casagrande, J., Gerkins, V., and Pike, M. C. (1974). An epidemiologic study of breast cancer. *Journal of the National Cancer Institute*, **53**, 609-614.
- Jacobsen, O. (1946). *Heredity in Breast Cancer: A Genetic and Clinical Study of Two Hundred Proband*s. H. K. Lewis, London.
- Jinks, J. L. (1964). *Extrachromosomal Inheritance*. Prentice-Hall, New Jersey.
- Kalter, S. S., Heberling, R. L., Smith, G. C., Panigel, M., Kraemer, D. C., Helmke, R. J., and Hellman, A. (1975). Vertical transmission of C-type viruses: their presence in baboon follicular oocytes and tubal ova. *Journal of the National Cancer Institute*, **54**, 1173-1176.
- Kalter, S. S., Helmke, R. J., Heberling, R. L., Panigel, M., Fowler, A. K., Strickland, J. E., and Hellman, A. (1973). C-type particles in normal human placentas. *Journal of the National Cancer Institute*, **50**, 1081-1084.
- l'Heritier, P. (1970). *Drosophila* viruses and their role as evolutionary factors. *Evolutionary Biology*, **4**, 185-209.
- l'Heritier, P., and Teissier, G. (1938). Un mécanisme héréditaire aberrant chez la drosophile. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, **206**, 1193-1195.
- Macklin, M. T. (1959). Comparison of the number of breast-cancer deaths observed in relatives of breast-cancer patients, and the number expected on the basis of mortality rates. *Journal of the National Cancer Institute*, **22**, 927-951.
- Murphy, D. P., and Abbey, H. (1959). *Cancer in Families: A Study of the Relatives of 200 Breast Cancer Proband*s. Harvard University Press, Cambridge.
- Namazaki, Y., Yano, N., Murizuka, T., Taka, S., and Ishida, N. (1970). Primary infection with human cytomegalovirus: virus isolation from healthy infants and pregnant women. *American Journal of Epidemiology*, **91**, 410-417.
- Penrose, L. S., Mackenzie, H. J., and Karn, M. N. (1948). A genetical study of human mammary cancer. *Annals of Eugenics*, **14**, 234-266.
- Rongey, R. W., Hlavackova, A., Lara, S., and Gardner, M. B. (1973). Types B and C RNA virus in breast tissue and milk of wild mice. *Journal of the National Cancer Institute*, **50**, 1581-1589.
- Snedecor, G. W., and Cochran, W. G. (1967). *Statistical Methods*, 6th ed. Iowa State University Press, Ames, Iowa.
- Staff of Roscoe B. Jackson Memorial Laboratory (1933). The existence of non-chromosomal influence in the incidence of mammary tumors in mice. *Science*, **78**, 465-466.
- Stevens, C. E., Beasley, R. P., Tsui, J., and Lee, W. C. (1975). Vertical transmission of hepatitis B antigen in Taiwan. *New England Journal of Medicine*, **292**, 771-774.
- Wallace, D. C. (1970). Leber's optic atrophy: a possible example of vertical transmission of a slow virus in man. *Australasian Annals of Medicine*, **3**, 259-262.
- Weiss, R. A. (1975). Genetic transmission of RNA tumor viruses. *Perspectives in Virology*, **9**, 165-205.

Requests for reprints to Dr Paul E. M. Fine, Ross Institute of Tropical Hygiene, London School of Hygiene and Tropical Medicine, Keppel Street (Gower Street), London WC1E 7HT.