A NOTE ON SEGREGATION ANALYSIS OF MULTIPLEX FAMILY[1]

Norikazu YASUDA

*Division of Genetics, National Institute of Radiological Sciences,
Chiba 260, Japan*

Summary   In segregation analysis of multiplex family, a method was presented for testing genetic hypothesis by utilizing the scores of maximum likelihood.   This source of materials is very informative to genetic analysis of high-risk cases.

INTRODUCTION

In segregation analysis of inherited diseases, data on high-risk cases sampled from multiplex families may be collected by omitting isolated cases.   In practice a selection of families with two or more affected siblings, ascertained by linking hospital records, death certificates, inquiries to physicians, examination of a population sample, or other direct means of ascertainment.   Then simplex families will not enter the sample, and the problem remaining is to develop method of analysis appropriate to multiplex sibships.   In this communication we shall suggest a simple method when the ascertainment probability is not known.

MODEL

Suppose that p and π are respectively the segregation frequency and the probability of ascertainment.   Distribution of r affected siblings in sibships with the size s will *a priori* be binomial; that is,

$$Q(r) = {}_sC_r p^r (1-p)^{s-r}.$$

After sampling multiplex families with two or more affected sibs, the *posterior* distribution $P(r|F)$ becomes

$$P(r|F) = \frac{w(F|r) Q(r)}{\sum_{r=2}^{s} w(F|r) Q(r)}$$

where $w(F|r)$ is the probability of selecting a family (F) with r affected sibs in the sample. If the ascertainment probability is uniform among sibships, the weight $w(r|F)$ will be $1-(1-\pi)^r$.

A model of segregation analysis for multiplex families thus will be

$$P(r) = \frac{{}_sC_r\, p^r (1-p)^{s-r}[1-(1-\pi)^r]}{1-(1-\pi p)^s - sp\pi(1-p)^{s-1}} \quad (r=2,\cdots,s) \qquad (1)$$

which was first derived by Morton (1959). The scores for p are

$$U = \frac{1}{p(1-p)}[A - \sum_s A_s n_s]$$

and

$$K = \frac{1}{[p(1-p)]^2}\sum_s B_s n_s$$

(definition and derivation of scores may be found in Cavalli-Sforza and Bodmer (1971), pp. 836–841), in which n_s is the observed number of sibships with the size s, A is the total number of affected individuals in multiplex families, and

$$A_s = \frac{sp[1-\pi(1-p)^{s-1}-(1-\pi)(1-p\pi)^{s-1}]}{1-(1-p\pi)^s - sp\pi(1-p)^{s-1}}$$

and

$$B_s = \frac{sp(1-p)}{E}\left[D + (s-1)p\pi F - \frac{sp\pi DG}{E}\right]$$

with $D = 1-\pi(1-p)^{s-1}-(1-\pi)(1-p\pi)^{s-1}$, $E = 1-(1-p\pi)^{s-1}-sp\pi(1-p)^{s-1}$, $F = (1-p)^{s-2}+(1-\pi)(1-p\pi)^{s-2}$ and $G = (1-p\pi)^{s-1}-(1-p)^{s-1}+(s-1)p(1-p)^{s-2}$. A_s and B_s have been tabulated in Tables 1 and 2 respectively for $p=1/4$ and $1/2$, and $s=2,3,\cdots,15$ and π is from 0.0 to 1.0 with the increment 0.1.

Single ascertainment $(\pi \to 0)$. When every family contains only one proband and the other affected sibs are secondary cases through the proband, the model becomes

$$P(r) = \frac{{}_{s-1}C_{r-1}\, p^{r-1}(1-p)^{s-r}}{1-(1-p)^{s-1}} \quad (r=2,\cdots,s) \qquad (2)$$

with

$$A_s = 1 + \frac{(s-1)p}{1-(1-p)^{s-1}}$$

and

$$B_s = (A_s-1)(1-p)[1-(A_s-1)(1-p)^{s-2}].$$

Table 1. The value of A_s as a function of s, p and π.

$$A_s = \frac{sp[1-\pi(1-p)^{s-1} - (1-\pi)(1-p\pi)^{s-1}]}{1-(1-p\pi)^s - sp\pi(-p)^{s-1}}$$

p = 1/4

s	π = 0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
2	2.00000	2.00000	2.00000	2.00000	2.00000	2.00000	2.00000	2.00000	2.00000	2.00000	2.00000
3	2.14286	2.13680	2.13090	2.12521	2.11980	2.11475	2.11017	2.10619	2.10299	2.10082	2.10000
4	2.29730	2.28434	2.27183	2.25989	2.24864	2.23827	2.22897	2.22100	2.21471	2.21051	2.20896
5	2.46286	2.44224	2.42250	2.40382	2.38641	2.37053	2.35648	2.34461	2.33538	2.32933	2.32713
6	2.63892	2.60998	2.58249	2.55670	2.53291	2.51145	2.49270	2.47709	2.46514	2.45744	2.45469
7	2.82477	2.78695	2.75130	2.71815	2.68787	2.66086	2.63756	2.61844	2.60403	2.59491	2.59171
8	3.01958	2.97245	2.92835	2.88769	2.85093	2.81851	2.79090	2.76857	2.75201	2.74170	2.73816
9	3.22250	3.16573	3.11301	3.06483	3.02169	2.98409	2.95249	2.92730	2.90893	2.89769	2.89389
10	3.43266	3.36604	3.30462	3.24899	3.19970	3.15724	3.12203	3.09439	3.07455	3.06264	3.05868
11	3.64918	3.57261	3.50252	3.43962	3.38447	3.33754	3.29916	3.26950	3.24957	3.23623	3.23219
12	3.87127	3.78469	3.70606	3.63613	3.57550	3.52456	3.48349	3.45225	3.43058	3.41804	3.41401
13	4.09814	4.00160	3.91461	3.83798	3.77229	3.71782	3.67457	3.64220	3.62014	3.60761	3.60365
14	4.32909	4.22268	4.12756	4.04461	3.97434	3.91687	3.87193	3.83887	3.81675	3.80441	3.80059
15	4.56349	4.44735	4.34438	4.25551	4.18116	4.12122	4.07511	4.04177	4.01988	4.00789	4.00425

p = 1/2

s	π = 0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
2	2.00000	2.00000	2.00000	2.00000	2.00000	2.00000	2.00000	2.00000	2.00000	2.00000	2.00000
3	2.33333	2.32224	2.31122	2.30041	2.28994	2.28000	2.27083	2.26276	2.25620	2.25170	2.25000
4	2.71429	2.68998	2.66621	2.64328	2.62153	2.60140	2.58137	2.56804	2.55609	2.54827	2.54545
5	3.11333	3.09475	3.05755	3.02229	2.94957	2.96006	2.93445	2.91345	2.89775	2.88797	2.88462
6	3.58062	3.52759	3.47720	3.43031	3.38781	3.35055	3.31929	3.29464	3.27701	3.26656	3.26316
7	4.04762	3.99047	3.91770	3.86050	3.80998	3.76704	3.73232	3.70608	3.68820	3.67812	3.67500
8	4.52756	4.44696	4.37294	4.30701	4.25041	4.20397	4.16790	4.14188	4.12501	4.11601	4.11336
9	5.01569	4.92235	4.83825	4.76521	4.70449	4.65652	4.62089	4.59644	4.58142	4.57384	4.57171
10	5.50881	5.40338	5.31034	5.23173	5.15861	5.12078	5.08693	5.06491	5.05213	5.04604	5.04442
11	6.00489	5.88794	5.78697	5.70417	5.64012	5.59373	5.56255	5.54338	5.53290	5.52820	5.52701
12	6.50269	6.37467	6.26669	6.18089	6.11712	6.07310	6.04508	6.02885	6.02052	6.01700	6.01616
13	7.00147	6.86278	6.74859	6.66083	6.59829	6.55724	6.53257	6.51914	6.51267	6.51012	6.50954
14	7.50079	7.35175	7.23209	7.14325	7.08269	7.04496	7.02359	7.01269	7.00777	7.00595	7.00556
15	8.00023	7.84132	7.71684	7.62768	7.56962	7.53538	7.51713	7.50841	7.50474	7.50346	7.50321

Table 2. The value of B_s as a function of s, p and π.

$$B_s = A_s \left\{ (1-p) + \frac{(s-1)p(1-p)\pi[(1-p)^{s-2}+(1-\pi)(1-p)^{s-2}]}{1-\pi(1-p)^{s-1}-(1-\pi)(1-p)^{s-1}} \right\} - \frac{sp(1-p)\pi[(1-p\pi)^{s-1}-(1-p)^{s-1}+(s-1)p(1-p)^{s-2}]}{1-(1-p\pi)^s-sp\pi(1-p)^{s-1}}$$

$p = 1/4$

s \ π	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	s
2	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	2
3	0.12245	0.11809	0.11377	0.10954	0.10545	0.10159	0.09803	0.09491	0.09238	0.09065	0.09000	3
4	0.26297	0.25313	0.24355	0.23435	0.22564	0.21760	0.21039	0.20426	0.19945	0.19629	0.19514	4
5	0.42005	0.40387	0.38834	0.37367	0.36007	0.34778	0.33707	0.32823	0.32154	0.31732	0.31586	5
6	0.59178	0.56865	0.54677	0.52642	0.50793	0.49161	0.47778	0.46672	0.45868	0.45384	0.45224	6
7	0.77595	0.74558	0.71722	0.69129	0.66819	0.64831	0.63194	0.61933	0.61057	0.60557	0.60403	7
8	0.97024	0.93260	0.89793	0.86677	0.83962	0.81686	0.79876	0.78539	0.77658	0.77191	0.77059	8
9	1.17237	1.12767	1.08709	1.05130	1.02086	0.99612	0.97721	0.96394	0.95580	0.95190	0.95095	9
10	1.38015	1.32882	1.28293	1.24329	1.21050	1.18480	1.16609	1.15383	1.14704	1.14431	1.14386	10
11	1.59167	1.53425	1.48378	1.44121	1.40711	1.38156	1.36408	1.35369	1.34887	1.34766	1.34779	11
12	1.80526	1.74236	1.68814	1.64364	1.60934	1.58504	1.56981	1.56209	1.55975	1.56030	1.56108	12
13	2.01956	1.95183	1.89471	1.84930	1.81591	1.79393	1.78188	1.77750	1.77805	1.78051	1.78195	13
14	2.23352	2.16159	2.10240	2.05707	2.02566	2.00701	1.99892	1.99843	2.00211	2.00654	2.00862	14
15	2.65755	2.37081	2.31033	2.26604	2.23758	2.22313	2.21966	2.22340	2.23033	2.23669	2.23937	15

$p = 1/2$

s \ π	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	s
2	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	2
3	0.22222	0.21840	0.21436	0.21016	0.20588	0.20160	0.19748	0.19372	0.19056	0.18835	0.18750	3
4	0.48980	0.48175	0.47348	0.46518	0.45709	0.44951	0.44275	0.43715	0.43301	0.43054	0.42975	4
5	0.78222	0.77119	0.76016	0.74957	0.73991	0.73164	0.72516	0.72070	0.71826	0.71745	0.71746	5
6	1.08221	1.07016	1.05868	1.04849	1.04029	1.03459	1.03165	1.03125	1.03273	1.03487	1.03601	6
7	1.37868	1.36749	1.35777	1.35058	1.34671	1.34648	1.34966	1.35527	1.36176	1.36715	1.36937	7
8	1.66656	1.65746	1.65116	1.64893	1.65146	1.65859	1.66920	1.68136	1.69267	1.70076	1.70375	8
9	1.94485	1.93846	1.93651	1.94042	1.95048	1.96565	1.98369	2.00167	2.01665	2.02640	2.02972	9
10	2.21470	2.21112	2.21393	2.22448	2.24233	2.26524	2.28969	2.31194	2.32902	2.33928	2.34255	10
11	2.47798	2.47707	2.48470	2.50194	2.52725	2.55689	2.58614	2.61083	2.62843	2.63827	2.64122	11
12	2.73656	2.73805	2.75041	2.77412	2.80618	2.84120	2.87346	2.89885	2.91571	2.92452	2.92700	12
13	2.99194	2.99559	3.01254	3.04238	3.08030	3.11922	3.15280	3.17749	3.19278	3.20024	3.20222	13
14	3.24524	3.25085	3.27227	3.30785	3.35067	3.39208	3.42556	3.44854	3.46182	3.46788	3.46939	14
15	3.49722	3.50467	3.53048	3.57139	3.61816	3.66079	3.69307	3.71375	3.72489	3.72964	3.73075	15

The results are in fact the same as those from distribution of a probands among r affected sibs (model (14) in Yasuda (1979)), replacing π by p, r by s $-$ 1 and a by r $-$ 1.

Complete ascertainment $(\pi = 1)$. If all affected sibs are proband, the method of ascertainment is said to be complete or $\pi = 1$. The model (1) will be

$$P(r) = \frac{{}_sC_r\, p^r (1-p)^{s-r}}{1-(1-p)^s - sp(1-p)^{s-1}} \quad (r = 2, \cdots, s) \tag{3}$$

with

$$A_s = \frac{sp[1-(1-p)^{s-1}]}{1-(1-p)^s - sp(1-p)^{s-1}}$$

and

$$B_s = \frac{[1-(1-p)^{s-1}]^2 - (s-1)^2 p^2 (1-p)^{s-2}}{[1-(1-p)^s - sp(1-p)^{s-1}]^2} .$$

These results are again mathematically equivalent to those from distribution of multiplex probands ((21) in Yasuda (1979)).

NUMERICAL EXAMPLE

A distribution of congenital deafs among sibs born to normal parents has been reported for Kanto district (Furusho and Yasuda, 1973). Apart from sporadic cases the data were compatible with the hypothesis of autosomally recessive modes of inheritance. As an illustration, the hypothesis will be tested by the multiplex families only. Table 3 is reproduced a part of table for observed distribution of affected sibs within sibships or "SR table" (Table 4 in Furusho and Yasuda (1973)) taking only multiplex deafs among sibs born to normal parents. The total number of deafs is thus $A = \sum_{s,r} rn_{sr} = 565$, where n_{sr} is the number of sibships with the size s containing r affected sibs.

Multiple ascertainment. The data have been gathered through sibs' probands and contained the secondary cases. The distribution of probands among affected siblings was observed as "RA table," from which the probability of ascertainment was estimated to be 0.5011. For illustrative purpose we assume $\pi = 0.5$. Following Cavalli-Sforza and Bodmer (1971), their notation θ_0 corresponds to p $= 1/4 = 0.25$ in our terminology. Also $S(\theta_0)$ and $I(\theta_0)$ correspond respectively to U and K computed by assuming p $= 0.25$ and $\pi = 0.5$ by the aid of Tables 1 and 2, and the procedure of calculations is summarized in Table 4. We found U $= -53.22$ and K $= 2,259$ so that $\chi^2 = U^2/K = 1.25$ with one degree of freedom being not significant at 5% level. Then we have estimate of segregation frequency $\hat{p} = 0.25 + U/K = 0.23$ with the standard error $\sqrt{I/K} = 0.02$. It should be emphasized here that those calculations do not themselves prove the hypothesis of recessive mode of inheritance

Table 3.   Distribution of multiplex deafs among sibs born to normal parents (n_{sr}). (Furusho and Yasuda, 1973)

s \ r	2	3	4	5	6	n_s
2	25					25
3	42	7				49
4	46	7	2			55
5	37	10	2	1		50
6	17	4	1	1		23
7	19	4	2		1	26
8	6	3	2			11
9		2	2			4
10		2				2
12			1			1
sum	192	39	12	2	1	246

Table 4.   Segregation analysis of multiplex deafs among sibs born to normal parents ($p = 1/4$ and $\pi = 0.5$).

s	n_s	A_s	B_s
2	25	2.00000	0.00000
3	49	2.11475	0.10159
4	55	2.23827	0.21760
5	50	2.37053	0.34778
6	23	2.51145	0.49161
7	26	2.66086	0.64831
8	11	2.81851	0.81686
9	4	2.98409	0.99612
10	2	3.15724	1.18480
12	1	3.52456	1.58504

$A = 565$. $\sum A_s n_s = 574.97882$ $\sum B_s n_s = 79.42258$

$$U = \frac{565. - 574.97882}{0.25(1 - 0.25)} = -53.22037$$

$K = 79.42258/[0.25(1 - 0.25)]^2 = 2{,}259.13116$

$\sqrt{1/K} = 0.0210$, $U/K = -0.0235$

$\hat{p} = 0.25 - 0.02 = 0.23$, $U^2/K = 1.25$.

A_s, from Table 1; B_s, from Table 2.

in deafs, but do infer the distribution of affected individuals in siblings does not differ significantly from that would be expected if the disease were due to such a gene.

Single ascertainment ($\pi \to 0$).   Very often in practice each family was selected through a single proband or was assumed simply single ascertainment.   Let us suppose that data of deafs were sampled by the method of single ascertainment for an illustrative purpose.   Taking the similar steps as in the previous case by using the values for $\pi = 0$ and $p = 0.25$ in Tables 1 and 2, we arrive at $U = -162.54$ and $K = 2{,}707$.   A value of chi-square for p will be $U^2/K = 9.76$ with one degree of freedom being highly significant ($p < 0.005$).   Maximum likelihood estimate for the segregation frequency becomes $\hat{p} = 0.25 + U/K = 0.19$ with the standard error 0.02, and which agrees with Weinberg's estimator* of 0.20.   This result simply indicates that the method of ascertainment is critical for testing genetic hypothesis.

Complete ascertainment ($\pi = 1$).   The other ideal extreme in method of ascertainment is complete or every affected individual being proband.   Computations using the value for $\pi = 1$ and $p = 0.25$, we have $U = -4.26$ and $K = 2{,}081$ so that $U^2/K = 0.01$ and $\hat{p} = 0.25 \pm 0.02$.   It is not significant at five percent level.

---

* Weinberg's estimator of p (Weinberg, 1912) becomes $\sum (r-1)(r-2) n_{sr} / \sum (r-1)(s-2) n_{sr}$ for multiplex family data.

DISCUSSION

The above numerical example illustrates an importance of parameter π in testing genetic hypothesis. Careful designs of family selection is therefore of important. In most practice however definition of proband among affected siblings is vague or even not defined explicitly. The probability of ascertainment is then not able to estimate quantitatively, but is rather guessed by investigators the quality of their survey and/or the experience of surveyors. Following suggestion then may be practical. Since the parameter p changes almost linearly when π varies, minimum at $\pi=0$ and maximum at $\pi=1$ (Yasuda, 1981), the necessary scores may be obtained from averaging the ones calculated under the assumptions of single and complete ascertainments, respectively. In the present example we have $U=[(-162.54)+(-4.36)]/2=-83.45$, and $K=(2,707+2,081)/2=2,394$, which yield $\chi^2=U^2/K=2.91$ being not significant at 5% level and $\hat{p}=0.22\pm0.02$. The results are practically the same those obtained by multiple ascertainment. This is actually due to an implicit assumption of $\pi=0.5$ in the above conventional method. In general approximate formulae for the scores may be used in practice; namely,

$$U=(1-\pi)\,U(0)+\pi U(1) \; \rbrace$$
and
$$K=(1-\pi)\,K(0)+\pi K(1) \; \rbrace \qquad (4)$$

in which (0) and (1) designate respectively single and complete ascertainments.

It can be shown for using (4) that $\chi^2=U^2/K$ will be less than 3.84, a critical value to be significant at 5% level, for $0.42\leq\pi<1$ in the present example, indicating a robustness for testing segregation frequency p for ascertainment probability π. The following procedures thus may be instructive for testing genetic hypothesis when no appropriate estimator of π would be available. First, calculate the scores under the hypothesis of single and complete ascertainment, respectively. Then, calculate U and K from (4) and/or use of Tables 1 and 2, assuming various values of π, for instance, 0.2 for moderately single ascertainment, 0.4 for moderately multiple, and 0.6 for moderately complete ascertainment. For each value of π, $\chi^2=U^2/K$ can be computed for testing the hypothesis. This procedure may give a clue for decision making on genetic hypothesis in addition to experiences in family selection. This approach however is no use to the material in which $U(0)>0$ and $\chi^2=U^2/K$ was significant, because the segregation frequency p almost linearly increases when π becomes large. Most likely explanations for such observations is the sampling bias towards to many multiplex families because of interests in high-risk cases, and/or wrong genetic hypothesis being tested. In former situation inclusion of simplex families in analysis might give a clue. It is strongly recommended that several methods of analysis should be applied for a given set of data for establishing mode of inheritance in diseases.

The use of multiplex families might be less common practice in industrialized

countries where the size of sibship has been reduced, but the method is still very informative in genetic analysis of high-risk cases and such data can be gathered from intensive surveillance systems which would be developed in industrialized nations.

REFERENCES

Cavalli-Sforza, L.L., and Bodmer, W.F. 1971.   *The Genetics of Human Populations.*   Freeman, San Francisco, pp. 836-841.
Furusho, T., and Yasuda, N. 1973.   Genetic studies on inbreeding in some Japanese population. XIII. A genetic study of congenital deafness.   *Jpn. J. Human Genet.* **18**: 47–65.
Morton, N.E. 1959.   Genetic tests under incomplete ascertainment.   *Am. J. Hum. Genet.* **11**: 1–16.
Weinberg, W. 1912.   Uber Methode und Fehlerquellen der Untersuchung auf Mendelsche Zahlen beim Menschen.   *Arch. Rass. -u Ges. Biol.* **6**: 165–174.
Yasuda, N. 1979.   Estimation of the ascertainment probability of rare disease.   *Jpn. J. Human Genet.* **24**: 265–291.
Yasuda, N. 1981.   A statistical analysis of the heterogeneity of inherited diseases.   *Jpn. J. Human Genet* **26**: 1–17.