American Genetic Association

Journal of Heredity, 2017, 1–9
doi:10.1093/jhered/esx032
Original Article

OXFORD

Original Article

Estimating Inbreeding Rates in Natural Populations: Addressing the Problem of Incomplete Pedigrees

Mark P. Miller, Susan M. Haig, Jonathan D. Ballou, and E. Ashley Steel

From the US Geological Survey Forest and Rangeland Ecosystem Science Center, 3200 SW Jefferson Way, Corvallis, OR 97331 (Miller and Haig); Smithsonian National Zoological Park, Conservation Biology Institute, Front Royal, VA (Ballou); and USDA Forest Service Pacific Northwest Research Station, Seattle, WA (Steel).

Address correpondence to M. P. Miller at the address above, or e-mail: mpmiller@usgs.gov.

Received November 8, 2016; First decision January 4, 2017; Accepted April 5, 2017.

Corresponding Editor: Oliver Ryder

Abstract

Understanding and estimating inbreeding is essential for managing threatened and endangered wildlife populations. However, determination of inbreeding rates in natural populations is confounded by incomplete parentage information. We present an approach for quantifying inbreeding rates for populations with incomplete parentage information. The approach exploits knowledge of pedigree configurations that lead to inbreeding coefficients of $F = 0.25$ and $F = 0.125$, allowing for quantification of $\Pr(I|k)$: the probability of observing pedigree I given the fraction of known parents (k). We developed analytical expressions under simplifying assumptions that define properties and behavior of inbreeding rate estimators for varying values of k. We demonstrated that inbreeding is overestimated if $\Pr(I|k)$ is not taken into consideration and that bias is primarily influenced by k. By contrast, our new estimator, incorporating $\Pr(I|k)$, is unbiased over a wide range of values of k that may be observed in empirical studies. Stochastic computer simulations that allowed complex inter- and intragenerational inbreeding produced similar results. We illustrate the effects that accounting for $\Pr(I|k)$ can have in empirical data by revisiting published analyses of Arabian oryx (*Oryx leucoryx*) and Red deer (*Cervus elaphus*). Our results demonstrate that incomplete pedigrees are not barriers for quantifying inbreeding in wild populations. Application of our approach will permit a better understanding of the role that inbreeding plays in the dynamics of populations of threatened and endangered species and may help refine our understanding of inbreeding avoidance mechanisms in the wild.

Subject areas: Reproductive strategies and kinship analysis; Conservation genetics and biodiversity
Keywords: coefficient, inbreeding conservation, management, mating between related individuals, natural populations, pedigree

Inbreeding and its consequences, or the avoidance thereof, play an important role in ecological and evolutionary processes. Progeny of related individuals often are at selective disadvantages within populations (Ralls et al. 1988; Crnokrak and Roff 1999; Keller and Waller 2002; but see Ballou 1997), leading to the evolution of strategies to minimize inbreeding in the wild (Pusey and Wolf 1996). Although inbreeding avoidance is common, inbreeding may be inevitable in small or isolated populations (Ballou 1995). Thus, it is important to consider the consequences of inbreeding when managing threatened and endangered species to better understand the role that it may play in population declines (Ralls et al. 1979; Lande 1988; Hedrick and Kalinowski 2000; Frankham 2005; O'Grady et al. 2006).

Inbreeding is most clearly documented when an individual's pedigree is known (Pemberton 2004). Whether inferred through direct observation or via genetic analyses, complete knowledge of an individual's pedigree may be difficult or impossible to obtain (Pemberton 2008), especially in species capable of long distance movement, inhabiting large geographic ranges, demonstrating high reproductive rates, or where parental care is minimal. Consequently, most pedigrees obtained in natural populations are incomplete. It has long been recognized that incomplete pedigrees complicate attempts to quantify inbreeding rates in the wild (Howard 1949; Bulmer 1973; Van Noordwijk and Scharloo 1981). In this article, we outline an approach for estimating inbreeding rates in wild populations. Derivation of this approach will facilitate future analyses that have applications for management of threatened and endangered wild populations or in any wild population of management concern.

Estimating Inbreeding Rates from Incomplete Pedigrees

In a population where the pedigrees of all individuals are known, the frequency of inbreeding (f) associated with a specific type of mating (i.e., parent-offspring, full-siblings, half siblings, etc) can be obtained as

$$f = o / n, \tag{1}$$

where n is the number of individuals examined and o is the observed number of inbred individuals produced by the parental pairings of interest. When information about the parentage of individuals is missing, this quantity will underestimate the true inbreeding rate because not all of the o inbred individuals will be identified as such. To address this issue, Marshall et al. (2002) suggested an approach for estimating inbreeding when there are incomplete pedigrees. They itemized the specific ancestors required to potentially identify inbreeding events for three types of inbreeding associated with $F = 0.25$ and 11 types of inbreeding associated with $F = 0.125$ (Table 1), where F is the inbreeding coefficient that quantifies both the probability that an individual possesses 2 alleles at a locus that are identical by descent and the severity of an inbreeding event (Ballou 1983). Lower values of F were not considered due to the large number and complexity of pedigrees capable of producing values of F less than 0.125. The estimator was verbally defined as follows: "In this analysis, we categorized inbreeding events … and counted the number of offspring born for whom each type of inbreeding event could have been detected." Based on this statement, we formally define the Marshall et al. estimator for category i as

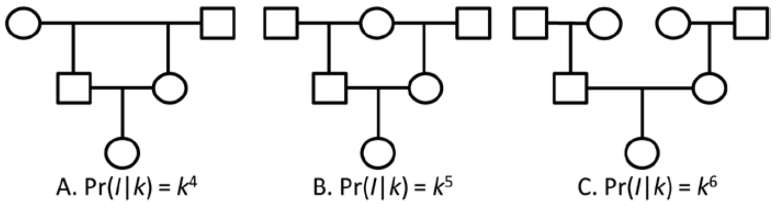
$$\hat{f}_i = o_i / c_i \tag{2}$$

where o_i represents the observed number of individuals with pedigrees demonstrating type i inbreeding and c_i represents the number of individuals with pedigrees capable of detecting a type i inbreeding event. As an example, Figure 1 illustrates 3 of the myriad pedigrees that could be observed in an empirical data set. If evaluated using the criteria for category 3 as outlined in Table 1, pedigrees that demonstrate inbreeding arising from a full sibling pair (Figure 1A) would contribute to o_3, whereas all 3 pedigrees types would be counted to

Table 1. Pedigree information required to detect inbreeding events associated with 14 different categories that produce inbreeding coefficients (F) of 0.25 or 0.125

Relationship between parents of inbred individual	F	Ancestors of male parent needed to detect inbreeding	Ancestors of female parent needed to detect inbreeding	Pr(I|k)		Pr(I|k_m, k_f)	
				Inbred pedigree	Noninbred pedigree	Inbred pedigree	Noninbred pedigree
1. Father/daughter	0.25	—	Father	k^2	k^3	$k_m \times k_f$	$k_m^2 \times k_f$
2. Mother/son	0.25	Mother	—	k^2	k^3	$k_m \times k_f$	$k_m \times k_f^2$
3. Full siblings	0.25	Both parents	Both parents	k^4	k^6	$k_m^2 \times k_f^2$	$k_m^3 \times k_f^3$
4. Paternal half siblings	0.125	Father	Father	k^3	k^4	$k_m^2 \times k_f$	$k_m^3 \times k_f$
5. Maternal half siblings	0.125	Mother	Mother	k^3	k^4	$k_m \times k_f^2$	$k_m \times k_f^3$
6. Grandson & paternal grandmother	0.125	Father, paternal grandmother	—	k^3	k^4	$k_m^2 \times k_f$	$k_m^2 \times k_f^2$
7. Grandson & maternal grandmother	0.125	Mother, maternal grandmother	—	k^3	k^4	$k_m \times k_f^2$	$k_m \times k_f^3$
8. Paternal grandfather & granddaughter	0.125	—	Father, paternal grandfather	k^3	k^4	$k_m^2 \times k_f$	$k_m^3 \times k_f$
9. Maternal grandfather & granddaughter	0.125	—	Mother, maternal grandfather	k^3	k^4	$k_m \times k_f^2$	$k_m^2 \times k_f^2$
10. Paternal uncle & niece	0.125	Both parents	Father, paternal grandparents	k^5	k^7	$k_m^3 \times k_f^2$	$k_m^4 \times k_f^3$
11. Maternal uncle & niece	0.125	Both parents	Mother, maternal grandparents	k^5	k^7	$k_m^2 \times k_f^3$	$k_m^3 \times k_f^4$
12. Nephew & paternal aunt	0.125	Father, paternal grandparents	Both parents	k^5	k^7	$k_m^3 \times k_f^2$	$k_m^4 \times k_f^3$
13. Nephew & maternal aunt	0.125	Mother, maternal grandparents	Both parents	k^5	k^7	$k_m^2 \times k_f^3$	$k_m^3 \times k_f^4$
14. Double first cousins	0.125	Both parents, all grandparents	Both parents, all grandparents	k^{10}	k^{14}	$k_m^5 \times k_f^5$	$k_m^7 \times k_f^7$

Pr(I|k) reflects the probability of detecting a given pedigree (I) given the proportion of known parents (k). Pr(I|k) is greater for inbred versus noninbred pedigrees. Pr(I|k_m, k_f) represents a refined approach for understanding the probability of observing a given pedigree when differences in the proportion of known male (k_m) and female (k_f) parents exists. See Supplementary Figure 1 for illustrations of the pedigrees listed in this table.



A. $\Pr(I|k) = k^4$ B. $\Pr(I|k) = k^5$ C. $\Pr(I|k) = k^6$

Figure 1. Examples of pedigrees that fulfill requirements for inbreeding category #3 (full sibling pairings; Table 1). In all cases, both parents and all grandparents are known. However, the inbred pedigree in panel **A** requires detection of fewer ancestors relative to the half-sibling inbred pedigree (**B**) and noninbred pedigree (**C**). For this reason, pedigrees associated with full sibling pairings are more likely to be resolved when true relative to the alternatives if incomplete parentage information exists. $\Pr(I|k)$ associated with each pedigree reflects the probability of resolving pedigree I given the fraction of known parents (k).

determine c_3. In the latter case, all 3 pedigrees contain similar information and fulfill requirements for category 3 as outlined in Table 1: both parents and all grandparents in the pedigree are known. By contrast, if evaluated using criteria for category 1, none of the observed pedigrees would contribute to o_1, whereas all 3 pedigrees would still count toward c_1 since both parents and the maternal grandfathers are known (Table 1). Illustrations of inbred and noninbred pedigrees from each of the 14 categories in Table 1 are presented in Supplementary Figure 1. Note that $\hat{f}_i$ takes on values of 1.0 when all individuals associated with category i are inbred and that $\hat{f}_i$ is undefined when there are no observed individuals with pedigrees capable of detecting a type i inbreeding event. Furthermore, $o_i \le c_i$ and the set of individuals with pedigrees contributing to o_i is a subset of or equal to the set of individuals contributing to c_i. Because the pedigrees that contribute to c are not independent and may vary across categories, Marshall et al. proposed a combined estimator across categories as

$$\hat{f}_{\text{tot}} = 1 - \prod(1 - \hat{f}_i) \tag{3}$$

which we continue to use in our analyses.

Revising Estimators to Account for Unequal Detection Probabilities of Different Pedigrees

While $\hat{f}_i$ is a reasonable estimator when pedigrees are largely complete, we show that it will be biased when pedigrees contain moderate to high levels of incomplete parentage information. Again considering pedigrees pertinent to category 3 (Figure 1), fewer individuals need to be detected in the case of inbreeding due to a full sibling pair relative to the other 2 relevant pedigree types that fulfill the requirements outlined in Table 1. For this reason, the pedigree associated with the full sibling pairing is more likely to be detected when true relative to the alternatives if pedigrees contain incomplete parentage information. Differences in detection probabilities of different pedigrees can be defined if we know k: the conditional probability of knowing the identity of a detected individual's parent. k can be quantified in general terms as the proportion of known parents in a data set. Pedigree data sets can be minimally represented as lists that contain information about individuals and their respective maternal and paternal parents (Haig and Ballou 2002). Thus, for any given known individual in a data set, k can be represented as 0 (no known parents), 0.5 (one known parent), or 1 (both parents known). If we assume k to instead represent the average of these values across all detected individuals in the data set, then the probabilities associated with resolving the pedigrees in Figure 1A–C are k^4, k^5, and k^6, respectively, due to the different number of ancestors associated with each configuration. More generally, for any pedigree of interest, the probability of detecting pedigree I given the fraction of known parents is

$$\Pr(I|k) = k^a \tag{4}$$

where a is the number of ancestors in the pedigree. For simplicity, we assume that k is essentially invariant over time. However, in empirical studies, less information will be available in the early years of pedigree development and data collection. We must therefore assume that k, as estimated from the data, is an average representation of the time frame over which inferences of inbreeding are being made.

Because detecting any individual pedigree will be a probabilistic event when $k < 1$, a sample of individuals will contain a fraction of pedigrees of a given type that is observed and a fraction that is not observed. The relative abundance of observed versus unobserved is unknown, but can be represented in general terms as

$$T = T\beta + T(1 - \beta) \tag{5}$$

where T represents the true number of pedigrees of a given type in the sample and $\beta = \Pr(I|k)$, thereby decomposing T into observed ($T\beta$) and unobserved $[T(1 - \beta)]$ components. Substituting τ for the observed quantity, τ then represents the quantity $T\beta$, leading to

$$T = \tau + T(1 - \beta) \tag{6}$$

which can be rearranged to provide an estimator for the true number of pedigrees of that type in the sample as

$$T = \tau / \beta. \tag{7}$$

In essence, this equation reveals that observed counts of pedigrees of a given type should be revised upward by dividing by their probability of detection. This relationship therefore suggests a framework for revising $\hat{f}_i$, where we propose use of

$$\hat{f}_i' = o_{i_{est}} / c_{i_{est}}. \tag{8}$$

If we define B and C to be the sets of pedigrees contributing to o_i and c_i, respectively, then

$$o_{i_{est}} = \frac{o_i}{Pr(I \mid k)} = \sum_{x \in B} \frac{1}{Pr(x \mid k)} \tag{9}$$

and

$$c_{i_{est}} = \sum_{x \in C} \frac{1}{Pr(x \mid k)}. \tag{10}$$

Again using pedigrees in Figure 1 as examples, when evaluating inbreeding category 3, the full sibling inbreeding depicted in Figure 1A would add $1/k^4$ to both $o_{3_{est}}$ and $c_{3_{est}}$ because $\Pr(I|k) = k^4$ for this pedigree since four ancestors are required to fulfill the requirements from Table 1. Likewise, the pedigrees in Figure 1B, C would add $1/k^5$ and $1/k^6$ to $c_{3_{est}}$, respectively, based on the number of ancestors involved when evaluating category 3. Alternatively, when evaluating category 1, none of the pedigrees would contribute to $o_{1_{est}}$, but all 3 would add $1/k^3$ to $c_{1_{est}}$ given the requirements for category 1. Note that under this framework, the original estimator of Marshall et al. is a special case of our revised estimator where $k = 1$ because $\Pr(I|k = 1) = 1$. As with $\hat{f}_i$, combined inbreeding rate estimates over the categories from Table 1 can be obtained by substituting $\hat{f}_i'$ for $\hat{f}_i$ in Equation 3.

This framework can be extended to create an estimator that will be appropriate when differences in the proportions of known male versus female parents exist. The sole required change is the use of $\Pr(I|k_m, k_f)$ instead of $\Pr(I|k)$. $\Pr(I|k_m, k_f)$ can be defined in general terms as

$$\Pr(I|k_m, k_f) = k_m^{a_m} \times k_f^{a_f} \tag{11}$$

where k_m and k_f are the respective proportions of known male and female parents in the data set and a_m and a_f represent the respective number of males and females required to detect pedigree I (Table 1, Supplementary Figure 1). In general, we recommend use of $\Pr(I|k_m, k_f)$ instead of $\Pr(I|k)$ for empirical analyses. However, for the sake of clarity, we retain the use of $\Pr(I|k)$ in the development of analytical expressions outlined below, recognizing that $\Pr(I|k)$ is a simplified representation of $\Pr(I|k_m, k_f)$ where $k_m = k_f$.

Analytical Expressions Defining Behaviors of Estimators

The general behavior of the estimators can be approximated by assuming that only one form of inbreeding occurs in a population (i.e., only one category from Table 1) and that all noninbred pedigrees capable of detecting the inbreeding event are identical. Under this scenario, for a sample of n individuals and a true inbreeding rate of f, there will be $n_1 = n \times f$ inbred individuals and $n_2 = n \times (1 - f)$ noninbred individuals in the sample. However, when $k < 1$, only a fraction of these individuals will be detected depending on $\Pr(I|k)$ for each pedigree type. Determination of $\Pr(I|k)$ for each pedigree type requires knowing the appropriate exponent to use for Equation 4, and we therefore define a_1 to be the exponent for the inbred pedigree and a_2 to be the exponent for the noninbred pedigrees. Because incomplete knowledge of parentage in the sample reduces the observed number of pedigrees of each type, we then expect to only detect $v_1 = n_1 \times k^{a_1}$ inbred pedigrees and $v_2 = n_2 \times k^{a_2}$ noninbred pedigrees in the sample, leading to an expected frequency of detected inbred pedigrees of $\phi_1 = v_1 / n$ and an expected frequency of detected noninbred pedigrees of $\phi_2 = v_2 / n$. Note that in the context of the original Marshall et al. estimator (Equation 2), $o_i = v_1$ and $c_i = v_1 + v_2$ because observed counts of pedigrees are used to calculate $\hat{f}$. Therefore, a simple expression for $\hat{f}$ is

$$\hat{f} = v_1 / (v_1 + v_2). \tag{12}$$

For comparison, an equivalent expression for $\hat{f}'$ is

$$\hat{f}' = \frac{v_1 / k^{a_1}}{v_1 / k^{a_1} + v_2 / k^{a_2}} \tag{13}$$

because $o_{i_{est}} = v_1 / k^{a_1}$ and $c_{i_{est}} = v_1 / k^{a_1} + v_2 / k^{a_2}$ under this simplified scenario where only 2 pedigree types can possibly be detected.

Equations 12 and 13 do not consider that observed counts of pedigrees can only take on integer values. If we define $x_1 \sim \mathit{Binomial}(n, \phi_1)$ and $x_2 \sim \mathit{Binomial}(n, \phi_2)$ for a sample of n individuals and their pedigrees, it becomes possible to enumerate over all possible values of x_1 and x_2 for a given sample size n to derive expectations for $\hat{f}$ and $\hat{f}'$ under different scenarios. Thus, by incorporating the discrete nature of observations into calculations,

$$E[\hat{f}] = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{i}{i+j} \times \Pr(x_1 = i|\phi_1) \times \Pr(x_2 = j|\phi_2) \tag{14}$$

and

$$E[\hat{f}'] = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{i / k^{a_1}}{i / k^{a_1} + j / k^{a_2}} \times \Pr(x_1 = i|\phi_1) \times Pr(x_2 = j \mid \phi_2). \tag{15}$$

We also note that situations exist where both estimators will be undefined. This will occur when x_1 and x_2 are both zero, reflecting detection of neither inbred pedigrees nor pedigrees capable of resolving the inbreeding event. Thus,

$$\Pr(\hat{f} \text{ or } \hat{f}' \text{ is undefined}) = \Pr(x_1 = 0|\phi_1) \times \Pr(x_2 = 0 \mid \phi_2). \tag{16}$$

Finally, as a consequence of the difference in $\Pr(I|k)$ associated with inbred versus noninbred pedigrees, there will be a range of values for k where there are increased chances of detecting more inbred pedigrees than noninbred pedigrees, leading to overestimates of inbreeding rates regardless of the estimator used. Specifically,

$$\Pr(x_1 > x_2) = \sum_{i=1}^{n} \sum_{j=0}^{i-1} \Pr(x_1 = i|\phi_1) \times Pr(x_2 = j \mid \phi_2). \tag{17}$$

Methods

Evaluating Analytical Expressions

Analytical expressions that illustrate the behavior of estimators (Equations 12–17) were evaluated in a computer program written in Python. Expressions were evaluated using varying inbreeding rates, sample sizes (n) of either 1000 or 200 pedigrees, and values of k ranging from 1.0 to 0.01 in increments of 0.01. Each unique set of exponents associated with values of $\Pr(I|k)$ for the 14 inbreeding categories itemized in Table 1 were considered to demonstrate differences among the different pedigree configurations.

Simulations

We used Monte Carlo simulations to compare properties of $\hat{f}$ and $\hat{f}'$ to determine their utility for empirical analyses. Simulations were performed using idealized populations that were tracked over 15 generations. Separate simulations were performed using generation sizes of 100 individuals (50 male and 50 females) or 20 individuals (10 males and 10 females). In each generation, parents for new individuals were chosen using a simple set of rules. Individuals in the first generation were created de novo with no parental information, and second generation progeny were created as offspring of first generation parents. Starting with the third generation, a parent for an individual was selected with equal probability from either of the previous 2 generations, thereby allowing for inbreeding to arise from intergenerational pairings (parent–offspring, uncle–niece, and aunt–nephew). Use of this strategy eliminated the possibility for grandparent–grandchild pairs to form, however, our primary interest was in evaluating overall values across categories for $F = 0.25$ and $F = 0.125$ rather than individual categories. Inbreeding rates were controlled by introducing a parameter that specified the width of a spatial window that determined the number of potential male and female parents for each new individual created in each generation. The widest window size was determined by generation size, and could maximally be set to one half of the generation size (equal to the number of males or females) to allow for random selection of any male or female parent from an appropriate generation. When window sizes were less than the maximum window size, higher inbreeding rates were produced as a consequence of restricting the number of potential parents. Thus, observed inbreeding rates recorded for simulations were emergent properties of the simulation as opposed to prespecified parameters. In simulations based on generation sizes of 100, we evaluated 5000 replicates of 4 window sizes (5, 10, 20, and 50), which yielded overall inbreeding rates of 3.38% ($F = 0.25$) and 10.21% ($F = 0.125$) for a windows size of 5 down to inbreeding rates of 0.50% ($F = 0.25$) and 1.57% ($F = 0.125$) for a window size of 50 (See Results). In analyses with generation sizes of 20, we evaluated 5000 replicates each of simulations using window sizes of 5 (yielding inbreeding rates of 3.4% and 10.1% for $F = 0.25$ and $F = 0.125$, respectively) and 10 (yielding inbreeding rates of 2.7% and 7.7% for $F = 0.25$ and $F = 0.125$, respectively).

True inbreeding rates from each inbreeding category for each simulation replicate were quantified as a population proportion using

$$f_i = o_i / N \tag{18}$$

where o_i is defined as above and N reflects the number of individuals in the simulated population with pedigrees of sufficient depth to meet the detectability requirements of all categories outlined in Table 1. We therefore excluded the first 3 generations from our analyses because they lacked pedigree depth as a consequence of being virtual organisms of spontaneous origin that possessed no pedigrees that could have been sampled, leaving $N = 1200$ (12 generations of 100) or $N = 240$ (12 generations of 20) when determining the true inbreeding rate for each simulation replicate.

After determining the true inbreeding rate for each replicate, we evaluated each of 19 different average levels of missing data (from 5% down to 95% missing in increments of 5%). For each level of missing data, individuals were deleted at random with a probability equivalent to the average missing data rate and all references to a deleted individual's role as a parent were likewise recorded as unknown information. For each level of data deletion, we then calculated average values of $\hat{f}_i$ and $\hat{f}_i^{\prime}$ across 5000 simulation replicates associated with each window size and level of data deletion and used Equation 3 to estimate overall inbreeding rates for the $F = 0.25$ and $F = 0.125$ levels.

Analyses of Empirical Data Sets

We reanalyzed 2 data sets that were originally analyzed by Marshall et al. (2002). One was based on 1767 Arabian oryx (*Oryx leucoryx*) pedigrees, of which 26 individuals were identified as inbred at one of the 14 categories listed in Table 1. The second data set included pedigrees from 2294 red deer (*Cervus elaphus*), of which 19 individuals were inbred. For each data set, we calculated $\hat{f}_i$ and 2 variants of $\hat{f}_i^{\prime}$. The first variant was based on Equations 8–10 and relied on the combined value of k for both male and female parents when determining $\Pr(I|k)$. The second variant used $\Pr(I|k_m, k_f)$ (Equation 11), and therefore accounted for differences in the proportions of known male (k_m) and female (k_f) parents.

Confidence limits for $\hat{f}_i^{\prime}$ were obtained using a beta distribution formulation of the Clopper and Pearson (1934) exact confidence limit for a binomial proportion (Thulin 2014). We define the upper and lower confidence limit as

$$\mathrm{CL}_{\text{lower}} = \mathrm{Beta}(0.025, o_{i_{est}}, c_{i_{est}} - o_{iest} + 1) \tag{19}$$

and

$$\mathrm{CL}_{\text{upper}} = \mathrm{Beta}(0.975, o_{i_{est}} + 1, c_{iest} - o_{i_{est}}), \tag{20}$$

which assumes that $o_{i_{est}}$ represents the number of "successes" and $c_{i_{est}}$ represents the number of trials. This approach has the beneficial attribute of allowing for calculation of an upper confidence limit when the estimated frequency of an event is zero, thereby quantifying the uncertainty associated with cases where no inbreeding was observed for a category.

Results

Analytical Expressions

Our analyses illustrate that $\hat{f}$ is biased and systematically overestimates true inbreeding rates. For example, $\mathrm{E}[\hat{f}]$ from Equation 14 demonstrates an upward bias, however, there is a value of k where maximum bias occurs (Figure 2). The location of the peak bias is not affected by inbreeding rates, but is instead determined by values of a_1 and a_2 associated with inbred and noninbred pedigrees, and to a much lesser extent, by the number of pedigrees examined (n). By contrast, through incorporating information on the probability of detecting different pedigree configurations, our revised estimator $\mathrm{E}[\hat{f}^{\prime}]$ (Equation 15) is largely unbiased over a wide range of values of k that may be encountered in empirical studies (Figure 2), but also shows peaks that are affected by a_1, a_2, and n. With the exception of pedigrees associated with inbreeding by double first-cousins ($a_1 = 10$, $a_2 = 14$; Table 1), $\hat{f}^{\prime}$ remains nearly unbiased as long as $k > 0.5$ for the $n = 1000$ case, and is likewise unbiased for the $n = 200$ case when $k > 0.6$ (Figure 2). Lower values of a_1 and a_2 result in unbiased estimates of $\hat{f}^{\prime}$ for values of k down to 0.2–0.3 for the $a_1 = 2$ and $a_2 = 3$ cases (Figure 2; categories 1 and 2 in Table 1).

The bias in $\hat{f}^{\prime}$ across some values of k can be attributed to the joint probabilities associated with observing different combinations of inbred and noninbred pedigrees. Specifically, bias is most pronounced at values of k where there is the greatest probability of detecting more inbred than noninbred pedigrees, even when inbreeding rates are low (Equation 17; Supplementary Figure 2). Note that, although bias will be most pronounced for pedigrees associated with double first-cousins, those pedigrees have the lowest probability of being detected even with modest levels of unknown parentage (Equation 4; Supplementary Figure 3). Likewise, pedigrees that require more ancestors to resolve also have higher probabilities of generating undefined inbreeding rate estimates (Equation 16; Supplementary Figure 4), in which case no estimates of an inbreeding rate will be possible.

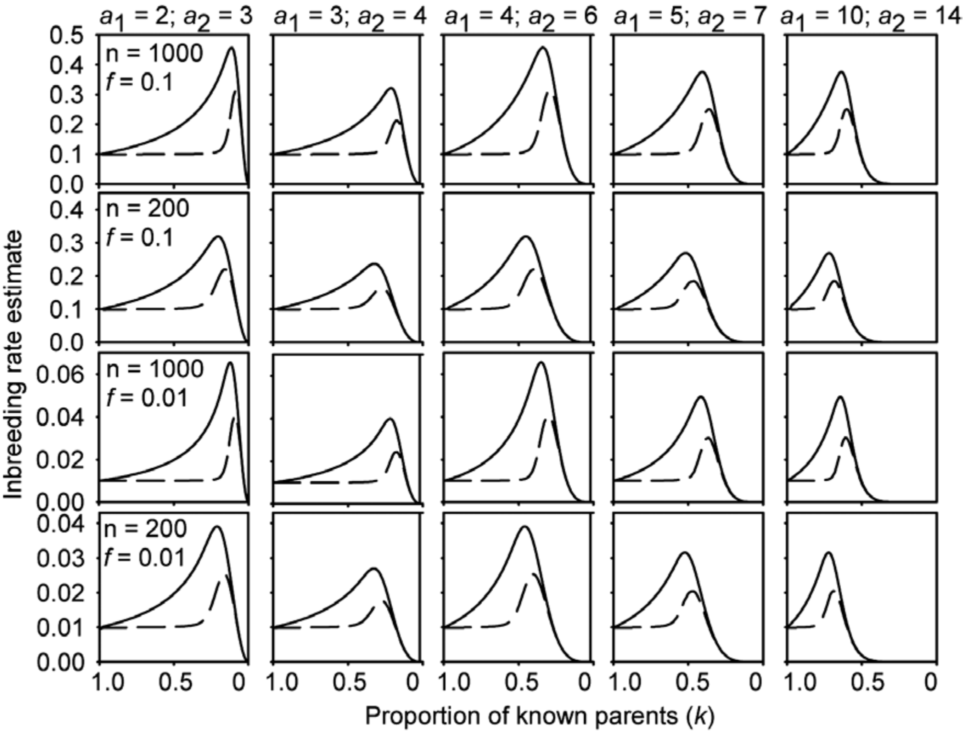


Figure 2. Analytical expectations for $\hat{f}$ (Equation 14: solid line) and $\hat{f}'$ (Equation 15: dashed line) as the proportion of known parents decreases. Values of a_1 and a_2 are exponents associated with inbred and noninbred pedigrees, respectively, as categorized in Table 1.

Computer Simulations

We used computer simulations to further explore properties of the estimators. Simulations were performed in a manner that allowed for diverse inter- and intragenerational combinations of parents to be chosen, thereby resulting in simulated pedigrees containing multiple forms of inbreeding that permitted evaluation of combined estimates of inbreeding for $\hat{f}$ and $\hat{f}'$ across categories associated with $F = 0.25$ and $F = 0.125$ (Equation 3). Simulation results (Figure 3) were similar to those from our evaluation of the analytical expressions (Figure 2). Specifically, as with the evaluation of individual inbreeding categories, the combined inbreeding rate estimator is highly biased when $\hat{f}$ is used. However, there is a broad range of values of k where $\hat{f}'$ is essentially unbiased. For a given population size, the overall inbreeding rates have little impact on the range of values of k where $\hat{f}'$ is unbiased (Figure 3), and this range is slightly reduced for smaller sample sizes than for larger sample sizes. Simulation results for individual inbreeding categories were also similar to our analytical findings (Supplementary Tables 3–8). We observed broader ranges of k where $\hat{f}'$ is unbiased for inbreeding categories associated with higher values of $\Pr(I|k)$ relative to lower values of $\Pr(I|k)$.

Examples with Empirical Data Sets

Analysis of Arabian oryx data produced expected results in light of outcomes from our analytical models and computer simulations. After accounting for the fraction of known parents ($k = 0.88$), inbreeding rate estimates were revised downward for all categories where an inbreeding event had been detected (Supplementary Table 1). Considering category totals, the estimated frequency of pairings that produce inbreeding at $F = 0.25$ was reduced from 9.6% down to 8.1% whereas inbreeding at $F = 0.125$ was reduced from 13.3% to 11.5%. We also generated estimates where the proportion of known parents varied by sex ($k_m = 0.822$, $k_f = 0.938$), however, accounting for this variation had minimal effects on category totals (Supplementary Table 1) relative to estimates based on a single value of k.

Analyses of the red deer data (Supplementary Table 2) also conformed to expectations. When considering that only 60% of parents were known ($k = 0.603$), the total inbreeding rate for $F = 0.25$ was revised from 1.5% to 0.9%, whereas inbreeding leading to $F = 0.125$ was reduced from 11.8% to 7.4%. Unlike the Arabian oryx data set, the red deer data had a substantially smaller fraction of known male ($k_m = 0.243$) versus female ($k_f = 0.964$) parents. When accounting for this variation by using $\Pr(I|k_m, k_f)$, the estimated inbreeding rates were further reduced to 0.378% and 3.988% for $F = 0.25$ and $F = 0.125$, respectively.

Discussion

Quantifying inbreeding rates in the wild has long been recognized as an important goal for wildlife conservation and management, however, the substantial information required to reconstruct pedigrees has made generation of accurate inbreeding rate estimates difficult (Haig and Ballou 2002). Our analyses illustrate that inbreeding rates can be obtained from data sets comprised of incomplete pedigrees if the probability of detecting different pedigree configurations is taken into account. If not incorporated into analyses, estimates will become increasingly biased as the fraction of unknown parentage increases.

We developed analytical expressions to illustrate properties of pedigree data sets that contain incomplete parentage information. Because different pedigrees require different numbers of ancestors to resolve (Table 1; Supplementary Figure 1), the probability of observing a specific pedigree when it is true will vary depending on the fraction of known parents and the number of ancestors needed

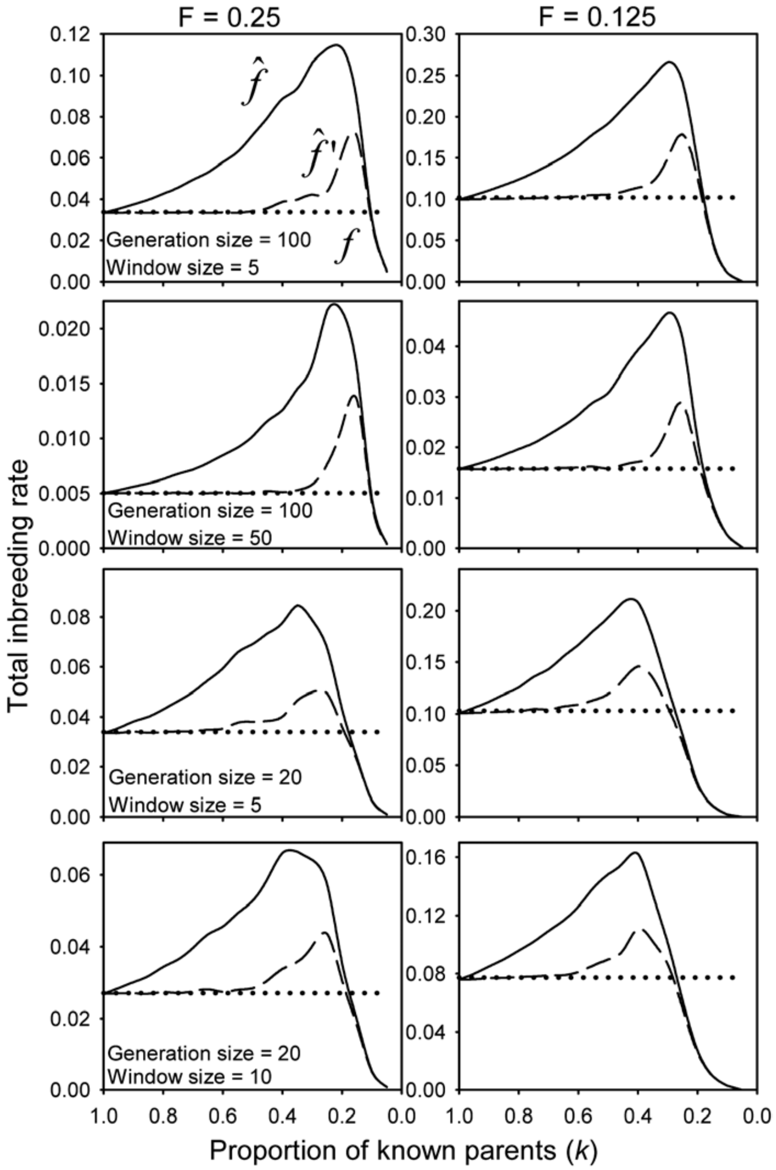


Figure 3. Simulation results illustrating the behavior of inbreeding rate estimators. True inbreeding rates (f) are indicated by the dotted horizontal lines whereas solid and dashed lines reflect inbreeding rate estimators $\hat{f}$ and $\hat{f}'$, respectively. Results shown here reflect category totals for $F = 0.25$ and $F = 0.125$. Results for individual inbreeding categories and other window sizes not displayed are provided in Supplementary Tables 3–8.

to complete the pedigree (Equations 4 and 11). This understanding highlights practical issues that should be considered when estimating inbreeding rates in empirical investigations. Because some of the pedigree types in Table 1 require comparatively large numbers of ancestors to document, small empirical data sets may be limited in their ability to reasonably estimate inbreeding rates for a given category. For example, assuming a modest value of $k = 0.8$, the probability of detecting an inbred pedigree associated with aunt/nephew or uncle/niece pairings is only ~0.33, whereas a comparable noninbred pedigree as outlined in Table 1 will be observed with a probability of ~0.21. This situation becomes more unfavorable with lower values of k: assuming $k = 0.5$, the probabilities of detecting inbred and noninbred pedigrees will be 0.03 and 0.008, respectively, indicating that a large pedigree data set will be required to observe these pedigree types when k is small. Consequently, researchers should appraise their data sets prior to attempting to estimate inbreeding rates to ensure that there are realistic opportunities to detect relevant pedigree configurations. For smaller data sets, restricting analyses to examining parent/offspring pairs or half sibling pairs may be necessary, whereas observing sufficient pedigrees to inform the double-first cousin category could be problematic. For this reason, we advocate reporting values of o_i and c_i, such as we have in Supplementary Tables 1 and 2, since these represent the raw counts that form the basis for our revised estimator. This information may help others better understand the extent that a given pedigree data set will be capable of providing reasonable inbreeding rate estimates for individual inbreeding categories.

Many pedigree configurations require different numbers of male versus female ancestors to resolve (Table 1). Thus, a beneficial attribute of our framework is the flexibility that allows it to accommodate data where differences in the fractions of known male versus female parents exist. Differences in sex-specific survival (Promislow et al. 1994; Loison et al. 1999; Spidle et al. 1998; Toïgo and Gaillard 2003; Sperry and Weatherhead 2009) and dispersal (Greenwood 1980;

Johnson 1986; Clarke et al. 1997) occur in many animal species, which may translate in empirical pedigrees to tractable differences in the proportion of known parents from each sex. Consequently, in most cases, use of $\Pr(I|k_m, k_f)$ for calculations will be preferred over the more general $\Pr(I|k)$, noting that $\Pr(I|k_m, k_f) = \Pr(I|k)$ when $k_m = k_f$. In our analyses of empirical data sets, use of $\Pr(I|k_m, k_f)$ resulted in only minor revision to estimates for individual inbreeding categories in the Arabian oryx data set (Supplementary Table 1), however, the proportion of known parents was relatively high for these data. This pattern is in contrast to the red deer data set (Supplementary Table 2), where substantial differences in the proportion of known male and female parents existed ($k_\mathrm{m} = 0.243$, $k_f = 0.964$). In this case, using $\Pr(I|k_m, k_f)$ for calculations resulted in a further substantial downward revision of overall inbreeding rates, even relative to estimates based on $\Pr(I|k)$.

There are numerous applications for the inbreeding rate estimator described in this article. First, because the estimates are largely unbiased, comparisons may be made among inbreeding categories (Table 1) to identify those that are over or under-represented in a data set. If differences are identified, then results may be pointing to mechanisms leading to avoidance of some forms of inbreeding. For example, if the rate of inbreeding due to father/daughter pairings is higher than the rate for mother/son pairs, then one potential inference is that juvenile dispersal of males, but not females, is common in the species being examined. Such observations could be corroborated with field observations on individual behavior to develop a concrete understanding of the basis for inbreeding avoidance and further our understanding of conditions where inbreeding is more likely to occur. Second, in the case of threatened and endangered species, an understanding of inbreeding rates may facilitate the development of management strategies and population models to support conservation efforts. Identification of populations with high inbreeding rates may suggest that translocations could be used to promote genetic rescue of a population (Tallmon et al. 2004; Trinkel et al. 2008; Hedrick and Fredrickson 2010; Weeks et al. 2011; Heber et al. 2013). Similarly, robust inbreeding rate estimates may facilitate the development of more refined population viability models (Brook 2000; Lacy 2000a, 2000b; Brook et al. 2002; Haig and Ballou 2002), leading to better predictions about the effects or consequences of alternative management actions for individual populations.

Supplementary Material

Supplementary data are available at *Journal of Heredity* online.

Funding

Project funding was provided by the USGS Forest and Rangeland Ecosystem Science Center, the USDA Forest Service Pacific Northwest Research Station, and Sierra Pacific Industries.

Acknowledgments

The authors thank Tristan Marshall for allowing us to reexamine the Arabian oryx and red deer data. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.

Data Availability

Source code and simulated data for this study are available at http://dx.doi.org/10.5066/F7QR4V85.

References

Ballou J. 1983. Calculating inbreeding coefficients from pedigrees. In: Schonewald-Cox CM, Chambers SM, MacBryde B, Thomas L, editors. *Genetics and conservation: a reference for managing wild animal and plant populations*. Menlo Park (CA): Benjamin/Cummings Publishing Company. p. 509–520.

Ballou JD. 1995. *Genetic management, inbreeding depression and outbreeding depression in captive populations* [PhD dissertation]. [College Park (MD)]: University of Maryland.

Ballou JD. 1997. Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations. *J Hered*. 88:169–178.

Brook BW. 2000. Pessimistic and optimistic bias in population viability analysis. *Conserv Biol*. 14:564–566.

Brook BW, Tonkyn DW, O'Grady JJ, Frankham R. 2002. Contribution of inbreeding to extinction risk in threatened species. *Conserv Ecol*. 6:16.

Bulmer MG. 1973. Inbreeding in the great tit. *Heredity (Edinb)*. 30:313–325.

Clarke AL, Sæther B-E, Røskaft E. 1997. Sex biases in avian dispersal: a reappraisal. *Oikos*. 79:429–438.

Clopper CJ, Pearson ES. 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 26:404–413.

Crnokrak P, Roff DA. 1999. Inbreeding depression in the wild. *Heredity*. 83(Pt 3):260–270.

Frankham R. 2005. Genetics and extinction. *Biol Conserv*. 126:131–140.

Greenwood PJ. 1980. Mating systems, philopatry, and dispersal in birds and mammals. *Anim Behav*. 28:1140–1162.

Haig SM, Ballou JD. 2002. Pedigree analyses in wild populations. In: Beissinger SR, McCullough DR, editors. *Population viability analysis*. Chicago (IL): University of Chicago Press. p. 388–405.

Heber S, Varsani A, Kuhn S, Girg A, Kempenaers B, Briskie J. 2013. The genetic rescue of two bottlenecked South Island robin populations using translocations of inbred donors. *Proc Biol Sci*. 280:20122228.

Hedrick PW, Kalinowski ST. 2000. Inbreeding depression in conservation biology. *Annl Rev Ecol Syst*. 31:139–162.

Hedrick PW, Fredrickson R. 2010. Genetic rescue guidelines with examples from Mexican wolves and Florida panthers. *Conserv Genet*. 11:615–626.

Howard WE. 1949. Dispersal, amount of inbreeding, and longevity in a local population of Prairie deermice on the George Reserve, southern Michigan. *Contrib Lab Vert Biol*. 43:1–50.

Johnson CN. 1986. Sex-biased philopatry and dispersal in mammals. *Oecologia*. 69:626–627.

Keller LF, Waller DM. 2002. Inbreeding effects in wild populations. *Trends Ecol Evol* 17:230–241.

Lacy RC. 2000a. Considering threats to the viability of small populations using individual-based models. *Ecol Bull*. 48:39–51.

Lacy RC. 2000b. Structure of the VORTEX simulation model for population viability analysis. *Ecol Bull*. 48:191–203.

Lande R. 1988. Genetics and demography in biological conservation. *Science*. 241:1455–1460.

Loison A, Festa-Bianchet M, Gaillard J-M, Jorgensom JT, Jullien J-M. 1999. Age-specific survival in five populations of ungulates: evidence of senescence. *Ecology*. 80:2539–2554.

Marshall TC, Coltman DW, Pemberton JM, Slate J, Spalton JA, Guinness FE, Smith JA, Pilkington JG, Clutton–Brock TH. 2002. Estimating the prevalence of inbreeding from incomplete pedigrees. *P R Soc B*. 269:1533–1539.

O'Grady JJ, Brook BW, Reed DH, Ballou JD, Tonkyn DW, Frankham R. 2006. Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biol Conserv*. 133:42–51.

Pemberton J. 2004. Measuring inbreeding depression in the wild: the old ways are the best. *Trends Ecol Evol*. 19:613–615.

Pemberton JM. 2008. Wild pedigrees: the way forward. *Proc Biol Sci*. 275:613–621.

Promislow D, Montgomerie R, Martin TE. 1994. Sexual selection and survival in North American waterfowl. *Evolution*. 48:2045–2050.

Pusey A, Wolf M. 1996. Inbreeding avoidance in animals. *Trends Ecol Evol*. 11:201–206.

Ralls K, Brugger K, Ballou J. 1979. Inbreeding and juvenile mortality in small populations of ungulates. *Science*. 206:1101–1103.

Ralls K, Ballou JD, Templeton A. 1988. Estimates of lethal equivalents and the cost of inbreeding in mammals. *Conserv Biol*. 2:185–192.

Sperry JH, Weatherhead PJ. 2009. Sex differences in behavior associated with sex-biased mortality in an oviparous snake species. *Oikos*. 118:627–633.

Spidle AP, Quinn TP, Bentzen P. 1998. Sex-biased marine survival and growth in a population of coho salmon. *J Fish Biol*. 52:907–915.

Tallmon DA, Luikart G, Waples RS. 2004. The alluring simplicity and complex reality of genetic rescue. *Trends Ecol Evol*. 19:489–496.

Thulin M. 2014. The cost of using exact confidence intervals for a binomial proportion. *Electron J Stat*. 8:817–840.

Toïgo C, Gaillard J-M. 2003. Causes of sex-biased adult survival in ungulates: sexual size dimorphism, mating tactic or environmental harshness? *Oikos*. 101:376–384.

Trinkel M, Ferguson N, Reid A, Reid C, Somers M, Turelli L, Graf J, Szykman M, Cooper D, Haverman P, *et al*. 2008. Translocating lions into an inbred lion population in the Hluhluwe-iMfolozi Park, South Africa. *Anim Conserv*. 11:138–143.

van Noordwijk AJ, Scharloo W. 1981. Inbreeding in an island population of the great tit. *Evolution*. 35:674–688.

Weeks AR, Sgro CM, Young AG, Frankham R, Mitchell NJ, Miller KA, Byrne M, Coates DJ, Eldridge MD, Sunnucks P, *et al*. 2011. Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evol Appl*. 4:709–725.