The Estimated Probability of Dizygotic Twins: A Comparison of Two Methods

Jill Hardin,[1,2] Steve Selvin,[2] Suzan L. Carmichael,[1] and Gary M. Shaw[1,2]

[1] March of Dimes, California Research Division, Children's Hospital Oakland Research Institute, Oakland, California, United States of America
[2] University of California, Berkeley, School of Public Health, Berkeley, California, United States of America

This study presents a general model of two binary variables and applies it to twin sex pairing data from 21 twin data sources to estimate the frequency of dizygotic twins. The purpose of this study is to clarify the relationship between maximum likelihood and Weinberg's differential rule zygosity estimation methods. We explore the accuracy of these zygosity estimation measures in relation to twin ascertainment methods and the probability of a male. Twin sex pairing data from 21 twin data sources representing 15 countries was collected for use in this study. Maximum likelihood estimation of the probability of dizygotic twins is applied to describe the variation in the frequency of dizygotic twin births. The differences between maximum likelihood and Weinberg's differential rule zygosity estimation methods are presented as a function of twin data ascertainment method and the probability of a male. Maximum likelihood estimation of the probability of dizygotic twins ranges from 0.083 (95% approximate CI: 0.082, 0.085) to 0.750 (95% approximate CI: 0.749, 0.752) for voluntary ascertainment data sources and from 0.374 (95% approximate CI: 0.373, 0.375) to 0.987 (95% approximate CI: 0.959, 1.016) for active ascertainment data sources. In 17 of the 21 twin data sources differences of 0.01 or less occur between maximum likelihood and Weinberg zygosity estimation methods. The Weinberg and maximum likelihood estimates are negligibly different in most applications. Using the above general maximum likelihood estimate, the probability of a dizygotic twin is subject to substantial variation that is largely a function of twin data ascertainment method.

Keywords: zygosity estimation methodology, probability of a male, maximum likelihood estimation, Weinberg's differential rule

In twin studies examination of concordance rates between zygosity groups is used as a means to identify the contributions of genetic and environmental factors to the development of a given condition. Where twin zygosity has not been established through biological means, accurate estimation of zygosity is critical. The primary methods used for zygosity estimation are the maximum likelihood procedure and Weinberg's differential rule. While there have been many publications regarding the validity of Weinberg estimation methods (Allen, 1981; Bulmer, 1976; Fellman & Eriksson, 2006; Husby et al., 1991; James, 1976; James, 1979; James, 1992; Orlebeke et al., 1991; Vlietinck et al., 1988) a lack of clarity remains in the relationship between the maximum likelihood procedure and Weinberg zygosity estimation methods.

The primary goal of this study is to clarify the relationship between maximum likelihood and Weinberg zygosity estimation methods through presentation of a general description of binary variables that yields estimates of the frequency of dizygotic twins. Variance formulas are presented that are substantially simplified in comparison to those recently published (Fellman & Eriksson, 2006). A formula quantifying the difference in estimates obtained using the maximum likelihood and Weinberg zygosity estimation methods is presented. We employed twin sex pairing data from 21 data sources to explore comparisons between zygosity estimation methods.

Methods

Data Sources

Twenty of the twin data sources were abstracted from publications of international twin registries (Table 1) while one twin data source was derived from California vital statistics records. Publications were included based on a review of the twin registry literature and were validated against electronic twin resources (Twin Registers and Research, 2007b) that enumerate major international twin registries. In selection of publications for inclusion in this study voluntary ascertainment was defined as recruitment of twins through convenience sampling while active ascertainment was defined as recruitment of twins at birth or through birth registers. The twin registries included in this analysis encompass a broad range of birth years, ranging from 1870 to 2005.

Jill Hardin, Steve Selvin, Suzan L. Carmichael, and Gary M. Shaw

Description of Binary Variables and Maximum Likelihood Estimation

The general properties and description of two binary variables applies directly to the estimation of the frequency of dizygotic twins. Thus, maximum likelihood procedures produce optimally efficient estimates of the probability of dizygotic twins and estimates of their associated variance. In addition, the classic Weinberg estimate becomes a special case and, therefore, its properties also follow from the maximum likelihood process.

Estimation of the probability that a twin pair is dizygotic (monozygotic) is derived from the general relationship between two binary variables denoted X and X′, where:

	X = 1	X = 0	
X′ = 1	$p^2 + \varepsilon$	$pq - \varepsilon$	p
X′ = 0	$pq - \varepsilon$	$q^2 + \varepsilon$	q
	p	q	1

The symbol p = probability (X = 1) = probability (X′ = 1) and $q = 1 - p$ = probability (X = 0) = probability (X′ = 0). The symbol ε represents the covariance between the two binary variables X and X′. Therefore, the correlation coefficient measuring the degree of association between X and X′ is given by:

$$f = \frac{\varepsilon}{pq}$$

and is an application to binary data of the classic Pearson correlation coefficient calculated between two variables where:

$$\text{correlation: } f = \frac{covariance(X, X')}{\sqrt{variance(X)}\sqrt{variance(X')}} = \frac{\varepsilon}{pq}$$

In the case of twin data, the correlation coefficient f indicates the proportion of monozygotic pairs among like-sex twin births. The correlation coefficient f, as always varies between –1 and 1 and when $f = 0$ ($\varepsilon = 0$) the variables X and X' are independent. The maximum likelihood estimates for the two parameters, p and f, denoted by $\hat{p}$ and $\hat{f}$, are:

$$\hat{p} = \frac{2a+b}{2n} \quad \text{and} \quad \hat{f} = 1 - \frac{b}{2n\hat{p}\hat{q}}$$

The symbol a = number of observed pairs (X = 1 and X′ = 1), b = number of observed pairs (X = 0 and X′ = 1) or (X = 1 and X′ = 0), c = number of observed pairs (X = 0 and X′ = 0), yielding a total number of pairs $n = a + b + c$.

An estimate of the variance of the distribution of the estimate of $\hat{p}$ is given by:

$$variance(\hat{p}) = \frac{2\hat{p}\hat{q}(1+\hat{f})}{2n}$$

and for the estimate $\hat{f}$, an estimate of the variance of its distribution is given by:

$$variance(\hat{f}) = \frac{(1-\hat{f})\left[1 - 2\hat{p}\hat{q}(1-\hat{f}) - \left(1 - 4\hat{p}\hat{q}(1-\hat{f})^2\right)\right]}{2n\hat{p}\hat{q}} .$$

A number of applications of the maximum likelihood estimate of $\hat{f}$ and its associated variance are important, particularly in genetics. The estimate $\hat{f}$ is sometimes referred to as the 'phi-correlation coefficient' and Wright (Wright, 1922) defined $\hat{f}$ as an estimate of the correlation between two uniting gametes (inbreeding coefficient) where p = frequency of a specific gamete and again $q = 1 - p$. In another application drawn from genetics, when p represents an allele frequency (Thomas, 2004) the parameter ε measures the degree of Hardy-Weinberg disequilibrium among a sample of observed genotypes. In fact, the quantity

$$X^2 = n\hat{f}^2$$

is the classic chi-square statistical criterion for measuring the likelihood that the observed deviance from independence arose by chance.

An application to twin data produces an estimate of the probability of a dizygotic twin (denoted $\hat{d} = 1 - \hat{f}$) and its variance. That is, when the probability of a dizygotic twin is the same regardless of the composition of the twin pair (like or unlike sex) and the sex composition of the dizygotic twin pair is determined by two independent events, the estimated probability of a dizygotic twin is $\hat{d}$.

Specifically, for twin data where a represents the observed number of male-male pairs, b represents the observed number of male-female pairs, and c represents the observed number of female-female pairs, estimates from twin data are given, as:

$$\text{estimated probability of male twin} = \hat{p} = \frac{2a+b}{2n}$$

where $n = a + b + c$ and

$$\text{estimated probability of dizygotic twin} = \hat{d} = \frac{b}{2n\hat{p}\hat{q}} .$$

When the probability of a male twin is 0.5 ($p = 0.5$), these expressions produce the classic Weinberg estimate of the probability of a dizygotic twin and its variance where:

$$\hat{d}' = \frac{2b}{n}$$

The variance of the distribution of either zygosity estimate is:

$$variance(\hat{d}') = variance(\hat{m}') = \frac{1 - \hat{m}'^2}{n}$$

where $\hat{m}' = \hat{d}'$.

That is, the Weinberg estimate is the special case of the maximum likelihood estimate when the value of p is replaced by 0.5. The difference between the

maximum likelihood and Weinberg estimates is extremely small when p is in the neighborhood of 0.5. The difference is given by:

$$\left|\hat{d} - \hat{d}'\right| = \hat{d}\left|4\hat{p}\hat{q} - 1\right|$$

When the probability of a male varies between 0.45 and 0.55, the difference in the two estimates is less than 0.01.

Statistical Methods

Estimated probabilities of dizygotic twins and 95% approximate confidence intervals were generated using the maximum likelihood method and applied to 21 twin data sources (Table 1 and Figure 1). Differences in maximum likelihood and Weinberg zygosity estimation methods were calculated for each twin data source and are presented as a function of the probability of a male in Figure 2.
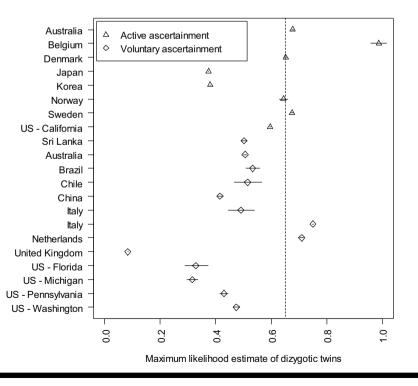
Results

Table 1 presents an overview of 21 international twin data sources. Thirteen of the sources obtained twins by voluntary ascertainment, while eight sources obtained twins by active ascertainment. In these data the frequency of males varies substantially depending on the ascertainment method, which influences the estimates of the probability of dizygotic twins. Twin data sources that used active ascertainment methods have a probability of a male that ranges from 0.476 to 0.516 with four out of the eight sources having a probability

of a male of essentially 0.51. In contrast, data sources that contained twins based on voluntary collection methods have a much broader range in the probability of a male, 0.165 to 0.543.

Also presented in Table 1 are estimates of the probability of dizygotic twins and their associated approximate confidence intervals derived from the maximum likelihood procedure. The estimate of the probability of dizygotic twins, $\hat{d}$ ranges from 0.083 (p = 0.165) to 0.987 (p = 0.514). The estimates of the probability of a dizygotic twin, $\hat{d}$ ranges from 0.374 to 0.987 for active ascertainment data sources and from 0.083 to 0.750 for voluntary ascertainment data sources. The probability of a male, p range from 0.477 to 0.516 but estimates of the probability of dizygotic twins, $\hat{d}$ spans a much broader range of 0.374 to 0.987 for the active ascertainment data sources. This illustrates that the distribution of males among twin pairs is crucially important in maximum likelihood zygosity estimation.

Figure 1 presents estimates of the probability of dizygotic twins and the associated approximate confidence intervals for each dataset. Bulmer reported (Bulmer, 1970) that the frequency of dizygotic twins varies with race-ethnicity, but on average dizygotic twins constitute about 65% of all twins. Therefore in Figure 1, the vertical line at 0.65 represents the probability of dizygotic twins across all race-ethnicity groups. Estimates of the difference in the estimation of the probability of dizygotic twins using maximum likelihood and Weinberg methods as a function of the
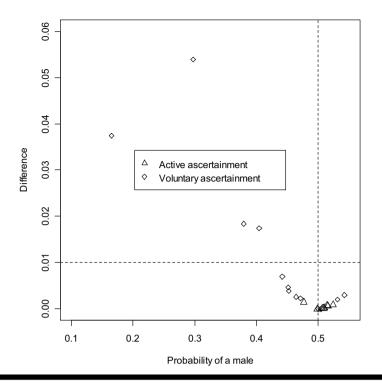

Figure 1

Maximum likelihood estimates of the probability of dizygotic twins and 95% approximate confidence intervals for 21 twin data sources.

Note: Dashed vertical line at 0.65 represents the literature reported (Bulmer, 1970) probability of dizygotic twins.

Table 1

Probability of Dizygotic Twins and 95% Approximate Confidence Intervals (CI) Using Maximum Likelihood Estimation

Reference	Year(s) of birth cohort	Country	Type	Twin pairs Male–male	Male–female	Female–female	Male probability[1]	Dizygotic twin probability	95% CI[2]
(Brackenridge, 1977)	1920–1969	Australia	Active	31,213	31,372	30,042	0.506	0.678	0.673, 0.682
(Derom et al., 2006)	1964–2005	East Flanders, Belgium	Active	1,208	2,224	1,078	0.514	0.987	0.959, 1.016
(Skythe et al., 2002)	1870–1996	Denmark	Active	23,872	22,008	21,738	0.516	0.652	0.647, 0.656
(Imaizumi & Nonaka, 1997)	1975–1994	Japan	Active	81,456	36,653	77,992	0.509	0.374	0.373, 0.375
(Sung et al., 2002)	1901–1999	Korea	Active	66,433	29,347	59,003	0.524	0.380	0.379, 0.382
(Harris et al., 2002)	1967–1979	Norway	Active	1,854	1,883	2,126	0.477	0.644	0.629, 0.659
(Lichtenstein et al., 2006)	1886–1958, 1959–1986, 1987–1992, 1992–2000	Sweden	Active	29,030	28,686	27,372	0.510	0.675	0.670, 0.679
Not applicable	1983–2003	USA — California	Active	44,678	38,098	44,990	0.499	0.596	0.593, 0.599
(2007a)	Unspecified	Sri Lanka	Voluntary	2,212	1,630	2,679	0.464	0.503	0.492, 0.513
(Hopper, 2002)	Various	Australia	Voluntary	8,998	6,907	11,670	0.452	0.506	0.501, 0.511
(Beiguelman et al., 1995)	1984–1993	Brazil	Voluntary	514	369	502	0.504	0.533	0.509, 0.558
(Ivanovic et al., 2006)	Unspecified	Chile	Voluntary	103	83	141	0.442	0.515	0.468, 0.566
(Li et al., 2006)	< 1929–2000	China	Voluntary	3,240	1,519	2,614	0.543	0.415	0.407, 0.423
(Lanni et al., 1994)	April–November 1993	Italy	Voluntary	124	74	105	0.531	0.490	0.445, 0.540
(Stazi et al., 2002)	1900–1995	Italy	Voluntary	207,262	248,893	207,259	0.500	0.750	0.749, 0.752
(Boomsma et al., 2006)	[3]1940–1987 [4]1986–2001	Netherlands	Voluntary	9,280	11,150	11,084	0.471	0.710	0.703, 0.718
(Spector & Williams, 2006)	1900–2000	United Kingdom	Voluntary	788	118	4,222	0.165	0.083	0.082, 0.085
(Taylor et al., 2006)	2002	USA — Florida	Voluntary	35	21	97	0.297	0.328	0.290, 0.373
(Coccaro & Jacobson, 2006)	1959–1978	USA — Michigan	Voluntary	164	80	294	0.379	0.316	0.297, 0.336
(Klump & Burt, 2006)	Unspecified	USA — Pennsylvania	Voluntary	725	446	924	0.453	0.430	0.415, 0.445
(Afari et al., 2006)	Unspecified	USA—Washington	Voluntary	1,248	984	2,072	0.404	0.475	0.463, 0.487

Note: [1] Probability of a male calculated as follows: [2*male–male + male–female]/[2*(male–male + male–female + female–female)]
[2] Confidence intervals are approximate.
[3] Adolescent and young adult twins (ANTR) recruited through city councils in 1990 and 1991.
[4] Young twins (YNTR) registered at birth by their parents who are approached through "birth felicitation" services.

Figure 2

Difference in maximum likelihood and Weinberg estimates of the probability of dizygotic twins as a function of the probability of a male for 21 twin data sources.

Note: The dashed vertical line at 0.5 represents the usual probability of a male. The dashed horizontal line at 0.01 illustrates that 17 out of 21 twin data sources have differences of <0.01.

probability of a male are presented in Figure 2. The figure illustrates that differences of 0.01 or less occur in 17 out of the 21 twin data sources, when the probability of a male falls between 0.45 and 0.55, indicating both methods result in almost identical zygosity estimates when the probability of a male is close to 0.5.

Discussion

This study uses a general model describing the statistical structure of binary traits to illustrate the estimation of the frequency of dizygotic twins using the maximum likelihood estimation procedure. The properties of Weinberg's differential rule follow directly from the maximum likelihood process. We present an intuitive general model that can be used to describe any binary trait and is useful for estimation of the probability of zygosity. We further illustrate that the probability of a male may deviate from 0.5 when ascertainment has resulted in selective samples (Table 1), which affects the agreement between the maximum likelihood procedure and the Weinberg's differential rule zygosity estimation methods. Our study shows that it would require large shifts away from the expected probability of male from 0.5 for substantial differences between these two zygosity estimation methods to exist.

Several investigators (Allen, 1981; James, 1971; James, 1976; James, 1979; James, 1992; Orlebeke et al., 1991) have suggested that the underlying assumptions of Weinberg's differential rule are often violated. These studies suggest that a lack of independence exists between zygotes in dizygotic pregnancies and that the probability of a male also differs depending on zygosity. Genetic theory assumes that dizygotic twins are as similar genetically as non-twin siblings, which implies independence between zygotes. The probability of a male in a singleton birth is universally close to 0.5 and therefore the probability of a male in dizygotic twins is similarly expected to be close to 0.5. Without persuasive evidence to the contrary, the binary variable model provides an accurate basis for useful estimates of zygosity. Furthermore, the robustness of Weinberg's rule has been illustrated through comparisons between Weinberg's estimates and known zygosity of observed data (Vlietinck et al., 1988; Husby et al., 1991), through more theoretical discussions (Bulmer, 1976), and recently through comparisons of estimates derived using a maximum likelihood function (Fellman & Eriksson, 2006).

We recommend that researchers examine the proposed general model describing the statistical structure of binary traits in estimation of zygosity. We further suggest that researchers carefully examine the ascertainment methods employed for recruitment of twins and the associated frequency of male twins, as they may substantially influence the accuracy of zygosity estimation methods.

Jill Hardin, Steve Selvin, Suzan L. Carmichael, and Gary M. Shaw

References

Sri Lankan Twin Registry (2007a). Available at: http://infolanka.com/org/twin-registry/Statistics.htm

Twin Registers and Research (2007b). Available at: http://www.ists.qimr.edu.au/links.html#Twin%20reg

Afari, N., Noonan, C., Goldberg, J., Edwards, K., Gadepalli, K., Osterman, B., Evanoff, C., & Buchwald, D. (2006). University of Washington Twin Registry: Construction and characteristics of a community-based twin registry. *Twin Research and Human Genetics, 9,* 1023–1029.

Allen, G. (1981). Errors of Weinberg's difference method. *Progress in Clinical and Biological Research, 69A,* 71–74.

Beiguelman, B., Franchi-Pinto, C., Dal Colletto, G. M., & Krieger, H. (1995). Annual variation of sex ratio in twin births and in singletons in Brazil. *Acta Geneticae Medicae et Gemellologiae, 44,* 163–168.

Boomsma, D. I., de Geus, E. J., Vink, J. M., Stubbe, J. H., Distel, M. A., Hottenga, J. J., Posthuma, D., van Beijsterveldt, T. C., Hudziak, J. J., Bartels, M., & Willemsen, G. (2006). Netherlands Twin Register: From twins to twin families. *Twin Research and Human Genetics, 9,* 849–857.

Brackenridge, C. J. (1977). The secular variation of Australian twin births over fifty years. *Annals of Human Biology, 4,* 559–564.

Bulmer, M. G. (1976). Is Weinberg's method valid? *Acta Geneticae Medicae et Gemellologiae, 25,* 25–28.

Bulmer, M. (1970). *The biology of twinning in man.* Oxford: Clarendon Press.

Coccaro, E. F., & Jacobson, K. C. (2006). PennTwins: A population-based cohort for twin studies. *Twin Research and Human Genetics, 9,* 998–1005.

Derom, C. A., Vlietinck, R. F., Thiery, E. W., Leroy, F. O., Fryns, J. P., & Derom, R. M. (2006). The East Flanders Prospective Twin Survey (EFPTS). *Twin Research and Human Genetics, 9,* 733–738.

Fellman, J., & Eriksson, A. W. (2006). Weinberg's differential rule reconsidered. *Human Biology, 78,* 253–275.

Harris, J. R., Magnus, P., & Tambs, K. (2002). The Norwegian Institute of Public Health Twin Panel: A description of the sample and program of research. *Twin Research, 5,* 415–423.

Hopper, J. L. (2002). *The Australian Twin Registry. Twin Research, 5,* 329–336.

Husby, H., Holm, N. V., Gernow, A., Thomsen, S. G., Kock, K., & Gurtler, H. (1991). Zygosity, placental membranes and Weinberg's rule in a Danish consecutive twin series. *Acta Geneticae Medicae et Gemellologiae, 40,* 147–152.

Imaizumi, Y. & Nonaka, K. (1997). The twinning rates by zygosity in Japan, 1975–1994. *Acta Geneticae Medicae et Gemellologiae, 46,* 9–22.

Ivanovic, D. M., Llop, E. R., Alvear, J. A., Perez, H. T., Diaz, N. S., Leyton, B. D., Almagià, A. F., Larraín, C. G., Alvarez, J. E., Herrera, Y. F., & Hazbun, E. L. (2006). Chilean school-age children twin registry: Planning, sampling and implications. *Twin Research and Human Genetics, 9,* 743–746.

James, W. H. (1971). Excess of like sexed pairs of dizygotic twins. *Nature, 232,* 277–278.

James, W. H. (1976). The possibility of a flaw underlying Weinberg's differential rule. *Annals of Human Genetics, 40,* 197–199.

James, W. H. (1979). Is Weinberg's differential rule valid? *Acta Geneticae Medicae et Gemellologiae, 28,* 69–71.

James, W. H. (1992). The current status of Weinberg's differential rule. *Acta Geneticae Medicae et Gemellologiae, 41,* 33–42.

Klump, K. L. & Burt, S. A. (2006). The Michigan State University Twin Registry (MSUTR): genetic, environmental and neurobiological influences on behavior across development. *Twin Research and Human Genetics, 9,* 971–977.

Lanni, R., Grimaldi, V., Corchia, C., & Mastroiacovo, P. (1994). The Mercurio Project and Twins. How to use a registry. *Acta Geneticae Medicae et Gemellologiae, 43,* 95–103.

Li, L., Gao, W., Lv, J., Cao, W., Zhan, S., Yang, H., & Hu, Y. (2006). Current status of the Chinese National Twin Registry. *Twin Research and Human Genetics, 9,* 747–752.

Lichtenstein, P., Sullivan, P. F., Cnattingius, S., Gatz, M., Johansson, S., Carlstrom, E. Björk, C., Svartengren, M., Wolk, A., Klareskog, L., de Faire, U., Schalling, M., Palmgren, J., Pedersen, N. L. (2006). The Swedish Twin Registry in the third millennium: an update. *Twin Research and Human Genetics, 9,* 875–882.

Orlebeke, J. F., Eriksson, A. W., Boomsma, D. I., Vlietinck, R., Tas, F. J., & de Geus, E. C. (1991). Changes in the DZ unlike/like sex ratio in The Netherlands. *Acta Geneticae Medicae Gemellologiae, 40,* 319–323.

Skytthe, A., Kyvik, K., Holm, N. V., Vaupel, J. W., & Christensen, K. (2002). The Danish Twin Registry: 127 birth cohorts of twins. *Twin Research, 5,* 352–357.

Spector, T. D. & Williams, F. M. (2006). The UK Adult Twin Registry (TwinsUK). *Twin Research and Human Genetics, 9,* 899–906.

Stazi, M. A., Cotichini, R., Patriarca, V., Brescianini, S., Fagnani, C., D'Ippolito, C., Cannoni, S., Ristori, G., & Salvetti, M. (2002). The Italian Twin Project: from the personal identification number to a national twin registry. *Twin Research, 5,* 382–386.

Sung, J., Cho, S. H., Cho, S. I., Duffy, D. L., Kim, J. H., Kim, H., Park, K. S., & Park, S. K. (2002). The Korean Twin Registry: Methods, current stage, and interim results. *Twin Research, 5,* 394–400.

Taylor, J. E., James, L. M., Reeves, M. D., & Bobadilla, L. (2006). The Florida State Twin Registry: Research aims and design. *Twin Research and Human Genetics, 9,* 958–962.

Thomas, D. C. (2004). *Statistical Methods in Genetic Epidemiology.* Oxford: Oxford University Press.

Vlietinck, R., Derom, C., Derom, R., Van den, B. H., & Thiery, M. (1988). The validity of Weinberg's rule in the East Flanders Prospective Twin Survey (EFPTS). *Acta Geneticae Medicae et Gemellologiae, 37,* 137–141.

Wright, S. (1922). Coefficients of inbreeding and relationship. *American Naturalist, 56,* 330–338.