Segregation of the Ellis-van Creveld Syndrome as Analyzed by the First Appearance Method

C. C. LI

Graduate School of Public Health, University of Pittsburgh.

DWARFISM IN MAN is usually a genetic dominant trait. The classical achondroplasia is a familiar example of dominance. There are, however, several different forms of dwarfism, some of which may be distinguished by clinical studies. One recessive form of dwarfism, the Ellis-van Creveld syndrome (EvC), in the Amish of Lancaster County, Pennsylvania, has been studied by McKusick *et al.* (1964). The features of EvC are well established: short-limbed, disproportionate dwarfism, polydactyly, and dysplasia of finger nails. For detailed description and photographic illustration, see McKusick, *et al.* (1964).

The recessive nature of the EvC syndrome is supported by 29 sibships of various sizes, each from two normal parents but with at least one affected member. The father of an additional family was affected, and this family is excluded from the following analysis. The purposes of the present communication are (1) to estimate the segregation ratio of the syndrome by a new and very simple method and (2) to compare its efficiency relative to the maximum likelihood method. Ascertainment is probably as complete as it can be practically. McKusick, *et al.* state: "An attempt was made to identify all cases of dwarfism in the Old Order Amish. Correspondence with acquaintances in many Amish communities, inquiries to over 500 physicians practicing in Amish areas, and information from Amish dwarfs and their families about other Amish dwarfs were the main approaches."

METHOD OF FIRST APPEARANCE TIME

In a sibship of any given size s, the first affected child may be the first ($t = 1$), second ($t = 2$), . . ., or the last ($t = s$) child of the family. Thus, t designates the position of the first affected child in the sequence of births in a family. We shall call t the first appearance time of an affected child. (If we wish, t may also be called the waiting time for the first affected child to appear.)

Two of the 29 sibships (McKusick *et al.*, 1964, Fig. 12, p. 323) have a single affected child and yield no information on segregation ratio. The remaining 27 sibships have been rearranged according to the first appearance time and according to sibship size in Fig. 1 and summarized in Table 1. The total number of sibships with first appearance time $t = 1$ (i.e., the sum of the first column of Table 1) is $A = 9$. The total number of sibships with the first appearance on the last child ($t = s$) is $B = 3$. This happened in pedi-

---

Received January 29, 1965.

Table 1. Classification of Sibships by Size and by First Appearance Time

Size of sibship s	First appearance time, t							Number of sibships
	1	2	3	4	5	6	9	
2	1							1
3	3		1					4
4	2	1	2	1				6
6	1	1	1			1		4
7		1			1			2
8	2		1		1			4
9				2				2
10				1				1
11			1					1
12			1					1
14							1	1
Total	$A = 9$,		$C = 15$,		$B = 3$			$N = 27$

grees 23, 2, and 6 in Fig. 1 and is recorded as the diagonal entries in Table 1. There are then $C = 15$ sibships with $1 < t < s$, while the grand total number of sibships is $N = 27$.

The method of estimating the segregation ratio by first appearance time involves so little arithmetic that as soon as the pedigrees are classified and tabulated, the work is done. Let p_1 be the estimate of p, the probability of a child being affected from two normal parents and $q = 1 - p$. Application of the method proposed by Li (1964) to the present set of data yields:

$$p_1 = \frac{A - B}{N - B} = \frac{9 - 3}{27 - 3} = \frac{6}{24} = .2500$$

with variance

$$V(p_1) = V_1 = q_1^2 \left(\frac{A + B}{(N - A)(N - B)} \right)$$

$$= \frac{p_1 q_1}{N - B} \left(\frac{A + B}{A - B} \right) = \frac{1}{64} = .015625$$

The standard error is $1/8 = .125$. It is seen that the estimate is of the right magnitude but the standard error is entirely too large for practical usage. This is, however, only the first stage of estimation. The second step is as follows.

SIMPLE COUNTING AFTER FIRST APPEARANCE

In classifying the sibships according to first appearance time (Fig. 1), we have ignored the phenotype of the children after the first affected child. As pointed out by Li (1964), the number of affected children after the first appearance should have a complete binomial distribution of degree $s - t$, and thus the segregation ratio may be determined by simple counting. Now, from Fig. 1, omitting the first affected member and those preceding him from

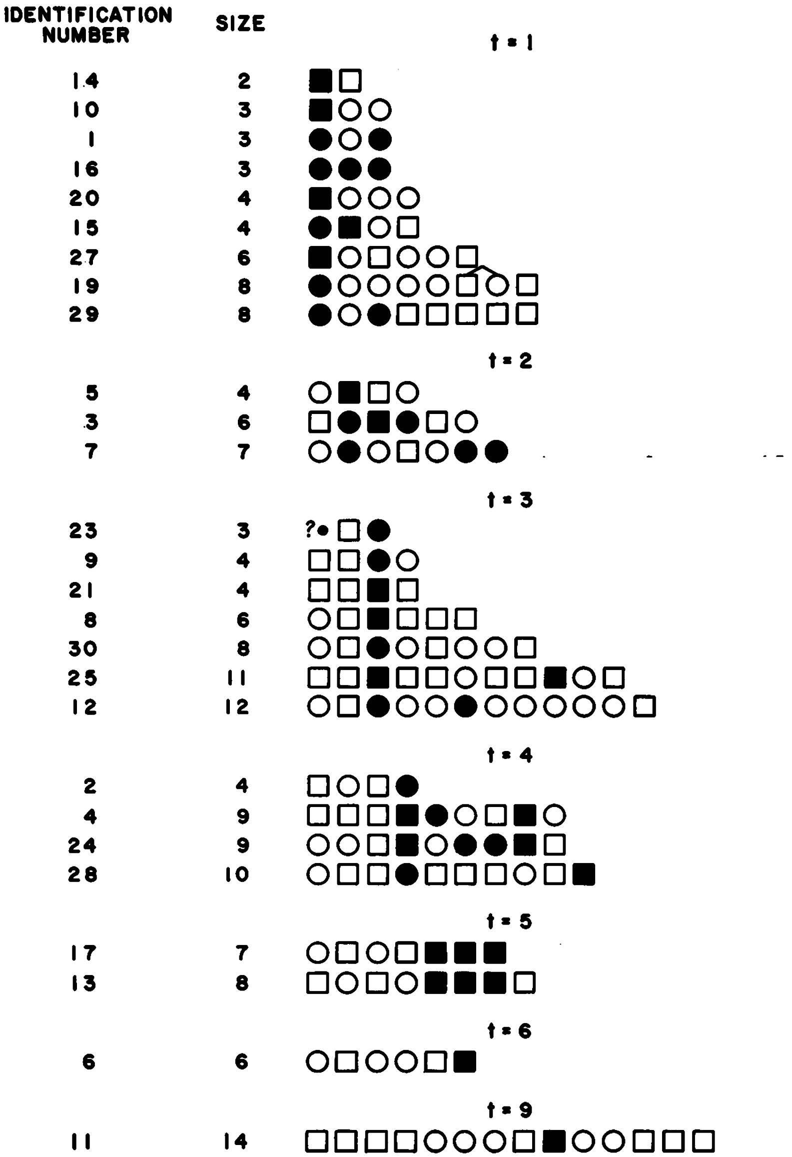

Fig. 1. Classification of 27 sibships by first appearance time. (Rearranged from McKusick *et al.*, 1964, Fig. 12, p. 383.)

each sibship, we count a total of $G = 96$ children, $R = 21$ of whom are affected. The estimate is then

$$p_o = \frac{R}{G} = \frac{21}{96} = .21875$$

$$V(p_o) = V_o = \frac{p_o q_o}{G} = \frac{(.21875)\,(.78125)}{96} = .001780$$

TABLE 2. ESTIMATING THE SEGREGATION RATIO BY LINEAR INTERPOLATION OF MAXIMUM LIKELIHOOD ESTIMATES

Size of sibship s	Number of sibships n_s	Number affected r_s	$n_s\bar{r}_s$		$n_s w_s$	
			$p = .225$	$p = .200$	$p = .225$	$p = .200$
2	1	1	1.13	1.11	3.64	3.86
3	4	7	5.05	4.92	31.10	32.75
4	6	7	8.45	8.13	74.20	77.80
6	4	6	6.89	6.50	91.04	94.96
7	2	6	3.79	3.54	56.92	59.30
8	4	7	8.28	7.69	137.66	143.42
9	2	7	4.50	4.16	81.16	84.62
10	1	2	2.44	2.24	46.90	48.96
11	1	2	2.63	2.41	53.32	55.77
12	1	2	2.83	2.58	59.81	62.71
14	1	1	3.24	2.93	72.87	76.79
Total	27	48	49.23	46.21	708.62	740.94
Interpolated value			$\hat{p} = .215$		$W = 721.55$	

This estimate, based on the children after the first affected, is independent of the previous estimate based on first appearance time only. Hence, a weighted average of the two estimates may be obtained, using the reciprocal of variance as weight:

$$w_1 = 1/V_1 = 64.00, \qquad w_o = 1/V_o = 561.74$$

The average estimate is

$$\bar{p} = \frac{w_1 p_1 + w_o p_o}{w_1 + w_o} = \frac{138.88}{625.74} = .222$$

$$V(\bar{p}) = \frac{1}{w_1 + w_o} = \frac{1}{625.74} = .001598$$

The standard error is .040 approximately. The first estimate p_1 is based on children up to and including the first affected, and the second estimate p_0 is based on children after the first affected, so that the phenotype of every child has been utilized in obtaining the average or pooled estimate $\bar{p}$.

MAXIMUM LIKELIHOOD ESTIMATE

A quantitative statement as to the efficiency of the estimate $\bar{p}$ relative to the maximum likelihood estimate is difficult except that the former is less efficient than the latter. It is thus of interest to compare empirically the efficiency for this particular set of data. The basic maximum likelihood method is given by Haldane (1932, 1938), and we shall adopt Lejeune's (1958) suggestion of linear interpolation. For an exposition of the procedure, the reader may refer to Li (1961, Chap. 5).

Table 2 gives the necessary computation for obtaining the maximum likelihood estimate of p. The first three columns give a summary of the raw data.

The fourth column shows the value of $n_s r_s = n_s \times sp/(1 - q^s)$ for the initial trial value $p = .225$. The last factor, $sp/(1 - q^s)$, has been tabulated by Li (1961, p. 66). The total of this column, 49.23, is the expected total number of affected children. It turns out to be slightly larger than the observed 48, so the next step is to try a smaller value of p. A similar calculation for $p = .200$ yields 46.21 which is smaller than 48. By linear interpolation we obtain the maximum likelihood estimate $\hat{p} = .215$.

The last two columns of Table 2 give the value of $n_s w_s$ where $w_s = 1/V(p)_s$, the weight for a sibship of size s. The value of w_s has also been tabulated by Li (1961, p. 66). The total weights are found to be 708.62 for $p = .225$ and 740.94 for $p = .200$. The total weight for $\hat{p} = .215$ is again obtained by simple interpolation, turning out to be 721.55. The variance of the estimate is $V(\hat{p}) = 1/721.55 = .001386$, and the standard error is .037 approximately.

The efficiency of the first appearance method relative to the maximum likelihood method is

$$R.E. = \frac{V(\hat{p})}{V(\bar{p})} = \frac{W(\bar{p})}{W(\hat{p})} = \frac{625.74}{721.55} = 86.7\%$$

The efficiency is higher than the author originally expected, considering the extreme simplicity of the first appearance method, which involves merely a new arrangement of the pedigrees and counting.

DISCUSSION

It is interesting to note that our first appearance estimate ($\bar{p} = .222$) is larger than the maximum likelihood estimate ($\hat{p} = .215$). Whether this is generally true or not must await further studies.

The ascertainment, though far more complete than ordinary surveys, is probably still incomplete. For instance, in Pennsylvania the birth certificate provides a space for recording congenital malformations only since 1948, and several cases of the EvC Syndrome were discovered from this source. Infant mortality rate is very high: 30 of the 52 cases died under six months of age. The syndrome is also a frequent cause of stillbirth. The possibility of early fetal death cannot be ruled out. All of these factors tend to make the ascertainment incomplete and thus the segregation ratio lower than ¼.

The present set of data consists of too many large families in comparison with the distribution of family size in the general population. Table 1 shows that 12 out of the 27 sibships are of size $s \geqq 7$, and the average size is 6.4. The efficiency of the first appearance method has been found to be approximately 87% of that of the maximum likelihood method for the entire set of families. One may wonder how the relative efficiency would change with more small and fewer large families. On this question, C. S. Chung (personal communication) has kindly pointed out that relative efficiency improves with smaller sibships. He writes, "For $s \leqq 6$, the relative efficiency is .96; and for $s \leqq 4$, R.E. = 1.06 arrived at empirically from the present set of data."

It may be noted that the first appearance method is, in fact, fully efficient for sibships of two members. To show this, let us consider the following seven sibships, where R denotes a recessive (affected) and D denotes a dominant (normal).

(R,R) (R,D) (R,D) (R,D) (D,R) (D,R) (D,R)

There are four sibships with $t = 1$ and three sibships with $t = 2$. The estimates are $p_1 = p_o = \frac{1}{4}$ with respective weights

$$w_1 = 64/21, \qquad w_o = 64/3, \qquad w_1 + w_o = 512/21 = 24.38.$$

On the other hand, if we use the maximum likelihood method, the data will be considered as containing six sibships with $r = 1$ and one sibship with $r = 2$, where r is the number of recessives in a sibship. The estimate will be $\hat{p} = \frac{1}{4}$ with weight

$$W = 7 \times \frac{2}{pq} \times \frac{(1 - q^2 - 2pq)}{(1 - q^2)^2} = \frac{512}{21} = 24.38$$

Thus, the two methods are equally efficient. In a similar way, it may be shown that for $s = 3$, the efficiency is 99.9%; and for $s = 4$, 98%.

The pooled estimate p_1 from families of sizes 2, 3, and 4, however, has a lower efficiency than 98%. This is apparently due to the inefficient pooling procedure of the estimate $p_1 = (A - B)/(N - B)$. This also suggests that efficiency may be increased by calculating p_1 and its variance from sibships of each fixed size separately and then combining them by proper weighting. But there are two practical drawbacks. One is that the first appearance method will lose much of its simplicity; another is that, for some of the large families, there are no sibships with $t = 1$ or $t = s$, so that no proper estimate of p can be obtained. For the kind of data represented in Table 1, the inefficient pooled estimate is probably the most practical.

In any case, it is safe to say that the efficiency would be higher than 87%, if there were more small sibships and fewer large ones in our set of data.

SUMMARY

Two methods of estimating the segregation ratio of the Ellis-van Creveld Syndrome have been applied to 27 sibships of the Old Order Amish reported by McKusick *et al.* (1964). The results are:

First appearance method: $\bar{p} = .222 \pm .040$
Maximum likelihood method: $\hat{p} = .215 \pm .037$

The efficiency of the former relative to the latter is 86.7%. Efficiency would be higher if we had fewer large and more small families.

ACKNOWLEDGMENT

I want to thank Dr. Victor A. McKusick for calling my attention to the study of the Amish by him and his colleagues and for suggesting the application of the method of first appearance time to their data. I also wish to express my gratitude to Dr. Chin S. Chung for pointing out the high efficiency of the first appearance method for small sibships and to Dr. Herman M. Slatis for his helpful comments.

POSTSCRIPT

A Method of Improving the Estimate

After the manuscript had been sent to the printer, a method of improving the estimate of p on the basis of first appearance time occurred to the author. Instead of writing a separate note or "letter," it was thought best to attach a postscript to the article for the convenience of reference.

The method to be described is an extension of the previous method, which may be regarded as the first step in the general procedure. The method may be best explained by the aid of a diagram. In terms of Fig. 2, the sum of the first column (without the vertex entry for $t = s = 1$), denoted by A previously, is now denoted by A_1. Similarly, the sum of the diagonal entries ($t = s \geqq 2$), denoted by B previously, is now denoted by B_1. The total number of sibships is accordingly N_1. Then the estimate is taken as $p_1 = (A_1 - B_1)/(N_1 - B_1)$. Thus we see this estimate has only utilized the outermost "layer" of the triangular array of data. It will yield a good estimate for small sibships ($s = 2, 3, 4$).

When there are large sibships, we may continue the same procedure with the inner triangular array of data which has not been utilized previously. Thus, with respect to the second layer shown in Fig. 2, we may obtain the column total A_2 without the vertex indicated by ( ), and the diagonal total B_2. The total number sibships on and below the second layer is denoted by N_2. The same argument advanced by Li (1964) yields a second estimate $p_2 = (A_2 - B_2)/(N_2 - B_2)$. It is important to notice that N_2 is *not* equal to $N_1 - A_1 - B_1$ owing to the omission of the vertex entry of the inner triangular array. Thus N_2 may be regarded as the size of another independent sample in calculating the variance of p_2. Proceeding the same way to the third, fourth, etc. layer, we obtain a series of estimates with the estimation and the variance formulas all of the same form. Then they may be combined with the reciprocal of variance as weight. To summarize:

$$p_i = \frac{A_i - B_i}{N_i - B_i}$$

$$w_i = \frac{1}{V_i} = \frac{N_i - B_i}{p_i q_i} \left(\frac{A_i - B_i}{A_i + B_i} \right)$$

where $i = 1, 2, 3, \ldots$ indicates the layer of data shown in Fig. 2.

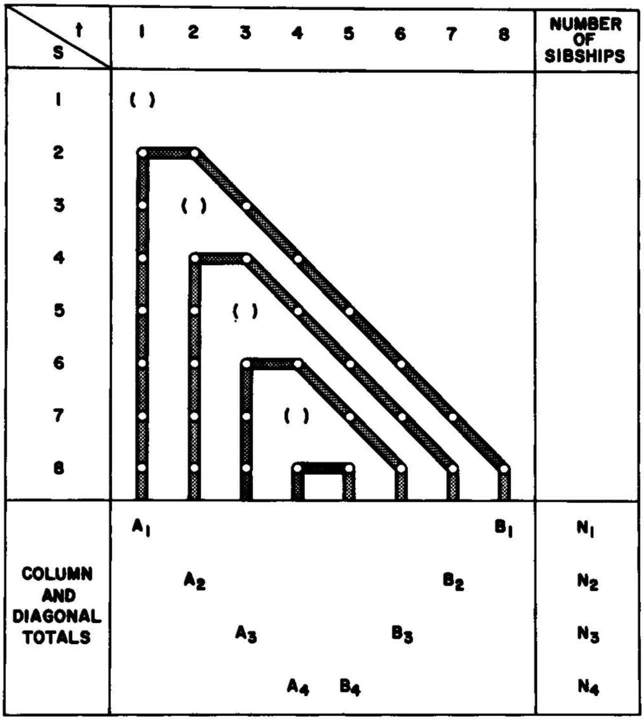

FIG. 2. Procedure of estimating the recessive proportion based on first appearance time by utilizing the successive "layers" of the triangular array of number of sibships. The vertex of each triangle, indicated by ( ), is not used in calculating the column totals (A_i) and the diagonal totals (B_i). s = size of sibship, t = first appearance time; the entries in the body of the table are n_{st} = number of sibships with size s and appearance time t.

The data on the segregation of the Ellis-van Creveld syndrome (Table 1) is too scarce to warrant a complete successive estimation procedure. Nevertheless, it would be interesting to see how the improved result would fare with the maximum likelihood estimate. Table 1 permits four successive estimates based on first appearance time. The simple counting of recessives after the first appearance remains the same as before. The five estimates and their weights are listed in Table 3. The average estimate is $\bar{p} = \Sigma w_i p_i / \Sigma w_i =$ .2214 with total weight 694.66. The relative efficiency with respect to maximum likelihood method is

$$\text{R.E.} = \frac{\text{weight of } \bar{p}}{\text{weight of } \hat{p}} = \frac{694.66}{721.55} = 96.3\%$$

The standard error of $\bar{p}$ is S.E. $(\bar{p}) = 1/\sqrt{694.66} = .0379$ while that for the maximum likelihood estimate is S.E. $(\hat{p}) = .0372$. It seems that the improved method of estimation described above is usable for all practical purposes.

TABLE 3. ESTIMATES OF RECESSIVE PROPORTION BASED ON FIRST APPEARANCE TIME AND ON COUNTING

Observations			Estimate p_i	Weight w_i	w_ip_i
$R = 21$	$D = 75$	$G = 96$	$p_0 = .21875$	561.74	122.88
$A_1 = 9$	$B_1 = 3$	$N_1 = 27$	$p_1 = .25000$	64.00	16.00
$A_2 = 3$	$B_2 = 2$	$N_2 = 15$	$p_2 = .07692$	36.62	2.82
$A_3 = 4$	$B_3 = 1$	$N_3 = 10$	$p_3 = .33333$	24.30	8.10
$A_4 = 3$	$B_4 = 1$	$N_4 = 5$	$p_4 = .50000$	8.00	4.00
			Total	694.66	153.80
			Estimate	s.e. $(\bar{p}) = .0379$,	$\bar{p} = .2214$

REFERENCES

HALDANE, J. B. S. 1932. A method for investigating recessive characters in man. *J. Genet.* 25: 251–255.

HALDANE, J. B. S. 1938. The estimation of the frequencies of recessive conditions in man. *Ann. Eugen.* (Lond.) 8: 255–262.

LEJEUNE, J. 1958. Sur une solution "a priori" de la methode "a posteriori" de Haldane. *Biometrics* 14: 513–520.

LI, C. C. 1961. *Human Genetics: Principles and Methods.* New York: McGraw-Hill.

LI, C. C. 1964. Estimate of recessive proportion by first appearance time. *Ann. Hum. Genet.* (Lond.) 28: 177–180.

MCKUSICK, V. A., EGELAND, J. A., ELDRIDGE, R., AND KRUSEN, D. E. 1964. Dwarfism in the Amish. I. The Ellis-van Creveld syndrome. *Bull. Johns Hopkins Hosp.* 115: 306–336.