Am. J. Hum. Genet. 50:1328–1332, 1992

Rh Segregation Distortion: An Artifact of Ascertainment Bias?

Swapan K. Nath,* Partha P. Majumder,* and Sumita Sarkar†

*Anthropometry and Human Genetics Unit, Indian Statistical Institute, Calcutta; and †Department of Anthropology, Haldia Government College, Haldia, India

Summary

On the basis of Rh blood group data on mother-child pairs collected from the maternity clinic of a hospital in Chile, Valenzuela and Harb postulated that there is a significant segregation distortion at the Rh locus. For data collected from a hospital, biases of ascertainment cannot be ruled out. For the Rh blood group locus, there is a strong possibility of preferential admission of Rh(−) pregnant women, especially of those Rh(−) women with Rh(+) husbands. We show that the evidence of segregation distortion vanishes when the possibility of such preferential admissions are taken into account.

Introduction

On the basis of data on Rh blood group phenotypes of mother-child pairs, Valenzuela and Harb (1982) claimed that there is a significant segregation distortion at the Rh blood group locus (also see Cifuentes et al. [1991]), and postulated the existence of another compatibility system closely linked or physiologically related to the Rh. Specifically, they noted that the frequency of Rh(−) children born to Rh(+) mothers is significantly less than the frequency of Rh(+) children born to Rh(−) mothers, although these two frequencies are expected to be equal under Hardy-Weinberg equilibrium (see table 1). They concluded that "something prevents (or delays) production of dd children from Dd mothers and allows a relatively higher production of Dd children from dd mothers" (Valenzuela and Harb 1982, p. 932). While segregation distortion at any genetic locus is possible, it is rather difficult to detect and requires much stronger evidence (of the type provided with respect to the mouse t-complex [Silver 1981] or the segregation distorter factor in *Drosophila* [Hartl and Hiraizumi 1976]) than that provided by Valenzuela and Harb (1982). In the past, initial claims of segregation distortion at specific loci in humans have often been refuted. For example, the initial claim of segregation distortion at the alpha-1-antitrypsin locus in humans (Chapuis-Cellier and Arnaud 1979; Immarino et al. 1979) was refuted on the basis of more extensive data and statistical analyses (Chakraborty et al. 1982; Constans et al. 1982).

The purpose of the present paper is to reexamine Valenzuela and Harb's (1982) data, especially because it is our belief that segregation distortion is an uncommon biological phenomenon and that the type of discrepancies between observed and expected phenotype frequencies found by them can also arise because of other reasons, including biases of ascertainment. For the purpose of self-containment of the present paper, we reproduce in table 1 their total data and present the probabilities of the various mother-child pairs under Hardy-Weinberg equilibrium. In this table, as also in the remainder of this paper, q denotes the population frequency of the Rh(−) allele, d. To recapitulate, the discrepancy noted by Valenzuela and Harb (1982) is that there are only 449 Rh(+)-Rh(−) mother-child pairs compared with 544 Rh(−)-Rh(+) mother child pairs (see table 1). As is also seen from table 1, the probabilities of each of these types of mother-child pairs is $q^2(1-q)$ under Hardy-Weinberg equilibrium. We note that the data presented in table 1 were collected from the Maternity Service of the Clinica Alemana, a private hospital in Santiago, Chile.

From the marginal frequencies of the data in table 1, it is evident that the number of Rh(−) mothers (804) is greater than the number of Rh(−) children (709); these two frequencies are expected to be equal

Address for correspondence and reprints: Partha P. Majumder, Ph.D., Anthropometry and Human Genetics Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta 700 035, India.

Table 1

Mother-Child Phenotypic Combinations

Mother's Blood Type[a]	Child's Blood Type: Rh(+)	Child's Blood Type: Rh(−)	Total
Rh(+):			
Observed	$n_{++} = 5{,}721$	$n_{+-} = 449$	$n_{+\cdot} = 6{,}170$
Probability	$1-2q^2+q^3$	$q^2(1-q)$	$1-q^2$
Expected	5,713.67	456.33	6,170
Rh(−):			
Observed	$n_{-+} = 544$	$n_{--} = 260$	$n_{-\cdot} = 804$
Probability	$q^2(1-q)$	q^3	q^2
Expected	553.60	250.40	804
Total:			
Observed	$n_{\cdot+} = 6{,}265$	$n_{\cdot-} = 709$	$n_{\cdot\cdot} = 6{,}974$
Probability	$1-q^2$	q^2	1
Expected	6,267.27	706.73	6,974

[a] Observed frequencies are from total data of Valenzuela and Harb (1982), probability is that under Hardy-Weinberg equilibrium, and expected frequencies are those under a proposed model of ascertainment bias (model 1).

under Hardy-Weinberg equilibrium. Our reexamination of these data was prompted by this observation and by the following statement made by Valenzuela and Harb (1982, p. 929): "There is a higher rate of admission of Rh(−) mothers than what is expected from a panmictic equilibrium." In spite of making this statement, Valenzuela and Harb (1982) have not considered, in any detail, its statistical implications and have provided only some indirect, nonquantitative arguments against any ascertainment bias (which in the present case relates to preferential admission of Rh(−) mothers to the clinic from where the data were collected). Our analyses rely on facts known about Rh incompatibility, one of which is the awareness of clinical problems arising in women who are themselves Rh(−) but have Rh(+) spouses may be an important factor in the preferential admission of Rh(−) women to hospitals for maternity service. Such preferential admissions to maternity clinics may be voluntary and/or on physicians' advice. Our main tenet, therefore, is that the observed discrepancy is possibly due to ascertainment bias rather than to any true biological phenomenon such as segregation distortion.

Models and Statistical Considerations

We first note that if, because of preferential admission or otherwise, there is indeed a higher frequency of Rh(−) mothers in the sample, then estimators of q should be constructed only on the basis of the observations n_{++}, n_{+-}, and $n_{+\cdot}$ (i.e., on the basis of the observations in the first row of table 1), unless all the observations are suitably adjusted to account for the ascertainment bias. Otherwise, the resulting estimates of q will be biased. We note that, in the context of estimation of allele frequencies from ABO blood group data collected from blood banks where there is usually an overrepresentation of individuals with the O blood group, adjusted estimators of gene frequencies were suggested by Smith (1967). Later, it was shown by Li (1986) that the same estimates of gene frequencies can also be obtained by discarding the data of O blood-group individuals. However, the estimates obtained by discarding the O blood-group data have larger variances. Valenzuela and Harb (1982) have considered seven estimators of q from the data of table 1. Of these, only one estimator, "q(d)" in the notation of Valenzuela and Harb (1982), is based solely on the observations in the first row of table 1. This estimator is defined as $q(\mathrm{d}) = n_{+-}/2n_{+\cdot} + (1/2)\sqrt{[4n_{+-}/n_{+\cdot} + (n_{+-}/n_{+\cdot})^2]}$. Their other estimator relevant to our discussion is q(c), which is the maximum-likelihood estimator based on the observations n_{++}, n_{+-}, n_{-+}, and n_{--}. The estimates from the data of table 1 were $q(\mathrm{c}) = .3291$ and $q(\mathrm{d}) = .3086$. The reason why the estimate q(c) is higher than q(d) is that there is an overrepresentation of Rh(−) mothers which has not been corrected for. Therefore, if only Rh(−) mothers are overrepresented in the sample, while the frequency of Rh(+) mothers in the sample is the same as that expected under Hardy-Weinberg equilibrium, then q(d) is a valid estimate of

q, while q(c) is not. The goodness-of-fit χ^2 value for the complete 2 × 2 table based on q(d) is 33.54, and the χ^2 value based on q(c) is 9.86, both of which are significant at the 5% level with 2 df. Thus, we see that the various "uncorrected" estimators of q that are obtained from the data do not provide good fits to the data of table 1. This led Valenzuela and Harb (1982) to postulate the hypothesis of segregation distortion at the Rh locus.

In an effort to investigate whether estimators obtained after correcting for the possible ascertainment bias can provide adequate fits to the observed data, we have considered two models. For each of these models, we assume that both parents of a child are drawn randomly from a population which is in Hardy-Weinberg equilibrium. We further assume that parents are unrelated.

Model 1

Suppose there is preferential admission of Rh(−) mothers by a factor ε_1. Then, in the sample there will obviously be an overrepresentation of Rh(−)-Rh(+) and Rh(−)-Rh(−) mother-child pairs, and there will be a proportional underrepresentation of Rh(+)-Rh(+) and Rh(+)-Rh(−) mother-child pairs. The frequencies of the various mother-child combinations in the sample will be as presented in table 2. By use of the data of the entire 2 × 2 table, it is easy to obtain maximum-likelihood estimates of the parameters q and ε_1. The corresponding goodness-of-fit χ^2 test has 1 df.

In this parametrization, $\hat{\varepsilon}_1$ is an estimate of the rate of preferential admission of Rh(−) mothers. We also note that the null hypothesis of no segregation distortion can be tested without the introduction of any parameter such as ε_1. If there is preferential admission of Rh(−) mothers, then the samples of Rh(+) and Rh(−) mothers can be viewed as independent samples. Within each of these samples, the proportions of offspring will be binomially distributed. Under the null hypothesis of no segregation distortion, the likelihood of the data (L_0) on mother-child pairs will be (see table 1)

$$L_0 \propto [(1+q-q^2)/(1+q)]^{n_{++}} [q^2/(1+q)]^{n_{+-}} (1-q)^{n_{-+}} q^{n_{--}}. \quad (1)$$

The unconstrained likelihood (L_1) will be

$$L_1 \propto (n_{++}/n_{+\cdot})^{n_{++}} (n_{+-}/n_{+\cdot})^{n_{+-}} (n_{-+}/n_{-\cdot})^{n_{-+}} (n_{--}/n_{-\cdot})^{n_{--}}. \quad (2)$$

The maximum-likelihood estimate of q can be obtained from equation (1), and the null hypothesis can be tested using the 1 df χ^2 statistic $-2(\ln\hat{L}_0 - \ln\hat{L}_1)$, where $\hat{L}_0$ and $\hat{L}_1$ are the maximum values of the likelihood functions. The estimate of q and the χ^2 values thus obtained will be approximately equal to those obtained under the model proposed above. (For the present data, these values were exactly the same as those presented in table 4.)

Model 2

The above model does not consider whether all Rh(−) mothers get preferentially admitted or whether only a subset of these mothers—specifically, those with Rh(+) spouses—get preferentially admitted. This can also be taken into account. Since it is more likely for Rh(−) mothers with Rh(+) spouses to get preferentially admitted, we assume that there is such an overrepresentation, by a factor ε_2, in our sample. Thus, the proportions of various mating types in the sample will be as given in table 3. From these mating

Table 2

Probabilities of Mother-Child Phenotypic Combinations under the Two Proposed Models of Ascertainment Bias

Mother's Blood Type	Child's Blood Type			
	Model 1[a]		Model 2[b]	
	Rh(+)	Rh(−)	Rh(+)	Rh(−)
Rh(+)	$1-2q^2+q^3$	$q^2(1-q)$	$1-2q^2+q^3$	$q^2(1-q)$
Rh(−)	$q^2(1-q)(1+\varepsilon_1)$	$q^3(1+\varepsilon_1)$	$q^2(1-q)(1+\varepsilon_2)$	$q^3[1+\varepsilon_2(1-q)]$

[a] All cell probabilities are multiplied by $(1+\varepsilon_1 q^2)$.
[b] All cell probabilities are multiplied by $[1+\varepsilon_2 q^2(1-q^2)]$.

Table 3

Mating Frequencies in the Sample, under a Proposed Model of Ascertainment Bias (Model 2)

Mother's Blood Type	Father's Blood Type	
	Rh(+)	Rh(−)
Rh(+)	$(1-q^2)^2$	$(1-q^2)q^2$
Rh(−)	$q^2(1-q^2)(1+\varepsilon_2)$	q^4

NOTE.—All probabilities are multiplied by $[1+\varepsilon_2 q^2(1-q^2)]$.

frequencies, the frequencies of various mother-child combinations can be calculated (Li and Sacks 1954); these are given in table 2. The probability model of table 2 can be used to obtain maximum-likelihood estimators of the parameters q and ε_2, and a 1 df goodness-of-fit χ^2 test can be performed.

Results and Discussion

Valenzuela and Harb (1982) partitioned their total data into three subsets based on the ethnic background of the mother. They found that there is a significant discrepancy between observed mother-child phenotype frequencies and those expected under Hardy-Weinberg equilibrium, for the total data and also for one of the three subsets (i.e., "One Chilean name" in table 4). Their postulation of the hypothesis of segregation distortion at the Rh locus is based on this finding.

Since the mother-child pairs were not sampled randomly from a population but, rather, from a maternity clinic, we considered it important to take into account possible biases of ascertainment that are due to preferential admission of mothers of a particular phenotype. We have, therefore, proposed two statistical models which take into account such biases of ascertainment. Maximum-likelihood estimates of parameters for each of the two models were obtained both from the total data and also separately from the data of the three subsets. The results are given in table 4. We find that the goodness-of-fit χ^2 values are all nonsignificant at the 5% level, for both the models. (We provide in table 1 the expected frequencies, under model 1, for the total data.) It may be noted that, even for those subsets for which Valenzuela and Harb (1982) did not find significant χ^2 values, our models provide better fits, as is indicated by the corresponding per-df χ^2 values. We therefore conclude that the data given by Valenzuela and Harb (1982) show no evidence of significant deviation from Hardy-Weinberg expectations if the possible ascertainment bias is taken into account. While this does not exclude the possibility of segregation distortion at the Rh locus, what we have been able to show in the present paper is that the conclusion of segregation distortion is not warranted by the present data and that the observed deviation may be an artifact of ascertainment bias.

An interesting related question is: What is the rate of preferential admission that can potentially lead to

Table 4

Results of Model-Fitting to Data Given by Valenzuela and Harb (1982)

Data Subset and Parameter	Hardy-Weinberg Equilibrium	Model 1	Model 2
Two Chilean names:			
$\hat{q}$	.3187 ± .0059	.3055 ± .0095	.3087 ± .0089
$\hat{\varepsilon}$	. . .	.1606 ± .1000	.1398 ± .0974
χ^2 (df)	3.58 (2)	.62 (1)	1.31 (1)
One Chilean name:			
$\hat{q}$	.3226 ± .0091	.2914 ± .0144	.2943 ± .0136
$\hat{\varepsilon}$	. . .	.4188 ± .1877	.4298 ± .1838
χ^2 (df)	8.29* (2)	.95 (1)	.21 (1)
Two foreign names:			
$\hat{q}$	.3654 ± .0101	.3518 ± .0166	.3581 ± .0150
$\hat{\varepsilon}$	. . .	.1456 ± .1529	.0920 ± .1458
χ^2 (df)	2.56 (2)	1.57 (1)	2.18 (1)
Total:			
$\hat{q}$	.3291 ± .0045	.3114 ± .0072	.3152 ± .0067
$\hat{\varepsilon}$	. . .	.2132 ± .0768	.1929 ± .0748
χ^2 (df)	9.86* (2)	.66 (1)	2.14 (1)

* $P < .05$.

the difference between observed and expected mother-child phenotypic frequencies of the magnitude found by Valenzuela and Harb (1982)? The answer is contained in table 4. Consider, for example, their total data. The estimate of ε_1 is .2131. This means that, in comparison with the Hardy-Weinberg equilibrium frequency, if 21.31% of Rh(−) mothers get preferentially admitted (i.e., if approximately 6 Rh(−) mothers get admitted instead of 5 Rh(−) mothers), then it is possible to obtain significant χ^2 values and to falsely conclude that there is segregation distortion.

Another interesting feature that is observed in these models is that not only is there a preferential admission of Rh(−) mothers with Rh(+) spouses, but there is also a preferential admission of Rh(−) mothers with Rh(−) spouses. We draw this conclusion because, if the preferential admission of Rh(−) mothers was solely due to the preferential admission of the subset with Rh(+) spouses, then the goodness-of-fit χ^2 values under model 2 would have been of the same magnitude as those under model 1. However, it is seen from table 4 that the χ^2 values under model 1 are generally about two or three times smaller than those under model 2. Sampling fluctuations may also lead to such differences; however, since three of the four χ^2 values show the same trend, our conclusion regarding the preferential admission even of Rh(−) mothers with Rh(−) spouses seems more plausible.

Acknowledgment

We are grateful to Professor C. C. Li and an anonymous reviewer for valuable comments on an earlier draft of this paper.

References

Chakraborty R, Constans J, Majumder PP (1982) Transmission of the Pi^Z allele for α_1-antitrypsin deficiency: population genetic considerations. Hum Genet 62:193–197

Chapuis-Cellier C, Arnaud P (1979) Preferential transmission of the Z deficient allele of α_1-antitrypsin. Science 245: 407–408

Cifuentes L, Nazer J, Valenzuela CY (1991) Segregation distortions of the ABO and Rh systems in malformed newborns. Hum Hered 41:195–200

Constans J, Chakraborty R, Majumder PP (1982) Transmission of Z allele from heterozygous males for α_1-antitrypsin deficiency: additional family data. Am J Hum Genet 34: 674–675

Hartl DL, Hiraizumi Y (1976) Segregation distortion in Drosophila melanogaster. Genetics 86:321–325

Immarino RM, Wagener DK, Allen RC (1979) Segregation distortion of the α_1-antitrypsin Pi-Z allele. Am J Hum Genet 31:508–517

Li CC (1986) A method of subdividing genetic data into self-contained subsets. Ann Hum Genet 50:259–270

Li CC, Sacks L (1954) The derivation of joint distribution and correlation between relatives by the use of stochastic matrices. Biometrics 10:61–81

Silver LM (1981) Genetic organization of the mouse t-complex. Cell 28:239–240

Smith CAB (1967) Notes on the gene frequency estimation with multiple alleles. Ann Hum Genet 31:99–107

Valenzuela YC, Harb Z (1982) A mother-child segregation distortion for the Rh system: new evidence for another compatibility system associated with Rh. Am J Hum Genet 34:925–936