

Analysis of sex distribution within families in a large Latin American sample

Mary Furlan Feitosa¹ and Henrique Krieger²

ABSTRACT

A sample of 59,496 Latin American individuals belonging to 28,545 sibships was taken from a large case-control Collaborative Study on Congenital Malformations (ECLAMC). Newborn babies with congenital malformations, and their respective controls, were excluded from the sample in order to avoid biases associated with malformation. On the assumption that Poisson and Markovian effects are not important in determining the overall population distribution of sibships, some statistical models were applied. The general model was a double binomial with one-tail excess. This model can be simplified to a double binomial, a simple binomial with one-tail excess or a simple binomial. The first three distributions fitted the data well ($\chi^2_{15} = 16.46$, $P = 0.35$; $\chi^2_{16} = 20.92$, $P = 0.18$; $\chi^2_{17} = 19.88$, $P = 0.28$), while the simple binomial did not ($\chi^2_{18} = 35.01$, $P = 0.009$). The same models were also applied to several studies carried out in different countries at different times, with similar results. The most parsimonious model that fits most data sets is the double binomial, the results from which suggest that 9% of couples segregate X-linked recessive lethals, since the two segregation proportions of the model are closely in line with the genetic prediction.

INTRODUCTION

Several studies have focused on familial distribution of sex, using a variety of statistical techniques and samples derived from census or church records (cf. Edwards, 1962 and James, 1987; for extensive reviews). In general, a binomial distribution does not fit large familial samples (Schutzenberger, 1949; Renkonen, 1956; Edwards, 1959; Renkonen *et al.*, 1961; Greenberg and White, 1967; Crouchley and Pickles, 1984), and several alternative models have also failed to provide a good fit with the data. Although some studies have found evidence that appears to explain this phenomenon (Renkonen, 1956; Edwards,

1959; Renkonen *et al.*, 1961; Crouchley and Pickles, 1984) it is unlikely that these explanations can be generalized to most of the human data on familial distribution of sex. In the present study, a few probabilistic models were applied to the sibship distribution of sex in a relatively large Latin American sample in an attempt to establish a similarity of pattern with the available data on other populations.

SUBJECTS AND METHODS

Data were taken from the Latin American Collaborative Study on Congenital Malformations - ECLAMC (Castilla and Orioli, 1983). This program is a clinical and epidemiological investigation of congenital defects, based on hospital births, with accumulated data from 147 hospitals in 64 towns and cities, in 11 countries (Uruguay, Chile, Argentina, Brazil, Bolivia, Peru,

¹ Departamento de Genética, Fundação Oswaldo Cruz, Rio de Janeiro, RJ, Brasil.

² Departamento de Parasitologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, Cidade Universitária, 05508-900 São Paulo, SP, Brasil. Send correspondence to H.K.

Paraguay, Ecuador, Venezuela, Colombia and Costa Rica). ECLAMC recorded 1,886,653 live-births in the period 1967-87, and 24,818 still-births in 1978-87 (Feitosa and Krieger, 1992). The following analyses were based on part of these data comprising 28,545 incomplete families, with sibship sizes of up to five live-births, since larger families were rather infrequent and some sibship types would introduce expected numbers under five. These are incomplete families, all the couples had at least one more child. ECLAMC is a case-control program, and both in the proband and control families the last outcome was discarded, in order to avoid bias due to sex-malformation association. The two types of families (proband and control families) were pooled, since there was no significant difference between the sex ratio of the proband sibship and the control ones: $\chi^2_1 = 0.20$, $P = 0.65$).

The distribution of sex was analyzed using the following models of segregation analysis, with complete ascertainment.

The general model, i.e. double binomial with one-tail excess is expressed by either of the two following equations:

$$L_1 = P(s, r = 0) = T + (1 - T) \{ (1 - w) (1 - m)^s + w (1 - p)^s \}$$

or

$$L_2 = P(s, r > 0) = (1 - T) (r^s) \{ (1 - w) (1 - m)^{s-r} m^r + w (1 - p)^{s-r} p^r \},$$

where r is the number of males among s sibs; m is the probability of a male child among families with a relatively low probability of having male offspring; p is the probability of a male child in high male probability families; w is the proportion of the population belonging to this last type of family, and T is the prevalence of families that are unable to segregate male offspring.

From this general model, others can be obtained with the following simplifications: fixing $T = 0$ gives the double binomial; fixing $w = 0$, the simple binomial with one tail excess, and, finally, fixing both $T = 0$ and $w = 0$, the simple binomial.

Parameters were estimated by maximum likelihood (ML), using iterative non-linear optimization procedures (Fletcher and Powell, 1963). A parameter vector q was attained in order to minimize the gradient of ML scores, with elements $\partial \ln L / \partial q_i$, where q_i is the i th parameter. At the maximum likelihood point, the inverted Hessian matrix furnishes the variance-covariance matrix. Usual goodness of fit tests were

applied, in order to infer the validity of the models. Classes with expected values lower than five were pooled in order to avoid inflation of type I errors.

RESULTS AND DISCUSSION

Table I gives the familial distribution of males according to sibship size. Segregation analysis models applied to these data showed (Table II) that only the simple binomial did not fit the data ($P = 0.009$). Among the other models that fit the data, the simple binomial with one-tail excess was the most parsimonious.

These segregation models were also applied to published data from France (Schutzenberger, 1949; Edwards, 1959), Sweden (Edwards and Fraccaro, 1960), Finland (Renkonen, 1956; Renkonen *et al.*, 1961; Crouchley and Pickles, 1984) and the United States (Greenberg and White, 1967) (Table III). For the Saxony families (Geissler, 1889) none of the models fit the data. This fact is probably due to ascertainment (Edwards, 1958), since several analyses were done on this material and no simple explanation could explain the results (Gini, 1908; Waaler, 1928; Harris and Gunstad, 1930a,b; Slater, 1944; Lancaster, 1950; Fisher, 1958; Stern, 1960).

Leaving aside small samples, which usually do not show significant deviations from the expected binomial distribution, the samples in the literature can be divided into two classes: a) one class that none of the

Table I - Family distribution of sex ratio.

Sibship size (s)	Males (r)	N _{s,r}
1	0	6164
1	1	6573
2	0	1816
2	1	3745
2	2	2040
3	0	541
3	1	1491
3	2	1677
3	3	571
4	0	166
4	1	567
4	2	911
4	3	583
4	4	191
5	0	56
5	1	215
5	2	489
5	3	449
5	4	241
5	5	59

Table II - General description of the models, of their respective goodness of fit tests and of estimated parameters in a Latin American sample.

Model	Goodness of fit				Estimated parameters and standard errors			
	χ^2	d.f.	-2 lnL	d.f.	T	w	m	p
Double binomial with one-tail excess	16.460	15	16.464	15	0.0098 (0.0040)	0.0072 (0.0156)	0.5130 (0.0067)	0.9950 (0.7836)
Double binomial	20.916	16	20.917	16		0.9703 (0.0325)	0.1657 (0.1739)	0.5222 (0.1116)
Simple binomial with one-tail excess	19.881	17	19.906	17	0.0120 (0.0033)		0.5176 (0.0027)	
Simple binomial	35.013	18	34.080	18			0.5114 (0.0020)	

T: Proportion of couples in the population without offspring; w: proportion of the population with a high probability of having a male child; p: probability of a male child in high male probability families; m: probability of a male child among families with a relatively low probability of having male offspring.

Table III - Segregation analysis of several large populational samples.

Sample	Size	Model	Goodness of fit			Parameters				Reference
			χ^2	-2 lnL	d.f.	T	w	m	p	
Latin America (present)	28,545	S.B.	35.01*	34.08*	18			0.511		
		S.B.O.E.	19.88	19.91	17	0.012		0.518		
		D.B.	20.92	20.92	16		0.970	0.166	0.522	
		D.B.O.E.	16.46	16.46	15	0.010	0.007	0.513	0.995	
France (1)	11,879	S.B.	48.79*	46.71*	28			0.506		Edwards (1959)
		S.B.O.E.	44.98*	43.55*	27	0.002		0.507		
		D.B.	26.76	26.92	26		0.847	0.387	0.527	
		D.B.O.E.	27.30	27.55	25	0.0001	0.719	0.421	0.539	
(2)	14,230	S.B.	129.55*	118.60*	75			0.507		Schutzenberger (1949)
		S.B.O.E.	96.61*	118.38*	74	0.00004		0.503		
		D.B.	59.83	80.42	73		0.759	0.413	0.536	
		D.B.O.E.					No convergence			
Sweden	19,145	S.B.	13.76	14.03	33			0.518		Edwards and Fraccaro (1960)
		S.B.O.E.	13.88	14.16	32	0.0001		0.518		
		D.B.	13.77	14.04	31		0.999	0.345	0.518	
		D.B.O.E.					No convergence			
Finland (1)	60,334	S.B.	54.45*	53.41*	10			0.503		Renkonen (1956)
		S.B.O.E.	13.42	13.41	9	0.013		0.510		
		D.B.	7.81	7.80	8		0.772	0.382	0.539	
		D.B.O.E.					No convergence			
(2)	123,445	S.B.	91.57*	89.59*	16			0.506		Renkonen and cols. (1961)
		S.B.O.E.	18.70	18.73	15	0.009		0.511		
		D.B.	12.59	12.60	14		0.967	0.208	0.517	
		D.B.O.E.	7.13	7.17	13	0.007	0.011	0.506	0.852	
(3)	51,768	S.B.	28.72*	27.85*	16			0.508		Crouchley and Pickles (1984)
		S.B.O.E.	27.03*	26.48*	15	0.002		0.509		
		D.B.	24.75*	24.23*	14		0.030	0.502	0.679	
		D.B.O.E.	24.83*	24.40*	13	0.0001	0.318	0.487	0.552	
USA	116,458	S.B.	54.05*	54.23*	31			0.514		Greenberg and White (1967)
		S.B.O.E.	53.97*	54.16*	30	0.0001		0.515		
		D.B.	53.89*	54.09*	29		0.999	0.155	0.514	
		D.B.O.E.	53.97*	54.10*	28	0.0001	0.909	0.515	0.515	

D.B.O.E. - Double binomial with one-tail excess; D.B. - double binomial; S.B.O.E. - simple binomial with one-tail excess; S.B. - simple binomial.

*Significant at the 5% level.

For other abbreviations see Table II.

models fits and b) the rest, for which at least the double binomial model, does fit (Table III).

Assuming that this last group of samples represents a real phenomenon, and that Poisson (variability within families, as parity effects, for instance) and Markovian (prior event dependency) variations are relatively unimportant, it might be possible to find biological interpretations of the statistical parameters. The simple binomial with one-tail excess is clear in the sense that a proportion T of couples is unable to segregate males, in a manner analogous to the sex ratio trait in *Drosophila*. Nevertheless, although this model fits most of the data from this group of samples, and although it is the most parsimonious, there are nonetheless four samples (France (1), France (2), Finland (1), Finland (2)), for which the three-parameter model fits better than the simple binomial with one-tail excess; Table III).

Thus, the double binomial was found to be the most suitable model, for expressing familial distribution of sex. An overall analysis of the double binomial model, with certain constraints, was applied to the six samples that fit the model (Table IV). A common value of w (the frequency of couples with high sex ratio) for all samples was estimated (since it would be expected that no substantial difference existed between large populations) and concomitantly forcing $m = p/(2 - p)$, i.e., assuming that the low sex ratio group is composed of X-linked lethal carrier couples. The absence of significant departures from the model for all six samples, together with no evidence of significant differences between the former analysis (with three parameters) and the ones with the constraint (two parameters, w and p) (Table IV, last column) supports

the hypothesis that there is a mixture of two types of families, one of them segregating X-linked lethals.

If this interpretation is correct, it is possible to infer that, aside from several other possible causes of departures from binomial expectations (cf. James, 1987), around 10% of females in the majority of human populations are carriers of X-linked lethals ($1 - w$), which is higher than would be expected if one takes into account the normal mutation rate and the proportion of the genome represented by X chromosome genes. The hypothesis suggested by this analysis is that, on average, important X-linked genes are longer than their autosomal counterpart, explaining, therefore, an apparent higher mutation rate. Also, these analyses allow the inference that the sex ratio prior to the action of these lethals is around 0.55, which poses an interesting question on both the primary sex ratio and on the true sex ratio of spontaneous abortions (Hassold *et al.*, 1980; James, 1987; Morton *et al.*, 1988; Bartels *et al.*, 1990). Finally, other models (such as incompatibility against male antigens, in particular families) could also be applied to these data and provide a different insight on this subjects. However, probably a more general model that involves most of these causes (and as shown here, the importance of X-linked recessive lethals cannot be neglected) is more suitable to apply to large samples in order to explain almost all the among-family variability of sex ratio.

ACKNOWLEDGMENTS

The authors are grateful to the ECLAMC Program (Dr. E.E. Castilla) for making these data available. This study was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Financiadora de Estudos e Projetos (FINEP), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). They also acknowledge some very pertinent suggestions made by a referee.

Publication supported by FAPESP.

RESUMO

Uma amostra de 59.496 indivíduos latino-americanos pertencentes a 28.545 irmandades foi tomada de um grande Estudo Colaborativo de Malformações Congênitas (ECLAMC). Recém-nascidos com malformações congênitas e seus respectivos controles foram excluídos da amostra com a finalidade de evitar vieses associados com malformações. Alguns modelos estatísticos foram aplicados, admitindo-se que os efeitos de Poisson e de Markovian não são importantes na determinação da distribuição geral das irmandades na

Table IV - Goodness of fit tests of segregation analyses of a model that forces the segregation of an X-linked lethal, in several samples.

Sample	$w = 0.89$		$w = 0.91$		Difference between forced lethal and double binomial	
	χ^2	d.f.	χ^2	d.f.	χ^2	d.f.
Present	22.95	16	23.91	16	1.21	1
France (1)	27.02	24	27.18	24	0.21	1
France (2)	60.50	75	61.49	75	0.67	1
Sweden	32.10	33	27.80	33	0.12	1
Finland (1)	13.51	10	17.50	10	0.01	1
Finland (2)	16.72	16	20.73	16	3.15	1

The first two columns refer to the double binomial, with the correspondent value of w and with the constraint: $m = p/(2 - p)$. The last column is the difference between the double binomial (with the constraint: $m = p/(2 - p)$) and the double binomial with no constraints.

população. O modelo geral foi o binomial duplo com excesso unicaudal. Este modelo pode ser simplificado para o binomial duplo, o binomial simples com excesso unicaudal ou o binomial simples. As primeiras três distribuições ajustaram-se bem aos dados ($\chi^2_{15} = 16,46, P = 0,35$; $\chi^2_{16} = 20,92, P = 0,18$; $\chi^2_{17} = 19,88, P = 0,28$, respectivamente), enquanto que o binomial simples não se ajustou ($\chi^2_{18} = 35,01, P = 0,009$). Os mesmos modelos foram aplicados a vários estudos realizados em diferentes países e em épocas diferentes, os quais apresentaram resultados semelhantes. O modelo mais parcimonioso que se ajustou à maioria dos conjuntos de dados foi o binomial duplo, sugerindo que cerca de 9% dos casais segregam letais recessivos ligados ao X, uma vez que as duas proporções de segregação do modelo estão próximas da previsão genética.

REFERENCES

- Bartels, I., Hansmann, I. and Eiben, B.** (1990). Excess of females in chromosomally normal spontaneous abortions. *Am. J. Med. Genet.* 35: 297-298.
- Castilla, E.E. and Orioli, I.M.** (1983). El Estudio Colaborativo Latinoamericano de Malformaciones Congénitas: ECLAMC/Monitor. *Interciência* 8: 271-278.
- Crouchley, R. and Pickles, A.R.** (1984). Methods for the identification of Lexian, Poisson and Markovian variations in the secondary sex ratio. *Biometrics* 40: 165-175.
- Edwards, A.W.F.** (1958). An analysis of Geissler's data on the human sex ratio. *Ann. Hum. Genet.* 23: 6-15.
- Edwards, A.W.F.** (1959). Some comments on Schutzenberger's analysis of data on the human sex ratio. *Ann. Hum. Genet.* 23: 233-238.
- Edwards, A.W.F.** (1962). Genetics and the human sex ratio. *Adv. Genet.* 11: 239-272.
- Edwards, A.W.F. and Fraccaro, M.** (1960). Distribution and sequences of sexes in a selected sample of Swedish families. *Ann. Hum. Genet.* 24: 245-252.
- Feitosa, M.F. and Krieger, H.** (1992). Demography of the human sex ratio in some Latin American countries, 1967-1986. *Hum. Biol.* 64: 523-530.
- Fisher, R.A.** (1958). *The Genetical Theory of Natural Selection*. 2nd edn. Dover Publications, New York, pp. 158-162.
- Fletcher, R. and Powell, M.J.D.** (1963). A rapidly convergent descent method for minimization. *Comput. J.* 6: 163-168.
- Geissler, A.** (1889). Beitrage zur Frage des Geschlechtsverhältnisses der Geborenen. *Z. K. Sachs. Stat. Bureaus* 35: 1-24.
- Gini, C.** (1908). *Il Sesso dal Punto di Vista Statistico*. 1st edn. Milano-Palermo-Napoli, Sandrom, pp. 380.
- Greenberg, R.A. and White, C.** (1967). The sexes of consecutive sibs in human sibships. *Hum. Biol.* 39: 374-404.
- Harris, J.A. and Gunstad, B.** (1930a). The correlation between the sex of human siblings. *Genetics* 15: 445-461.
- Harris, J.A. and Gunstad, B.** (1930b). The problem of the relationship between the number and the sex of human offspring. *Am. Naturalist* 64: 495-508.
- Hassold, T., Chen, N., Funkhouser, J., Jooss, T., Manuel, B., Matsuura, J., Matsuyama, A., Wilson, C., Yamane, J.A. and Jacobs, P.A.** (1980). A cytogenetic study of 1000 spontaneous abortions. *Ann. Hum. Genet.* 44: 151-178.
- James, W.H.** (1987). The human sex ratio. Part 1: A review of the literature. *Hum. Biol.* 59: 721-752.
- Lancaster, H.O.** (1950). The sex ratio in sibships, with special reference to Geissler's data. *Ann. Eugen.* 15: 153-158.
- Morton, N.E., Jacobs, P.A., Hassold, T. and Wu, D.** (1988). Maternal age in trisomy. *Ann. Hum. Genet.* 52: 227-235.
- Renkonen, K.O.** (1956). Is the sex ratio between boys and girls correlated to the sex of precedent children? *Ann. Med. Exp. Fenn.* 34: 447-451.
- Renkonen, K.O., Makela, O. and Lehtovaara, R.** (1961). Factors affecting the human sex ratio. *Ann. Med. Exper. Fenn.* 39: 173-184.
- Schutzenberger, M.P.** (1949). Résultats d'une enquête sur la distribution du sexe dans les familles nombreuses. *Sem. Hôp.* 25: 2579-2582.
- Slater, E.** (1944). A demographic study of a psychopatic population. *Ann. Eugen.* 12: 121-137.
- Stern, C.** (1960). *Principles of Human Genetics*. 2nd edn. W.H. Freeman & Co., San Francisco, pp. 424.
- Waalder, G.H.M.** (1928). Om seksualproportionen ved fodslen. *Nord. Statist. Tidsskr.* 7: 65-84.

(Received April 24, 1995)