

# 21

## Family-Based Selection

*Practical breeding programs must be commercially optimal,  
not theoretically maximal. — Fairfull and Muir (1996)*

*Draft Version 11 May 2013.*

Up to now, we have focused on **individual selection**, wherein selection decisions are based solely on the phenotypes of single individuals (this is also referred to as **mass** or **phenotypic selection**, and we use these terms interchangeably). Selection decisions can also incorporate the phenotypic values of an individual's measured relatives, and in fact most plant and animal breeding schemes do so. The focus of this chapter is **family-based** selection — using family information to select individuals. We restrict discussion in this chapter to using sib information. Chapters 31–32 develop the general theory of index selection, which can accommodate information from arbitrary types of relatives. The culmination of this approach is BLUP selection using an index based on the entire known pedigree of an individual, which is the major route for artificial selection in most domesticated animals. While our focus here is on short-term response (formally, the single-generation response), certain family-based schemes may give a greater long-term response than individual selection, even when their initial response is less. This long-term advantage arises because of larger effective population sizes associated with selection schemes that down-weight between-family differences, a point examined in detail in Chapter 24. Likewise, a number of family-based schemes in plant breeding use extensive inbreeding or crossbreeding and these are examined in Volume 3. Our focus here is on using family-based selection for the improvement of a single population.

There are a variety of reasons for using family-based schemes. Mass selection may be impractical in many settings due to difficulties in measuring trait values in individuals (e.g., grasses growing in a field). Family-based designs can also provide greater accuracy in predicting an individual's breeding value and hence can give a larger (short-term) response. In particular, an appropriately weighted index of an individual's family mean and phenotypic value has an expected response at least as large as mass selection. When significant environmental heterogeneity exists (e.g., crops planted across a broad climatic range), replication of families over environments provides a more efficient method than mass selection for choosing higher-performing genotypes. This is one major factor why crop breeders favor family-based schemes over individual selection.

The structure of this chapter is as follows. We start with a few overview remarks on the nature and types of family-based selection before considering extensions of the generalized breeders equation to accommodate family-based schemes. We next develop the variances and covariances required to apply these equations, and then consider a number of these schemes in detail. The relative efficiencies of within- and between-family selection compared to mass selection are examined next, followed by consideration of designs where families are replicated over environments, as is usually the case in plant breeding. We conclude by examining the properties of family index selection. While most of the concepts in this chapter are straightforward, the bookkeeping can be a bit tedious at times. Thus, we have tried to summarize key results at the end of various sections to allow the casual reader to more easily navigate through this material.

## INTRODUCTION TO FAMILY-BASED SELECTION SCHEMES

Family-based designs are based on two components: **between-family** schemes that choose entire families on the basis of their mean performance, and **within-family** schemes that choose individuals based on their relative performance within their families. While many designs are based on just one of these components, the most general approach is **family-index selection** where individuals are chosen based on a weighted index of between- and within-family components. Mass selection is a special case of a family index, where the within- and between-family components are weighted equally.

While we assume here that the parents for any particular family are from the same population, in some plant-breeding settings parents are from different populations. An example of this are **interpopulation improvement** schemes where the goal is to improve the performance of *hybrids* between populations (Volume 3). Our focus in this chapter are family-based schemes for **intrapopulation improvement** (increasing the performance of the population under selection).

### Overview of the Different Types of Family-Based Selection

The key to making sense out of the often bewildering number of family-based designs that appear in the literature is to consider the individual components that together define any particular design. The first component is the type of sib family providing information for selection decisions. A family may consist of half-sibs, full-sibs, or full-sibs nested within half-sibs (the NC Design I, see LW Chapter 18). Sibs can also be generated by one (or more) generations of selfing (e.g.,  $S_1$ ,  $S_2$ ), and we examine such families in Chapter 20. While the family-based schemes developed in this chapter are generally used with **allogamous** species (outcrossers, cross-pollination), they can also be applied to facultatively **autogamous** species (facultative selfers, self-pollination) through the use of either controlled pollination and/or the introduction of male-sterile genes under open pollination (e.g., Gilmore 1964, Doggett and Eberhart 1968, Brim and Stuber 1973, Burton and Brim 1981, Sorrells and Fritz 1982).

Once a particular family type has been chosen, the second component is how sib data are used for selection decisions. One could use between-family selection, choosing the best families (i.e., those with the largest family means). Alternatively, one could use within-family selection, choosing either the best individuals within each family (**strict within-family [WF] selection**), or the individuals with the largest overall family deviations (**family-deviations [FD] selection**). While WF and FD selection are very similar, there are subtle differences between the two schemes, as they do not necessarily select the same individuals. One could also consider an index weighting both family mean and family deviations.

Finally, a major design component is the relationship between the measured sibs and the individuals serving as parents for the next generation. Under either within-family or family-index selection, the selected individuals themselves are used to form the next generation. However, with between-family selection, we can use any number of relatives of the chosen families to form the next generation. The most straight-forward is to use some (or all) of the measured sibs from the chosen families (**family selection**). Some characters cannot be scored on living organisms, such as carcass traits in meat animals. In such cases, one can use unmeasured sibs from the best families as the parents of the next generation (**sib selection**). Sib selection can also be used to improve selection on sex-specific traits. For example, milk production can be selected in males by choosing males from families whose sisters show high levels of milk production. An important variant of sib selection is to use **remnant seeds** from the best families, which are planted and subsequently crossed to form the next generation. For perennial species and for annual species that can be asexually propagated (or cloned), one can select the best parents by the performance of their offspring (**parental selection** or **progeny testing**). Finally, an option available for facultatively autogamous species is to

both self an individual to generate  $S_1$  progeny ( $S_1$  seeds) and likewise outcross it to one or more individuals to generate a family for testing. For such species, one can grown up and intercross  $S_1$  seed from the chosen families to form the next generation (the  $S_1$  seed design).

### Plant vs. Animal Breeding

While animal breeders typically employ only a few standard sib-based designs (Turner and Young 1969), an array of options are available to plant breeders (e.g., Hallauer and Miranda 1981, Wricke and Weber 1986, Nyquist 1991). Further, the final product desired by a plant breeder can vary considerably, e.g., an open-pollinated population,  $F_1$  hybrid, pure line, or synthetic. Thus it is not surprising that the literatures on family-based selection in the two fields are rather divergent. Much of the animal breeding literature is expressed in terms of the phenotypic ( $t$ ) and additive-genetic ( $r$ ) correlations among sibs. In contrast, much of the plant breeding literature is expressed in terms of variance components. As our discussion will attempt to interweave both approaches, we will typically present response equations in both forms.

Reproductive differences between plants and animals underlie many of the differences in designs available to breeders. Historically, plant breeders have had more options than animal breeders because of the reproductive flexibility of many plants (i.e., selfing, stored seed, vegetation propagation, see Fehr and Hadley 1990). With the recent successes in cloning several domesticated animals, animal breeders now have the option of exploiting some of these classic plant breeding schemes.

One obvious difference between plants and many animals is the ability to easily store progeny for many generations in the form of seed. Generally speaking, plants also produce far more offspring than domesticated animals, providing more offspring per family, allowing for more extensive replication of families across environments. Another reproductive advantage of plants is that asexual propagation (cloning) is straightforward in many species, allowing individual genotypes to be preserved over many generations.

Another key difference is in control of crosses. While simple isolation will prevent most undesirable crosses in animals, either complete isolation or extensive manual control may be required to prevent pollination vectors from generating undesirable crosses in plants. For facultatively autogamous species, the investigator may be faced with either trying to prevent selfing or prevent outcrossing, or to allow for both while identifying which seed came from which type of cross. Options for controlled crosses range from complete manual control over pollination at one extreme to open pollination at the other. Given that most plants have multiple flowers (which are often both very numerous and very small), large-scale controlled crosses can be much more labor intensive than similar crosses among animals, as hand pollination and control of external and/or self pollinators may be required. Even under **open pollination** (allowing seed plants to be pollinated at random), the investigator still has different levels of control over the pollen spectrum that a seed plant experiences. In a **test cross** or **topcross** design, the population of plants supplying the pollen is controlled. For example, individual maize plants can be **detasseled** by hand (removing the pollen-producing flowers) or have their tassels bagged to prevent the plants from either selfing or pollinating other plants. Such plants serve only as seed plants and are intergrown with rows of the **tester** strain which provides the pollen. Under true open pollination, seed parents are randomly pollinated from the population with no control of the pollen parent. A consequence of open pollination is that while most half-sib families in animal breeding are paternal, most half-sib families in plant breeding are maternal (sharing the same seed parent).

There are also more subtle biological differences between plants and animals that flavor differences in designs. While one can usually score many traits in individual animals, this is often not the case in plants. For example, many traits of forage grasses, grains, and legumes are scored as *plot totals*, measuring an entire family (or line) instead of each separate individ-

ual. When individuals cannot be directly scored, between-family selection is possible, but within-family and family-index selection is not. Similarly, many selective traits in plants can be scored only after reproduction (seed or fruit yield being prime examples), and this may influence the types of relatives that can be used to form the next generation. Finally, there are always resource limitations that favor some designs over others.

### Between- vs. Within-Family Selection

When the heritability of a trait is high, an individual's phenotype is an excellent predictor of its breeding value, and mass selection is more efficient than either strict within- or between-family selection. When heritability is low, individual phenotypic value is a poor predictor of breeding value, in which case an individual's family mean or its relative performance within its family may better predict its breeding ability.

The relative efficiencies of between- vs. within-family selection depend on the relative magnitudes the common-family ( $E_c$ ) and individual-specific ( $E_s$ ) environmental values. A large common-family effect severely compromises the phenotype as a predictor of breeding value. However, *within* each family, all members share the same environmental effect  $E_c$  and differences between individuals more accurately reflect differences in breeding value. In this case, selection within families (for example, by choosing the largest individuals from each family) can give a larger response than individual selection. For example, many mouse selection experiments use within-family selection, especially for traits with suspected maternal effects such as body weight (Falconer and Latyszewski 1952; Falconer 1953, 1960a; Eisen and Hanrahan 1972; von Butler et al. 1984; Nielsen and Anderson 1987; Siewerdt et al. 1999), litter size (Falconer 1960b), and nesting behavior (Lynch 1980).

Conversely, suppose that environmental effects unique to each individual account for a large fraction of the phenotypic variance. In this case, selecting whole families as units can give a larger response than individual selection, as the family mean averages out differences based on environmental values, revealing those families with the most extreme breeding values. An important example of this family-averaging of environmental effects is the use of between-family selection to improve performance across multiple environment. Under mass selection, a genotype is represented by a single individual in a single environment, while family-based approaches allow the performance of different families to be compared over multiple environments. Such studies are by no means restricted to plant breeding, as animal selection experiments examining phenotypic plasticity (norms of reactions), in which genotypes must also be assessed over multiple environments, almost exclusively use between-family selection (e.g., Waddington 1960, Waddington and Robertson 1966, Kindred 1965, Druger 1967, Scharloo et al. 1972, Minawa and Birley 1978, Scheiner and Lyman 1991, Brumpton et al. 1997). Chapters 33–34 examines selection over multiple environments in some detail, while selection for phenotypic plasticity is reviewed in Volume 3.

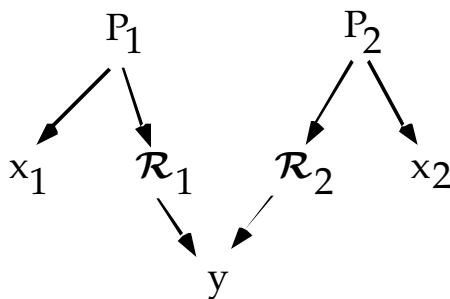
## DETAILS OF FAMILY-BASED SELECTION SCHEMES

### Selection and Recombination Units

Under mass selection, individuals are scored and those with the best phenotypic values are used as parents to form the next generation. Here group of individuals (or **unit**) upon which selection decisions are based and unit used for recombination (gamete production to form the next generation) are one in the same, and a single **cycle of selection** takes a single generation. In family-based selection schemes the individuals used for selection decisions may be entirely separate from those used to form the next generation. Further, a single cycle of selection may take two (or more) generations, as one must generate, score, and recombine

families.

Following the convention of plant breeders, we therefore distinguish between an individual  $x_i$  in the **selection unit** (those individuals upon which selection decisions are made, which we assume throughout this chapter are sibs) and an individual  $\mathcal{R}_i$  from the **recombination unit** (individuals serving as parents for the next generation) whose resulting offspring are  $y$ . Even though we may not *directly* select on the parents ( $\mathcal{R}_1, \mathcal{R}_2$ ) of  $y$ , we expect some response in  $y$  due to the a genetic correlation between  $x_i$  and  $\mathcal{R}_i$  arising from them sharing (at least) one common relative  $P_i$  (Figure 21.1). Equivalently, response occurs due to observations on the selection unit  $x$  providing information to predict the breeding value of  $\mathcal{R}$ .



**Figure 21.1.** Under family-based schemes, selection decisions are based on some function of the values of sibs  $x_i$  in the selection unit. An offspring  $y$  in the next generation has parents  $\mathcal{R}_1$  and  $\mathcal{R}_2$  that are chosen on the basis of the selection unit. Members of the selection ( $x_i$ ) and recombination ( $\mathcal{R}_i$ ) units are related as they both share a common relative  $P_i$ , the parent of sib  $x_i$ . Under within-family or family index selection,  $\mathcal{R}$  is just one of the measured sibs, while under between-family selection,  $\mathcal{R}$  is often an unmeasured relative.

As mentioned in the introduction, the variety of family-based schemes appearing in the literature arises by combining four specific components.

1. **Type of sib family** comprising the selection unit. Sibs can be half-or full-sibs, full sibs nested within half-sibs (NC Design I), or  $S_1$  selfed sibs (which are considered in Chapter 20).
2. **Nature of the selection decisions** based on the sib information. Selection can be based on sib means, the deviations of individuals within families, an index of both, or strict rank within families.
3. **Selection on one vs. both parents.** Often selection decisions involve only one sex, with the parents of the opposite sex chosen at random (and hence are unselected). For example, the best plants may not be able to be scored until after pollination, resulting in selection on females (seed parents) but not males (pollen parents). In such cases, we are only concerned with one side of the pedigree, for example involving  $\mathcal{R}_1$  but not  $\mathcal{R}_2$ . More generally, the parents ( $\mathcal{R}_1$  vs.  $\mathcal{R}_2$ ) may be chosen using different schemes, further increasing the variety of potential family-based schemes.
4. **Nature of the relationship between a measured sib  $x_i$  in the selection unit and a parent of the next generation  $\mathcal{R}_i$ .** Under within-family or family index selection,  $\mathcal{R}$  is one of the measured sibs ( $\mathcal{R}_i = x_i$ ), while under between-family selection,  $\mathcal{R}$  is often an unmeasured relative. For example it could be the parent of the sibs

( $\mathcal{R}_i = P_i$ ) so that the relationship between  $x$  and  $\mathcal{R}$  is that of parent-offspring or it could be an unmeasured sib so that the relationship is that of either half- or full-sib (depending on the type of family).

While the variety family-based selection schemes may seem a bit overwhelming at first (especially in plant breeding!), considering each design in terms of these four basic components greatly simplifies matters.

### Variations of the Selection Unit

Once the type of family (half sib, full-sib, nested,  $S_i$ ) has been specified, there is still the issue of how to incorporate sib information when making selection decisions. To distinguish between a particular sib  $x$  and the *trait value* of that sib, we use  $x$  to index the sib and  $z$  to denote its trait value. Suppose that we select the uppermost fraction  $p$  of the relevant population and let  $z_{ij}$  denote the  $i$ th individual from the  $j$ th family, with  $m$  families each with  $n$  sibs, for a total of  $N = mn$  scored individuals. Four different approaches for weighting sib information are commonly used:

1. **Between-family selection:** Individuals are selected solely on the basis of their family means,  $\bar{z}_i$ , so that all individuals from the same family have the same selective rank. Here, the best  $pm$  families are chosen.
2. **Strict within-family (WF) selection.** The best  $pn$  individuals from *each* family are chosen, so that individuals are ranked within each family.
3. **Selection on within-family deviations (FD):** Individuals are ranked solely on the basis of their within-family deviation  $z_{ij} - \bar{z}_i$ . The  $pN$  individuals with the largest deviations (regardless of family) are chosen.
4. **Family-index selection:** Individuals are ranked using an index weighting within- and between-family components,

$$\begin{aligned} I &= b_1 (z_{ij} - \bar{z}_i) + b_2 \bar{z}_i \\ &= b_1 z_{ij} + (b_2 - b_1) \bar{z}_i \end{aligned}$$

The  $pN$  individuals with the best index score are chosen. Note that the index with weights  $(b_1, b_2)$  chooses the same individuals as an index with weights  $(b_1, b_2)$ . As a result, one of the index weights is often set to one, as the indices with weights  $(b_1, b_2)$ ,  $(1, b_2/b_1)$ , and  $(b_1/b_2, 1)$  are all equivalent. Individual selection, between-family selection, and selection on family deviations (FD) are special cases, being indices with weights  $(b_1, b_2) = (1, 1)$ ,  $(0, 1)$ , and  $(1, 0)$ , respectively. Note, however, that strict within-family (WF) selection cannot be expressed in terms of an index. Family-index selection is also referred to as **combined selection**, which is unfortunate as the same term is also used by breeders to refer to schemes that combine different types of selection schemes in a single cycle (such as modified ear-to-row, discussed below).

The choice of a particular scheme has implications on the selection intensity. When the fraction saved  $p$  is fixed, between-family and strict within-family selection have lower selection intensities than family-deviation, index, or mass selection. The former selects the best  $pm$  of  $m$  families and  $pn$  of  $n$  sibs, while the latter three select the best  $pN$  of  $N$  individuals. Since  $N > n, m$  the finite-sample value for  $\bar{r}$  is larger when sampling from  $N$  than from  $n$  or  $m$ .

**Example 21.1.** Suppose that a total of 100 sibs are measured and the fraction selected is  $p = 0.2$ . As a benchmark, for this level of selection, the infinite-population value for the selection intensity is  $\bar{z} = 1.40$  (Equation 10.26). Suppose that the 100 total measured sibs are distributed into 20 families of five sibs each ( $m = 20, n = 5$ ). Under within-family selection, the top 1 of 5 within each family is selected. Under between-family selection, the top 4 of the 20 families are selected. Finally, under family-deviations or index selection, the top 20 of the 100 measured individuals are selected. Using the finite-size correction approximation offered by Equation 10.27b gives the following selection intensities:

Individual selection (infinite population)	Upper 20%	$\bar{z}_{\infty} = 1.40$
Individual selection, Index Selection, Family-deviations selection	Best 20 of 100	$\bar{z}_{100,20} = 1.39$
Between-family selection	Best 4 of 20	$\bar{z}_{20,4} = 1.33$
Strict within-family selection	Best 1 of 5	$\bar{z}_{5,1} = 1.16$

As will be shown later in the Chapter (Equations 21.40 and 21.57) that additional corrections to the selection intensity are required in some cases, as family members are correlated, which changes the variance relative to  $n$  unrelated individuals. Ignoring this correction (for now), note that strict within-family (WF) selection has only 83% of the selection intensity (for this example) as family-deviations (FD) selection.

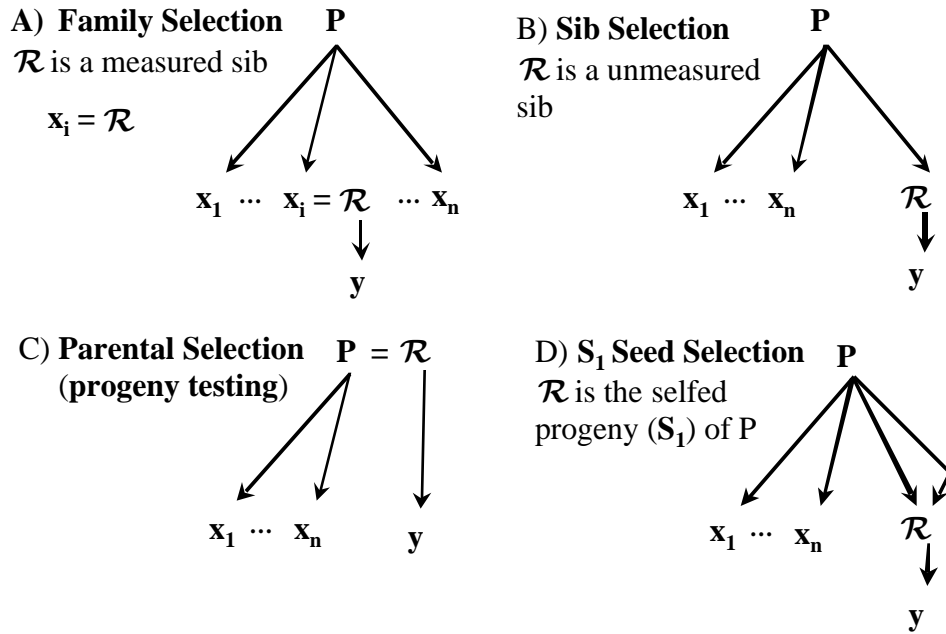
Selection decisions also influence the long-term effective population size (and hence the long-term response, see Chapter 24), with schemes placing more weight on between-family components having smaller effective population size than those placing more weight on within-family components (Chapter 3).

### Variations on the Recombination Unit Under Between-Family Selection

Under either within-family or index selection, measured individuals are the parents for the next generation, forming the recombination unit. By contrast, with between-family selection there are a variety of options for which relatives comprise the recombination unit (Table 21.1, Figure 21.2).

The most straightforward situation is **family selection**, using measured sibs from each chosen family as the parents for the next generation (Figure 21.2A). Under **sib selection**, unmeasured sibs are used to form the next generation (Figure 21.2B). Sib selection is often used for characters that are sex-limited or that cannot be scored without sacrificing the individual. Plants breeders routinely use sib selection in the form of **remnant seeds**. Here, seeds from a parental cross are split into two batches, one is planted and used to assess families, the other held in reserve. Seeds from the elite families are grown and crossed to form the next generation. Under this design, a single cycle of selection takes two generations — one to assess the families and a second to grow up and cross remnant seeds. Given this extra generation, what is the advantage to using remnant seeds? For annual plants, any traits that are expressed during or after flowering can only be directly selected in already pollinated females, with seeds from the best performing plants forming the next generation. Since these plants were pollinated at random, selection has occurred for the seed, but not pollen, parents. By using remnant seeds, one can choose the best families, grow up remnant seeds and allow the resulting plants to randomly intercross. Since both seed and pollen parents have now been selected (through their families), a single cycle of selection using remnant seed has double the response of family selection on seed from open pollinated plants. This doubling of response per cycle exactly counters the extra generation in each

cycle, so that open pollination family selection and sib selection using remnant seed both have the same expected response per generation. One potential advantage is that the extra generation to grow up the remnant seed can be used for selection on other characters. For example, culling those otherwise elite families that show poor disease or insect resistance.



**Figure 21.2.** Under between-family selection, decisions are made on the basis of observations from sibs, while the next generation is formed by crossing relatives  $\mathcal{R}$  of sibs from the chosen families. The measured sibs upon which selection decisions are based are denoted by  $x_1, \dots, x_n$ , while  $y$  denotes a random offspring from a random member  $\mathcal{R}$  from the recombination unit. Different types of relatives can be used as the recombination unit  $\mathcal{R}$ , with a few of the most common types illustrated here. Let  $P$  denote the shared parent(s) of  $x_i$  and  $\mathcal{R}$ . The pedigrees illustrated here all focus on just one parent of  $y$ , with corresponding pedigrees for the other parent. **A) Family selection:**  $\mathcal{R}$  is one of the measured sibs ( $x_i = \mathcal{R}$ ). **B) Sib selection:**  $\mathcal{R}$  is an unmeasured sib. **C) Parental selection** (also known as **progeny testing**):  $\mathcal{R}$  is the parent of the sibs ( $\mathcal{R} = P$ ). **D)  $S_1$  seed selection:**  $\mathcal{R}$  is the selfed progeny of the parent of the sibs.

Another common between-family design is **parental selection** (or **progeny testing**), where  $\mathcal{R} = P$ , the parent of the measured sibs (Figure 21.2C). This design typically involves evaluation of half-sib families with selection on just one sex. In animal breeding this is typically sires, choosing elite males by the performance of their half-sib families. This is greatly facilitated by the use of artificial insemination and frozen semen. The relatively recent ability to clone several domesticated animals (e.g., sheep, Campbell et al. 1996; cattle, Wells et al. 1999a,b; goats, Baguisi et al. 1999) is likely to further increase the importance of progeny testing in animal breeding settings. (The most elaborate, and widely used, extension of progeny testing is BLUP selection wherein the entire pedigree is used for information on selection decisions) Plant breeders typically perform progeny testing using maternal half-sib families (common seed parents). Vegetative propagation (cloning) allows even annual plants to be used as parents in future generations. If the species being selected is **monoecious** (single individuals produce both seed and pollen), one can obtain elite plants for both seed and pollen on the basis of female (seed) performance, and hence select on both sexes.



Finally, with self-compatible species, an alternative to vegetative propagation is the **S<sub>1</sub> seed design** (Figure 21.2D). For each parent, a subset of flowers are selfed to produce S<sub>1</sub> seed and the remainder outcrossed. The outcrossed seed is then grown to produce the sibs in which the trait of interest is assessed. Following selection of the best families, their S<sub>1</sub> seed are grown and the adults crossed to form the next generation. As with remnant seed, a single cycle takes two generations. One advantage of the S<sub>1</sub> design is that one can store seeds, rather than adults. In maize, the S<sub>1</sub> seed designs requires the use of **prolific plant** (those with more than one ear), as (at least) one ear is selfed and (at least) one other outcrossed. Hallauer and Mirana (1981) note that the use of such plants also results in selection for prolificacy, which by itself can increase yield. Another advantage of designs using remnant seed is that traits can be scored over several years before selection, providing the opportunity to select over temporal variation in the environment. As presented in this Chapter, the S<sub>1</sub> design has a random-mated family as the selection one. Obviously, one could collect only S<sub>1</sub> seed from a plant and use some for selection decisions and the rest for future breeding. Such designs are examined in Chapter 20.

**Table 21.1.** Family-based selection schemes using outbred sibs. Families are selected based on the sib values  $z_{i1}, \dots, z_{in}$ .  $\mathcal{R}_i$  denotes a relative of the selected family used to from the next generation. Here  $\bar{z}_{HS}$  and  $\bar{z}_{FS}$  denote the sample means and  $\mu_{HS}$  and  $\mu_{FS}$  the true means of half- and full-sib families, while  $P$  is the parent of the measured sibs and  $z_{ij}$  denotes the  $j$ th measured sib from family  $i$ .

Between-Family Selection	Recombination Unit $\mathcal{R}$	Selection Unit $x$
<b>Family Selection</b>	Measured sib	
Half-sib Family Selection		$\bar{z}_{HS}$
Full-sib Family Selection		$\bar{z}_{FS}$
<b>Sib Selection / Remnant Seed</b>	Unmeasured sib	
Half-sib Sib Selection		$\bar{z}_{HS}$
Full-sib Sib Selection		$\bar{z}_{FS}$
<b>Parental Selection/Progeny testing</b>	Parent $P$	$\bar{z}_{HS}$
<b>S<sub>1</sub> Seed Selection</b>	S <sub>1</sub> Seed of $P$	
Half-sib S <sub>1</sub> Seed Selection		$\bar{z}_{HS}$
Full-sib S <sub>1</sub> Seed Selection		$\bar{z}_{FS}$
<b>Within-Family Selection</b>		
<b>Family Deviation (FD) Selection</b>	Measured Sib	
Half-sib Family Deviation Selection		$z_{ij} - \bar{z}_{HS}$
Full-sib Family Deviation Selection		$z_{ij} - \bar{z}_{FS}$
<b>Strict Within-Family (WF) Selection</b>	Measured Sib	
Half-sib Strict Within-Family Selection		$z_{ij} - \mu_{HS}$
Full-sib Strict Within-Family Selection		$z_{ij} - \mu_{FS}$

## THEORY OF EXPECTED SINGLE-CYCLE RESPONSE

Response is typically given on a *per cycle*, rather than per generation, basis. A cycle begins with choosing the parents  $P$  to form the sib families and ends with the offspring  $y$  formed by crossing members  $\mathcal{R}$  from the recombination unit. The expected response is the difference in the means of these two populations ( $P$  vs.  $y$ ). When comparing the efficiencies of different schemes, response per cycle should be converted to a response per generation (for discrete generations) or per unit time (for overlapping generations).

Our treatment of the theory of response starts by developing several equivalent modifications of the Breeders' Equation (Chapter 12) to accommodate family-based selection. To apply these expressions, we require the **selection unit-offspring covariance**  $\sigma(x, y)$  and the **variance of the selection unit**  $\sigma_x^2$  for various family-based designs. The full development of these variances and covariances is straightforward but involves a bit of bookkeeping. The reader wishing to skip the details can find the final results summarized in Tables 21.3 and 21.4.

### Modifications of the Breeders' Equation for Predicting Family-Based Response

Response is a function of how selection decisions based on the sib families ( $x_1$  and  $x_2$ ) translates into selection on the corresponding parents ( $\mathcal{R}_1$  and  $\mathcal{R}_2$ ) of the offspring  $y$ . Phrased in terms of breeding values, we predict response by using the sib information to predict the breeding values of the parents  $\mathcal{R}$  for the next generation, as under random mating the expected mean of the offspring equaling the mean breeding value of the chosen parents.

Making the standard assumptions that all appropriate regressions are linear (which follows under the infinitesimal model assumptions, see Chapters 14 and 22), the expected response is given by the general form of the Breeders' Equation (12.4a,b),

$$R_y = \frac{\sigma(x_m, y)}{\sigma_{x_m}^2} S_{x_m} + \frac{\sigma(x_f, y)}{\sigma_{x_f}^2} S_{x_f} \quad (21.1a)$$

Here  $x_m$  and  $x_f$  correspond to the selection units associated with the male (sire/pollen) and female (dam/seed) parents ( $\mathcal{R}_m$  and  $\mathcal{R}_f$ ) of the offspring  $y$ . Equation 21.1a allows the male and female parents to be chosen by completely different schemes. For example, sib selection could be used on males and individual selection on females when selecting for a female-limited character (this would be an example of combined selection). The selection unit-offspring covariance  $\sigma(x, y)$  can be directly computed from the pedigree connecting  $P$ , a sib in  $x$ , and  $\mathcal{R}$  through the use of path analysis (LW Appendix 2). The path (or correlation) between selection on the unit  $x_f$  through the female parent  $\mathcal{R}_f$  and its offspring  $y$  is

$$x_f \leftarrow P \rightarrow \mathcal{R}_f \rightarrow y$$

Because the path connecting  $x_f$  and  $y$  is through  $\mathcal{R}_f$ , we often write  $\sigma(x, y | \mathcal{R}_f)$  in place of  $\sigma(x_f, y)$  to remind the reader of this fact. Path(s) connecting  $x_m$  and  $y$  through  $\mathcal{R}_m$  are similarly defined. If  $P$  consists of multiple relatives, each path connecting  $x_i$  and  $\mathcal{R}_i$  (and hence  $y$ ) needs to be counted. For example, if  $x_i$  and  $\mathcal{R}_i$  are full sibs, we must compute the paths through each of the common parents (e.g., Figure 21.3D). If the covariances are the same for both parents, Equation 21.1a simplifies to

$$R_y = \frac{\sigma(x, y)}{\sigma_x^2} S_x \quad (21.1b)$$

where  $S_x = (S_{x_m} + S_{x_f})/2$  is the average selection intensity on the unit(s) leading to the parents and

$$\sigma(x, y) = \sigma(x, y | \mathcal{R}_f) + \sigma(x, y | \mathcal{R}_m) = 2\sigma(x, y | \mathcal{R}) \quad (21.1c)$$

is the covariance between the value of selection unit  $x$  and the offspring  $y$ , counting the paths through both parents ( $\mathcal{R}_m$  and  $\mathcal{R}_f$ ). When covariances are equal, this is twice the single parent covariance,  $\sigma(x, y | \mathcal{R}_1)$ . By analogy with the Breeders' Equation, Equation 21.1b is often written as

$$R_y = h_{x,y}^2 S_x \quad (21.2a)$$

Here the **generalized heritability** of  $y$  given  $x$ ,

$$h_{x,y}^2 = \frac{\sigma(x,y)}{\sigma_x^2} \quad (21.2a)$$

is the slope of the regression of  $y$  on  $x$  (LW Chapter 3). Just as the individual heritability  $h^2$  is the accuracy in using an individual's phenotypic value to predict their breeding value (Chapter 12), the generalized heritability is the accuracy of using the sib data  $x$  to predict the breeding value of  $\mathcal{R}$ .

**Example 21.2.** Consider family selection, wherein the selection unit is the family mean ( $\bar{z}_i$ ) and the recombination unit are sibs from this family. Assuming the covariance between the sib mean and an individual sib is independent of sex, Equation 21.1b gives the response as

$$R_b = \frac{2 \sigma(\bar{z}_i, y | \mathcal{R}_i)}{\sigma^2(\bar{z}_i)} S_b$$

where  $\mathcal{R}_i$  is one of the sibs (say sib  $j$ , so that  $\mathcal{R}_i = x_{ij}$ ). This can be even more compactly written as  $R_b = h_b^2 S_b$ , where the **between-family heritability**  $h_b^2$  is

$$h_b^2 = \frac{2 \sigma(\bar{z}_i, y | \mathcal{R}_i)}{\sigma^2(\bar{z}_i)}$$

Similarly, for selection on within-family deviations, the value of selection unit is  $z_{ij} - \bar{z}_i$ , giving

$$R_{FD} = \frac{2 \sigma(z_{ij} - \bar{z}_i, y | \mathcal{R}_i)}{\sigma^2(z_{ij} - \bar{z}_i)} S_{FD}$$

where  $\mathcal{R}_i = x_{ij}$ . Response can also be expressed in terms of the **family deviations heritability**, with  $R_{FD} = h_{FD}^2 S_{FD}$ , where

$$h_{FD}^2 = \frac{2 \sigma(z_{ij} - \bar{z}_i, y | \mathcal{R}_i)}{\sigma^2(z_{ij} - \bar{z}_i)}$$

Tables 21.3 and 21.4 (below) give expressions for the above variances and covariances.

Other (equivalent) versions of Equations 21.1a and 21.2a appear in the literature. The **selection intensity version** allows for standardized comparisons of different selection schemes. Defining the selection intensity on  $x$  by  $\bar{l}_x = S_x/\sigma_x$ , Equation 21.1a becomes

$$R_y = \frac{\sigma(x_m, y)}{\sigma_{x_m}} \bar{l}_{x_m} + \frac{\sigma(x_f, y)}{\sigma_{x_f}} \bar{l}_{x_f} \quad (21.3a)$$

If the regressions are the same for both parents,

$$R_y = \frac{\sigma(x, y)}{\sigma_x} \bar{l}_x \quad (21.3b)$$

where  $\bar{l}_x = (\bar{l}_{x_m} + \bar{l}_{x_f})/2$  is the average selection intensity. The selection intensity version is frequently written in terms of the selection unit-offspring correlation  $\rho$ ,

$$R_y = \sigma_z \bar{l}_x \rho(x, y) \quad (21.4a)$$

where (counting both parents)  $\rho(x, y) = 2\rho(x, y | \mathcal{R})$ . Equation 21.4a follows immediately from Equation 21.3b by recalling that  $\rho(x, y) = \sigma(x, y)/(\sigma_x \sigma_y)$  and that the trait variance in the offspring  $y$  is just the phenotypic variance of the character ( $\sigma_y^2 = \sigma_z^2$ ). A variant of Equation 21.4a commonly seen in the literature is

$$R_y = \sigma_A \bar{r}_x \rho(x, A_{\mathcal{R}}) \quad (21.4b)$$

where  $\rho(x, A_{\mathcal{R}})$ , the correlation between the selection unit  $x$  and the breeding value of a parent  $\mathcal{R}$ , is referred to as the **accuracy of selection**. The accuracy of individual selection is  $\rho(z_{\mathcal{R}}, A_{\mathcal{R}}) = h$  (the correlation between an individual's phenotypic and breeding values). A particular family-based approach is favored over individual selection if  $x$  is a more accurate predictor of the breeding value of  $\mathcal{R}$  than is  $\mathcal{R}$ 's phenotypic value, i.e.,  $\rho(x, A_{\mathcal{R}}) > h$ . Strictly speaking, Equation 21.4b holds only in the absence of epistasis, while Equations 21.1-21.3 hold for arbitrary epistasis. To see this, recall that the mean value of an offspring is the average of its parental breeding values,  $y = \mu + A_{\mathcal{R}_m}/2 + A_{\mathcal{R}_f}/2 + e_y$ . Hence,

$$\sigma(x, y) = \frac{1}{2}\sigma(x, A_{\mathcal{R}_m}) + \frac{1}{2}\sigma(x, A_{\mathcal{R}_f}) + \sigma(x, e_y)$$

In the absence of epistasis, inbreeding, and shared environmental effects,  $\sigma(x, e) = 0$ . If the regression is the same for both sexes, then  $\sigma(x, y) = \sigma(x_1, A_{\mathcal{R}_1})$ . Recalling that  $\sigma_y = \sigma_z$ ,

$$\rho(x, y) = \frac{\sigma(x, y)}{\sigma_x \sigma_z} = \left( \frac{\sigma_A}{\sigma_z} \right) \frac{\sigma(x_1, A_{\mathcal{R}_1})}{\sigma_x \sigma_A} = h \rho(x, A_{\mathcal{R}_1}) \quad (21.5)$$

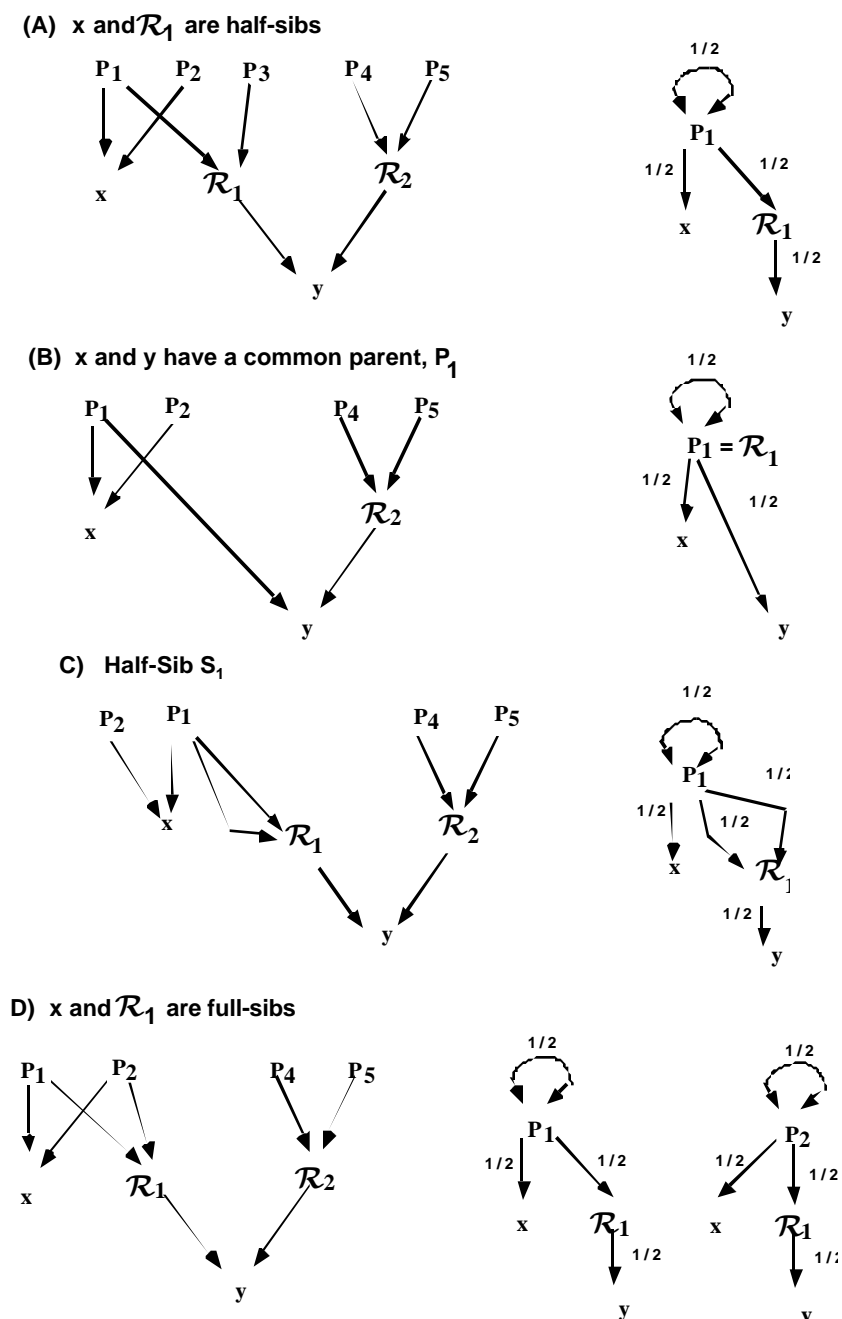
Substitution into Equation 21.4a gives Equation 21.4b (as  $\sigma_z h = \sigma_z (\sigma_A/\sigma_z) = \sigma_A$ ). Equations 21.1-21.4 provide equivalent expressions for computing the expected response. To apply these for any particular scheme, we need to compute the selection unit-offspring covariance  $\sigma(x, y)$  and the variance of the selection unit  $\sigma_x^2$ .

**Table 21.2.** Coefficients of coancestry  $\Theta$  between an offspring  $y$  (of parent  $\mathcal{R}_1$ ) and a member of the selection unit  $x_1$ . Genetic covariances  $\sigma_G(x_1, y)$  are computed assuming no epistasis. Derivations are given in Figure 21.3.

Relationship between $x_1$ and $\mathcal{R}_1$	$\Theta_{x_1 y}$	$\sigma_G(x_1, y) = 2\Theta_{x_1 y} \sigma_A^2$
$x_1 = \mathcal{R}_1$ (the sib is also the parent of $y$ )	1/4	$\sigma_A^2/2$
$x_1$ and $\mathcal{R}_1$ are half-sibs (Figure 21.3A)	1/16	$\sigma_A^2/8$
$x_1$ and $\mathcal{R}_1$ are full-sibs (Figure 21.3D)	1/8	$\sigma_A^2/4$
$\mathcal{R}_1$ is the parent of both $x_1$ and $y$ (Figure 21.3B)	1/8	$\sigma_A^2/4$
$\mathcal{R}_1$ is an $S_1$ offspring of the parent of $x_1$ (Figure 21.3C)	1/8	$\sigma_A^2/4$

### The Selection Unit-Offspring Covariance, $\sigma(x, y)$

Recall that the genetic covariance between two relatives is a function of their **coefficients of coancestry**  $\Theta$  and **fraternity**  $\Delta$  (LW Chapter 7). Ignoring epistasis (for now), the genetic covariance between a particular sib  $x_i$  and  $y$  is  $\sigma_G(x_i, y) = 2\Theta_{x_i y} \sigma_A^2 + \Delta_{x_i y} \sigma_D^2$  (LW Equation 7.12). In the absence of inbreeding (the parents  $\mathcal{R}_1$  and  $\mathcal{R}_2$  are from different, unrelated families),  $\Delta_{xy}$  is zero. Note that this holds even when  $\mathcal{R}_1$  and/or  $\mathcal{R}_2$  are themselves inbred, provided that they are unrelated. For dominance effects to be shared by relatives, there must



**Figure 21.3.** Derivation of the  $\Theta$  values presented in Table 21.2. Pedigrees (left) and associated path diagrams (right) for computing the coefficient of coancestry  $\Theta$  between a measured sib ( $x_1$ ) and an offspring  $y$  from the parent  $\mathcal{R}_1$ .  $P_1$  to  $P_5$  are assumed to be unrelated and non-inbred. **(A):**  $x$  and  $\mathcal{R}_1$  are half-sibs. Taking the product of the path coefficients gives the coefficient of coancestry as  $\Theta_{x_1y} = (1/2)^4 = 1/16$ . **(B):**  $x_1$  and  $y$  are half-sibs, with  $\Theta_{x_1y} = (1/2)^3 = 1/8$ . **(C):**  $\mathcal{R}_1$  is a selfed progeny from the common parent  $P_1$ . Here there are two separate paths between  $x$  and  $y$  (as there are two different routes through  $P_1$ ), giving  $\Theta_{x_1y} = 2 \cdot (1/2)^4 = 1/8$ . **(D):**  $x_1$  and  $\mathcal{R}_1$  are full-sibs. Again there are two paths between  $x_1$  and  $y$  (one through each parent), each being  $(1/2)^4$ , giving a total of  $\Theta_{x_1y} = 2 \cdot (1/2)^4 = 1/8$ .

be paths wherein *both* alleles from an individual in the selection unit are passed onto the offspring  $y$ , which cannot occur if the parents of  $y$  ( $\mathcal{R}_1$  and  $\mathcal{R}_2$ ) are unrelated.

The coefficient of coancestry between  $x_1$  and  $y$  depends upon the relationship between  $\mathcal{R}_1$  and  $x_1$ . The designs covered in Table 21.1 involve four different relationships (Figure 21.2): (i)  $x_1 = \mathcal{R}_1$  (measured sib is a parent of  $y$ ), (ii)  $x_1$  and  $\mathcal{R}_1$  are sibs, (iii)  $\mathcal{R}_1 = P_1$  (the parent of  $x_1$ ), (iv)  $\mathcal{R}_1$  is the selfed-progeny of the parent of  $x_1$ . The path diagrams for computing  $\Theta_{x_1 y}$  for these four relationships are given in Figure 21.3, and Table 21.2 summaries the resulting genetic covariances.

As an example of how the coefficients of coancestry given in Table 21.2 are used, consider family selection. Ignoring epistasis,

$$\begin{aligned}\sigma(\bar{z}_i, y | \mathcal{R}_1 = x_{ij}) &= \sigma\left(\frac{1}{n} \sum_k z_{ik}, y | \mathcal{R}_1 = x_{ij}\right) = \frac{1}{n} \sigma(z_{ij}, y) + \left(1 - \frac{1}{n}\right) \sigma(z_{ik}, y) \\ &= \sigma_A^2 \left(\frac{1/2}{n} + \left(1 - \frac{1}{n}\right) 2\Theta_{z_{ik} y}\right)\end{aligned}\quad (21.6)$$

This follows since the first covariance is for parent and offspring ( $\sigma_A^2/2$ ), while the second covariance follows using the appropriate value of  $2\Theta$  from Table 21.2 (1/8 for half-sibs, 1/4 for full-sibs). Using the results from Table 21.2, expressions for the sib selection, parental selection (progeny-testing), and  $S_1$  seed designs follow using similar logic. These are summarized in Table 21.3.

In much of the animal breeding literature, **Wright's coefficient of relationship**,  $r$ , is usually used in place of  $2\Theta$ . Assuming no inbreeding,  $r = 1/4$  for half-sibs and  $1/2$  for full sibs. Using Wright's coefficient, Equation 21.6 simplifies to

$$\sigma(\bar{z}_i, y | \mathcal{R}_1 = z_{ij}) = r_n \frac{\sigma_A^2}{2} \quad \text{where} \quad r_n = r + \frac{1-r}{n} \quad (21.7a)$$

Considering the paths through both parents ( $\mathcal{R}_1$  and  $\mathcal{R}_2$ ) of  $y$ ,

$$\sigma(\bar{z}_i, y) = 2 \sigma(\bar{z}_i, y | \mathcal{R}_1) = r_n \sigma_A^2 \quad (21.7b)$$

Likewise, the covariance between an individual's family deviation and its offspring's phenotypic value is

$$\sigma(z_{ij} - \bar{z}_i, y | \mathcal{R}_1 = z_{ij}) = \sigma(z_{ij}, y | \mathcal{R}_1) - \sigma(\bar{z}_i, y | \mathcal{R}_1) = (1 - r_n) \frac{\sigma_A^2}{2} \quad (21.8a)$$

as  $\sigma(z_{ij}, y | \mathcal{R}_1)$  is the parent-offspring covariance,  $\sigma_A^2/2$ . Doubling the single-parent contribution gives the total contribution (considering both parents of  $y$ ) as

$$\sigma(z_{ij} - \bar{z}_i, y) = (1 - r_n) \sigma_A^2 \quad (21.8b)$$

The covariance for strict within-family (WF) selection is slightly different, as the appropriate covariance here is  $\sigma(z_{ij} - \mu_i, y)$ , with  $\mu_i$  in place of  $\bar{z}_i$ . The rankings of individuals under WF selection is simply their ranking within each family, while the ranking under FD selection further depends on how much an individual actually deviates from its family mean. Thus the top ranked individuals in two families are always chosen under WF selection, but may not be under FD selection. As a consequence, FD selection is influenced by the observed family mean  $\bar{z}_i$ , while WF selection is a function of the true mean  $\mu_i$  (Dempfle 1975, 1990; Hill et al. 1996).

A few simple rules emerge from Table 21.3. The number  $n$  of measured sibs only influences the covariance for family selection and family deviations selection. Even in these cases its effect is small unless the number of sibs is small. Under sib selection (and family selection ignoring terms of order  $1/n$ ), the selection unit-offspring covariance contributed through one parent is  $\sigma_A^2/8$  for half-sibs and  $\sigma_A^2/4$  for full sibs. For parental selection and  $S_1$  seed designs, this covariance is  $\sigma_A^2/4$  (independent of whether full or half sibs are used). The covariance under WF within-family selection (and FD selection ignoring terms of order  $1/n$ ) is  $3\sigma_A^2/8$  for half-sibs and  $\sigma_A^2/4$  for full-sibs.

**Table 21.3.** Summary of the covariances between the selection unit and one parent ( $\mathcal{R}_1$ ) from the recombination unit. As given by Equation 21.7a,  $r_n = r + (1 - r)/n$ .

---

**Between-family Selection:**

**Family selection** ( $\mathcal{R}_1$  is a measured sib from family  $i$ )

$$\sigma(\bar{z}_i, y | \mathcal{R}_1) = r_n (\sigma_A^2/2) = \begin{cases} (1 + 3/n) (\sigma_A^2/8) & \text{half-sibs} \\ (1 + 1/n) (\sigma_A^2/4) & \text{full-sibs} \end{cases}$$

**Sib Selection / Remnant Seed** ( $\mathcal{R}_1$  is a unmeasured sib from family  $i$ )

$$\sigma(\bar{z}_i, y | \mathcal{R}_1) = r (\sigma_A^2/2) = \begin{cases} \sigma_A^2/8 & \text{half-sibs} \\ \sigma_A^2/4 & \text{full-sibs} \end{cases}$$

**Parental Selection / Progeny testing** ( $\mathcal{R}_1$  is a the parent of the sibs)

$$\sigma(\bar{z}_i, y | \mathcal{R}_1) = \sigma_A^2/4$$

**$S_1$  seed design** ( $\mathcal{R}_1$  is a selfed progeny of the parent of the sibs)

$$\sigma(\bar{z}_i, y | \mathcal{R}_1) = \sigma_A^2/4$$

**Within-family Selection:**

**Selection on Family Deviations (FD)**

$$\sigma(z_{ij} - \bar{z}_i, y | \mathcal{R}_1) = (1 - r_n) (\sigma_A^2/2) = \begin{cases} (1 - 1/n) (3/8) \sigma_A^2 & \text{half-sibs} \\ (1 - 1/n) (\sigma_A^2/4) & \text{full-sibs} \end{cases}$$

**Strict Within-family Selection (FW)**

$$\sigma(z_{ij} - \mu_i, y | \mathcal{R}_1) = (1 - r) (\sigma_A^2/2) = \begin{cases} (3/8) \sigma_A^2 & \text{half-sibs} \\ \sigma_A^2/4 & \text{full-sibs} \end{cases}$$


---

**Variance of the Selection Unit,  $\sigma_x^2$**

The variance  $\sigma_x^2$  of the selection unit is a function of the within- and between-family variances. To obtain these, we assume that the total environmental value is partitioned as  $E = E_c + E_s$ , a common family effect ( $E_c$ ) plus an individual-specific effect ( $E_s$ ). This decomposes the total environmental variances into between- and within-family components,  $\sigma_E^2 = \sigma_{E_c}^2 + \sigma_{E_s}^2$ . When families are replicated over plots/environments, the environmental variance contains additional structure and is usually partitioned into further components (Equations 21.41-21.40; Chapter 33).

For families with a large number of sibs, the between-family variance  $\sigma_b^2$  (the variance in the expected family means  $\mu_i$ ) is

$$\sigma_b^2 = \sigma^2(\mu_i) = \sigma_{GF}^2 + \sigma_{Ec}^2$$

where  $\sigma_{GF}^2$  is the between-family genetic variance (the variance in the expected mean genotypic value of each family), and is developed below (Equations 21.11a, 21.26a). Likewise, the within-family variance when the number of sibs is large is

$$\sigma_w^2 = \sigma^2(z_{ij} - \mu_i) = \sigma_{Gw}^2 + \sigma_{Es}^2$$

where  $\sigma_{Gw}^2$  is the within-family genetic variance (Equations 21.11b, 21.26b). Note that  $\sigma_b^2$  and  $\sigma_w^2$  are functions of  $\mu_i$ , while the variance of the selection unit usually requires the variances about the *observed* mean  $\bar{z}_i$  of each family. Replacing  $\mu_i$  by  $\bar{z}_i$  results in a slight inflation of the between-family variance and a slight reduction in the within-family variance (this is formally shown below in Example 21.3). With  $n$  sibs in each family, the between-family variance based on the observed means becomes

$$\sigma^2(\bar{z}_i) = \sigma^2(\mu_i + \bar{e}_i) = \sigma_b^2 + \sigma_w^2/n$$

namely the between-family variance plus the variance  $\sigma^2(\bar{e}_i)$  in the error from estimating  $\mu_i$  from  $\bar{z}_i$ . Since the total variance is the sum of the between- plus within-family variances ( $\sigma_b^2 + \sigma_w^2$ ), the within-family variance is corresponding reduced to

$$\sigma^2(z_{ij} - \bar{z}_i) = (1 - 1/n)\sigma_w^2$$

Thus

$$\sigma^2(\bar{z}_i) = \sigma_{GF}^2 + \sigma_{Ec}^2 + \frac{\sigma_{Gw}^2 + \sigma_{Es}^2}{n} \quad (21.9a)$$

In the animal breeding literature, this equation is often more compactly written in terms of  $t$ , the phenotypic correlation between sibs (the **intraclass correlation coefficient**, see LW Chapter 7). The phenotypic covariance between sibs equals  $t\sigma_z^2 = \sigma_b^2 = \sigma_{GF}^2 + \sigma_{Ec}^2$  (Example 21.3), implying that

$$\sigma^2(\bar{z}_i) = t_n \sigma_z^2 \quad (21.9b)$$

where

$$t_n = t + \frac{1-t}{n} \quad (21.9c)$$

Likewise, the within-family variance is

$$\sigma^2(z_{ij} - \bar{z}_i) = \left(1 - \frac{1}{n}\right) (\sigma_{Gw}^2 + \sigma_{Es}^2) \quad (21.10a)$$

which is usually written as

$$\sigma^2(z_{ij} - \bar{z}_i) = (1 - t_n) \sigma_z^2 \quad (21.10b)$$

**Example 21.3.** To obtain the within- and between-family variances for families with  $n$  sibs, decompose the phenotypic value of the  $j$ th individual from family  $i$  as

$$z_{ij} = G_{ij} + E_{ij} = \mu + GF_i + Gw_{ij} + Ec_i + Es_{ij}$$



where the genotypic value  $G_{ij} = \mu + GF_i + Gw_{ij}$  has both a family genotypic effect  $GF_i$  (the expected genotypic value for a random sib from that family) and a deviation  $Gw_{ij}$  of the  $j$ th individual's genotypic value from its family average. The environmental value is similarly decomposed, with  $E_{ij} = Ec_i + Es_{ij}$ , an environmental effect  $Ec_i$  common to family  $i$  and an environmental effect  $Es_{ij}$  unique to the  $j$ th individual from this family.  $GF_i + Ec_i = b_i$  are the effects common to a family, giving the between-family variance as

$$\sigma_b^2 = t \sigma_z^2 = \sigma_{GF}^2 + \sigma_{Ec}^2$$

The equality  $\sigma_b^2 = t \sigma_z^2$  follows from the ANOVA identity that the between-group variance equals the covariance between group members (e.g., LW Chapter 18).

Likewise,  $Gw_{ij} + Es_{ij} = w_{ij}$  are the within-family effects, giving the within-family variance (in an infinitely-large family) as

$$\sigma_w^2 = (1 - t) \sigma_z^2 = \sigma_{Gw}^2 + \sigma_{Es}^2$$

The equality  $\sigma_w^2 = (1 - t) \sigma_z^2$  again follows from ANOVA theory, as the total variance equals the between- plus within-group variances,  $\sigma_z^2 = \sigma_b^2 + \sigma_w^2 = t \sigma_z^2 + \sigma_w^2$ . Using these results, for a family of size  $n$ ,

$$\bar{z}_i = \frac{1}{n} \sum_{j=1}^n z_{ij} = \mu + GF_i + Ec_i + \sum_{j=1}^n \frac{(Gw_{ij} + Es_{ij})}{n}$$

Recalling that the  $Es_{ij}$  and  $Gw_{ij}$  (being deviations from the mean) are uncorrelated with each other gives

$$\begin{aligned} \sigma^2(\bar{z}_i) &= (\sigma_{GF}^2 + \sigma_{Ec}^2) + \frac{1}{n^2} \sum_{j=1}^n (\sigma_{Gw}^2 + \sigma_{Es}^2) = \sigma_b^2 + \frac{n \sigma_w^2}{n^2} \\ &= \left( t + \frac{1-t}{n} \right) \sigma_z^2 = t_n \sigma_z^2 \end{aligned}$$

Now consider the variance of within-family deviations. Recalling the expression for the variance of a sum (LW Equation 3.11a),

$$\sigma^2(z_{ij} - \bar{z}_i) = \sigma_z^2 + \sigma^2(\bar{z}_i) - 2\sigma(z_{ij}, \bar{z}_i)$$

To refine this further, first note that  $\sigma^2(\bar{z}_i) = t_n \sigma_z^2$ , and that the covariance term simplifies to

$$\sigma(z_{ij}, \bar{z}_i) = \frac{1}{n} \left( \sigma(z_{ij}, z_{ij}) + \sum_{k \neq j} \sigma(z_{ij}, z_{ik}) \right) = \frac{\sigma_z^2}{n} + \frac{n-1}{n} t \sigma_z^2 = t_n \sigma_z^2$$

as  $\sigma(z_{ij}, z_{ik}) = t \sigma_z^2$ . Thus the variance of within-family deviations reduces to

$$\sigma^2(z_{ij} - \bar{z}_i) = \sigma_z^2 + t_n \sigma_z^2 - 2 t_n \sigma_z^2 = (1 - t_n) \sigma_z^2$$

---

Table 21.4 gives these family variances in terms of genetic and environmental variance components. These follow by expressing the within- and between-family genetic variances in terms of additive and dominance variance components. Recalling from Analysis of Variance

(ANOVA) theory that the between-group variance equals the within-group covariance (e.g., LW Chapter 18), the between-family component  $\sigma_{GF}^2$  equals the genetic covariances between sibs. Ignoring epistasis,

$$\sigma_{GF}^2 = \begin{cases} \frac{1}{4}\sigma_A^2 & \text{half-sibs} \\ \frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2 & \text{full-sibs} \end{cases} \quad (21.11a)$$

Since the total genetic variance  $\sigma_G^2$  equals the between-family genetic variance plus the within-family variance,

$$\sigma_{Gw}^2 = \sigma_G^2 - \sigma_{GF}^2 = \begin{cases} \frac{3}{4}\sigma_A^2 + \sigma_D^2 & \text{half-sibs} \\ \frac{1}{2}\sigma_A^2 + \frac{3}{4}\sigma_D^2 & \text{full-sibs} \end{cases} \quad (21.11b)$$

where we have again ignored epistasis. When epistasis is present, Equations 21.26a,b replace 21.11a,b.

**Table 21.4.** Within- and between-family variances as functions of the genetic and environmental variance components. Epistasis is assumed absent and the environmental value is assumed to equal  $E = E_c + E_s$ , a common family plus an individual-specific value.  $n$  is the number of measured sibs.

---

Half-sib between-family variance

$$\sigma^2(\bar{z}_{HS}) = \frac{\sigma_A^2}{4} + \frac{(3/4)\sigma_A^2 + \sigma_D^2 + \sigma_{E_s}^2}{n} + \sigma_{E_c(HS)}^2$$

Full-sib between-family variance

$$\sigma^2(\bar{z}_{FS}) = \frac{\sigma_A^2}{2} + \frac{\sigma_D^2}{4} + \frac{(1/2)\sigma_A^2 + (3/4)\sigma_D^2 + \sigma_{E_s}^2}{n} + \sigma_{E_c(FS)}^2$$

Half-sib with nested full-sibs between-family variance

( $n_f$  females per male,  $n_s$  offspring/female,  $n = n_f n_s$ )

$$\sigma^2(\bar{z}_{HS(FS)}) = \frac{\sigma_A^2}{4} \left(1 + \frac{1}{n_f} + \frac{2}{n}\right) + \frac{\sigma_D^2}{4n_f} \left(1 + \frac{3}{n_s}\right) + \frac{\sigma_{E_s}^2}{n} + \frac{\sigma_{E_c(FS)}^2}{n_f} + \sigma_{E_c(HS)}^2$$

Half-sib within-family variance

$$\sigma^2(z_{ij} - \bar{z}_i | HS) = \left(1 - \frac{1}{n}\right) \left(\frac{3}{4}\sigma_A^2 + \sigma_D^2 + \sigma_{E_s}^2\right)$$

Full-sib within-family variance

$$\sigma^2(z_{ij} - \bar{z}_i | FS) = \left(1 - \frac{1}{n}\right) \left(\frac{1}{2}\sigma_A^2 + \frac{3}{4}\sigma_D^2 + \sigma_{E_s}^2\right)$$


---

Finally, under a **nested sib design** (the North Carolina Design I of Comstock and Robinson 1948) one sex (typically a sire/pollen plant) is mated to each of  $n_f$  dams/seed plants, each of which produces  $n_s$  sibs for a total of  $n = n_f n_s$  sibs. The expression in Table 21.4 for the between-family variance for this type of family follows using similar logic as in Example 21.3, with

$$\sigma^2(\bar{z}_{HS(FS)}) = \sigma_{GF(HS)}^2 + \frac{\sigma_{G(f|m)}^2}{n_f} + \frac{\sigma_{Gw(FS)}^2}{n_s n_f} + \sigma_{E_c(HS)}^2 \quad (21.12a)$$

where  $\sigma_{G(f|m)}^2$ , the genetic variances of females nested within males, is

$$\sigma_{G(f|m)}^2 = \sigma_{GF(FS)}^2 - \sigma_{GF(HS)}^2 = \frac{\sigma_A^2 + \sigma_D^2}{4} \quad (21.12b)$$

When epistasis is present, Equation 21.26a (below) provides the appropriate additional genetic variance terms in  $\sigma_{G(f|m)}^2$ . The between-family variance under a nested design is bounded below by the half-sib variance ( $n_f = n$  and  $n_s = 1$ ) and above by the full-sib variance ( $n_f = 1$  and  $n_s = n$ ).

## RESPONSE FOR PARTICULAR DESIGNS

The formal development of the response equations for any particular design follows from the generalized breeders' equation (21.1-21.4), using the appropriate variances (Table 21.4) and covariances (Table 21.3). Results for a number of standard between- and within-family designs are developed below, with family index selection examined at the end of the chapter.

### Overview of Between- and Within-Family Response

The selection response for a particular family-based scheme depends on how the additive-genetic (breeding value) and total (phenotypic) variances are partitioned within and between families. When the number of sibs per family ( $n$ ) is large, these variances are partitioned as

	Within-family	Between-family
Breeding values	$(1 - r) \sigma_A^2$	$r \sigma_A^2$
Phenotypic values	$(1 - t) \sigma_z^2$	$t \sigma_z^2$

where  $t$  and  $r$  are the phenotypic and additive-genetic correlations between sibs ( $r = 1/4$  for half-sibs,  $1/2$  for full-sibs). When the number of measured sibs within each family is small,  $t_n = t + (1 - t)/n$  replaces  $t$  and  $r_n$  (similarly defined) replaces  $r$ . Since the response to selection depends on the ratio of the available additive genetic variance to the phenotypic variance, the response to between-family selection is of the form

$$R_b = \frac{r_n \sigma_A^2}{t_n \sigma_z^2} S = \sigma_A \left( \frac{\sigma_A}{\sigma_z} \right) \left( \frac{r_n}{\sqrt{t_n}} \right) \left( \frac{S}{\sqrt{t_n} \sigma_z} \right) = \sigma_A h \frac{r_n}{\sqrt{t_n}} \bar{i} \quad (21.13a)$$

Equation 21.13a is the exact expression for family selection and is due to Lush (1947). The response under other designs (e.g., sib and parental selection,  $S_1$  seed) are very similar (see below).

Likewise, the response to within-family selection is a function of the within-family additive-genetic and phenotypic variances, leading us to expect that response is in the form of

$$R_{FD} = \frac{(1 - r_n) \sigma_A^2}{(1 - t_n) \sigma_z^2} S = \sigma_A h \frac{1 - r_n}{\sqrt{1 - t_n}} \bar{i} \quad (21.13b)$$

Indeed, this is the exact expression for selection on family deviations (FD), while the response under strict within-family (WF) selection is given by replacing  $r_n$  and  $t_n$  by  $r$  and  $t$ .

Equations 21.13a and 21.13b are the standard response equations that appear in much of the elementary animal breeding literature. The use of  $r$  and  $t$  allows these results to be presented in a very compact fashion. When the design is more complicated, such as replication of families or the use of nested sibs families, the response can not longer be cleanly expressed in terms of  $t$ , and expressions are given in terms of variance components.

**Between-Family Selection: Different Schemes**

Here the selection unit is  $\bar{z}$ , the mean of a half-, full-, or nested-sib family. The type of sib family, together with the relatives used to produce the next generation, specifies each particular between-family design (Table 21.1). Tables 21.3 and 21.4 and Equation 21.3a gives the response to a single cycle of selection as

$$R_b = \frac{\Theta}{\sqrt{t_n}} \frac{\sigma_A}{2} h (\bar{t}_{x_m} + \bar{t}_{x_f}) = \frac{\Theta}{\sigma(\bar{z})} \frac{\sigma_A^2}{2} (\bar{t}_{x_m} + \bar{t}_{x_f}) \quad (21.14a)$$

The left equality holds when sib families are not nested and families are not replicated, while the rightmost expression is completely general (using  $\sigma^2(\bar{z})$  in place of  $t_n \sigma_z^2$ ). The selection unit-offspring covariance is  $\Theta \sigma_A^2/2$ , where

$$\Theta = \begin{cases} r_n = r + (1-r)/n & \text{family selection} \\ r & \text{sib selection} \\ 1/2 & \text{parental, } S_1 \text{ seed selection} \end{cases} \quad (21.14b)$$

Recall that these different values arise because  $r$  is the genetic correlation among sibs, and that parental and  $S_1$  selection correspond to the case where  $r = 1/2$ . Under strict sib selection, no measured individual is the parent and hence all the correlations between an individual in the selection unit and an individual forming the parent for the next generation are the same (namely  $r \sigma_A^2$ ). Under family selection, one of the  $n$  measured individual sibs is also the parent, and hence has a genetic correlation of  $\sigma_A^2/2$  while the other  $n - 1$  are simply sibs with genetic correlations of  $r \sigma_A^2$ .

The variance of the selection unit  $\sigma^2(\bar{z}) = t_n \sigma_z^2$  depends only on the types of sibs measured and is independent of the types of relatives used to form the next generation. The theory of expected response to between-family selection traces back to Lush's classic 1947 paper, and Equation 21.14 is a generalization of his results. Table 21.5 expresses the response in terms of variance components.

**Table 21.5.** Variance-component expressions of the expected response with between-family selection schemes using outbred sibs. The number  $n$  of measured sibs is assumed sufficiently large that terms of order  $1/n$  can be ignored (i.e.,  $r_n \simeq r$  and  $t_n \simeq t$ ). We also assume no epistasis and the simple structure  $E = E_c + E_s$  for environmental values.

	Half-sibs	Full-sibs
Family, Sib	$\frac{(\sigma_A^2/8)(\bar{t}_{x_m} + \bar{t}_{x_f})}{\sqrt{\sigma_A^2/4 + \sigma_{E_c(HS)}^2}}$	$\frac{(\sigma_A^2/4)(\bar{t}_{x_m} + \bar{t}_{x_f})}{\sqrt{\sigma_A^2/2 + \sigma_D^2/4 + \sigma_{E_c(FS)}^2}}$
Parental, $S_1$ seed	$\frac{(\sigma_A^2/4)(\bar{t}_{x_m} + \bar{t}_{x_f})}{\sqrt{\sigma_A^2/4 + \sigma_{E_c(HS)}^2}}$	$\frac{(\sigma_A^2/4)(\bar{t}_{x_m} + \bar{t}_{x_f})}{\sqrt{\sigma_A^2/2 + \sigma_D^2/4 + \sigma_{E_c(FS)}^2}}$

Several variants of Equation 21.14 are common in the literature. Noting that  $\sigma_A h = \sigma_z h^2$ , the response is often expressed as

$$R_b = \frac{\Theta}{\sqrt{t_n}} \sigma_z h^2 \bar{t} \quad (21.15a)$$

where  $\bar{i} = (\bar{i}_{x_f} + \bar{i}_{x_m})/2$ . Similarly, response can be expressed in terms of the between-family heritability,

$$R_b = h_{b,\Theta}^2 S, \quad \text{where} \quad h_{b,\Theta}^2 = \frac{\Theta}{t_n} h^2 \quad (21.15b)$$

and  $S = (S_f + S_m)/2$  is the average selection differential on the parents.

Turning now to particular between-family designs, we start with family selection. Here measured sibs (all or a random subset) from the chosen families form the parents for the next generation. To reduce the effects of inbreeding, crosses between sibs from the same family are typically avoided. With family selection, Equation 21.14 becomes

$$R_b = \begin{cases} \frac{(1+3/n)}{\sqrt{t_n(HS)}} \frac{\sigma_A}{8} h (\bar{i}_{x_m} + \bar{i}_{x_f}) & \text{half-sibs} \\ \frac{(1+1/n)}{\sqrt{t_n(FS)}} \frac{\sigma_A}{4} h (\bar{i}_{x_m} + \bar{i}_{x_f}) & \text{full-sibs} \end{cases} \quad (21.15c)$$

as first obtained by Lush (1947). While full-sibs have twice as much usable between-family additive variance ( $\sigma_A^2/2$  vs.  $\sigma_A^2/4$ ), this advantage is reduced because half-sibs have a smaller between-family phenotypic variance than full sibs, with  $t_{HS}/t_{FS} < 1$ . This inequality follows by recalling that  $\sigma^2(\bar{z}) = t\sigma_z^2$  and noting that  $(t_{FS} - t_{HS})\sigma_z^2 = \sigma^2(\bar{z}_{FS}) - \sigma^2(\bar{z}_{HS})$ , where

$$\sigma^2(\bar{z}_{FS}) - \sigma^2(\bar{z}_{HS}) = \frac{\sigma_A^2 + \sigma_D^2}{4} + (\sigma_{Ec(FS)}^2 - \sigma_{Ec(HS)}^2) > 0 \quad (21.16)$$

Given that full-sibs share a common mother (and hence potentially share maternal effects), we expect  $\sigma_{Ec(FS)}^2 \geq \sigma_{Ec(HS)}^2$  and hence  $\sigma^2(\bar{z}_{FS}) > \sigma^2(\bar{z}_{HS})$ . The ratio of response for full- vs. half-sib family selection is thus  $2\sqrt{t_n(HS)/t_n(FS)} < 2$ .

If the character can only be measured after reproduction, females (seed parents) from the chosen families have already been pollinated, and hence selection has occurred on only one sex. Planting these seeds and evaluating the resulting families allows for half-sib selection. Full-sib selection can also be accomplished, but each cycle takes an additional generation. Here seeds from open-pollinated selected females are grown up and controlled crosses made between the offspring from different seed parents to create full-sib families for the next cycle of selection.

**Example 21.4.** Clayton et al. (1957) examined family selection on abdominal bristle number in *Drosophila melanogaster* (LW Figure 14.1). Their estimated intraclass correlations for half- and full-sibs were 0.121 and 0.265, while the estimated additive variance and heritability were 5.59 and 0.52. Hence

$$t_{HS} = 0.121, \quad t_{FS} = 0.265, \quad \sigma_A h = \sqrt{5.59 \cdot 0.52} = 1.70$$

Clayton et al. performed selection in two different settings: (i) the top two of ten half-sib families were saved, and (ii) the top four of 20 full-sib families were saved. The expected selection intensities under these two schemes are  $\bar{i}_{HS} = \bar{i}_{10,2} = 1.27$  and  $\bar{i}_{FS} = \bar{i}_{20,4} = 1.33$ . The family sizes used were 20 half-sibs and 12 full-sibs. Because of the design, there was a one in ten chance that the half-sibs are actually full-sibs, resulting in a slight inflation of  $r$  from 0.25 to 0.275 ( $= 0.25 + 0.5/10$ ). Summarizing:

	Half-sibs	Full-sibs
$r$	0.275	0.5
$n$	20	12
$t_n$	0.165	0.326
$r_n$	0.311	0.542

Equation 21.13a gives the expected response to half-sib family selection as

$$R_b(HS) = \bar{i}_{HS} \cdot (\sigma_A h) \frac{r_n}{\sqrt{t_n}} = 1.27 \cdot 1.70 \frac{0.311}{\sqrt{0.165}} = 1.67,$$

while the expected response to full-sib family selection is

$$R_b(FS) = 1.33 \cdot 1.70 \frac{0.542}{\sqrt{0.326}} = 2.15$$

Clayton et al. obtained slightly different estimated responses (1.33 and 2.02 for half- and full-sibs, respectively). This occurred because they used  $R = h_b^2 S_b$  with  $S_b = \sigma_b \bar{i}$  computed by taking the observed between-family variance  $\sigma_b^2$  (in place of the estimates  $\sigma_A^2$ ,  $t$ , and  $h^2$ ). The observed responses (averaged over the first five generations) were 1.38 and 0.94 for up- and down-selected half-sibs, and 1.62 and 1.36 for up- and down-selected full sibs. The authors noticed a fairly sizable reduction in the estimated additive variance during generations two through four, and this (in addition to sampling error) likely accounts for the discrepancy between observed and predicted response.

Under sib selection, unmeasured sibs from each chosen family are used to form the next generation. The most common response equation for sib selection in the literature, due to Robertson (1955), is

$$R_{sib} = \bar{i} \sigma_A h \frac{n r}{\sqrt{n(1 + [n-1]t)}} \quad (21.17)$$

where  $\bar{i}$  denotes the average selection intensity on both parents. Equation 21.17 follows from Equation 21.15a, by noting that  $\Theta = r$  for sib selection and using  $\sigma_A h$  in place of  $\sigma_z h^2$ . As mentioned, the use of remnant seed is a variant of sib selection, which allows for selection on both sexes of parents, at the expense of an extra generation per cycle. As mentioned, this approach can be useful for traits that are manifested after pollination (such as yield).

Under parental selection (progeny testing), parents are chosen based on the performance of a trial set of their offspring. Typically half-sib families are used and selection is on a single sex. In this case, the expected response is

$$R_{pt} = \frac{\sigma_A/4}{\sqrt{t_n(HS)}} h \bar{i} \quad (21.18a)$$

In monoecious species, the expected response is double that given by Equation 21.18a if one uses the selected parents for both seed and pollen. The use of maternal half-sib families (as commonly occurs in plant breeding) is expected to inflate  $t(HS)$  relative to paternal half-sibs (and hence reduce response) as the common family environmental effects can be rather significant due to maternal effects.

If males (sires or pollen plants) are progeny-tested using a nested sib design, wherein each male is crossed to  $n_f$  females (dams or seed plants) each of which has  $n_s$  sibs, the appropriate between-family variance is given in Table 21.4, and the response becomes

$$\begin{aligned} R_{pt} &= h \bar{i} \frac{\sigma_A^2/4}{\sqrt{\sigma_{GF(HS)}^2 + \sigma_{G(f|m)}^2/n_f + \sigma_{GW(FS)}^2/n_f n_s + \sigma_{Ec(HS)}^2}} \\ &= h \bar{i} \frac{\sigma_A^2/4}{\sqrt{\frac{\sigma_A^2}{4} \left(1 + \frac{1}{n_f} + \frac{2}{n_f n_s}\right) + \frac{\sigma_D^2}{4n_f} \left(1 + \frac{3}{n_s}\right) + \frac{\sigma_{Es}^2}{n_f n_s} + \sigma_{Ec(HS)}^2}} \quad (21.18b) \end{aligned}$$

For progeny testing of females using a nested design, the roles of males and females are exchanged in the above expression. Since  $\sigma^2(\bar{z}_{HS}) \leq \sigma^2(\bar{z}_{HS(FS)}) \leq \sigma^2(\bar{z}_{FS})$ , the response using a progeny test is intermediate to that for schemes using half- or full-sibs. All above comments for parental selection equally apply to the  $S_1$  seed design, as the expected response is the same.

### Between-Family Selection: Which Scheme is Best?

Given the number of options for between-family selection, which scheme should be used? Biological and/or economic restriction may preclude certain designs and make others more feasible. These logistical concerns aside, there are three issues that must be weighted: (i) cycle time versus selection on one or both sexes, (ii) performance evaluation using half- vs. full-sib families (the value of  $t_n$ , and more generally  $\sigma^2[\bar{z}]$ ), and (iii) choice of relatives for the recombination unit (the value of  $\Theta$  in Equation 21.14a). As mentioned above, a common reason for using a two-generation cycle (e.g., remnant seed) is the inability to select on both sexes. In such cases, doubling the cycle time is countered by selection on both sexes doubling the response per cycle, giving both approaches the same rate of progress on a per-generation basis. In many cases, a multigeneration method is used because selection on other characters beside the primary one of interest is also performed in one (or both) generations of the cycle.

The second choice is the type of family. While the type of sibs changes the value of  $\Theta$  under family- and sib-selection, it does not influence relatives under parental or  $S_1$  selection. Indeed, for these last two designs it is more efficient to use half-sib families, as the ratio of response to parental half-sib versus parental full-sib is  $t(FS)/t(HS) > 1$ .

Provided the same type of families (half-, full-, nested-sibs) are measured, choosing relatives that increase the recombination unit-offspring covariance (by increasing  $\Theta$ ) increases the expected response. For half-sib families, parental and  $S_1$  selection gives twice the response per cycle as sib or family selection (assuming the same number of sexes are selected in comparisons). With full-sibs, Table 21.5 shows that, given the same selection intensity, the response per cycle under all four methods is the same. While the response to selection using full-sib families is greater than that of family or sib selection using half-sibs, the use of full-sibs does not result in a doubling the response as  $2\sqrt{t_n(HS)/t_n(FS)} < 2$ . This less than two-fold increase in response per cycle using full-sibs is thus not sufficient to cover the cost of the extra generation often required to create full-sib families.

Once one has chosen a particular design, there is also the issue of allocation of the number of sibs per family  $n$  given constraints on the total number of sibs  $N$  measured. One increases the accuracy of the method by increasing the number of sibs per family, but does so by decreasing the selection intensity. Robertson (1957, 1960), Rendel (1959), and Lindgren et al. (1997) have examined this problem of optimal family size. To maximize the response, the breeder usually has two fixed constraints: the total number of sibs  $N$  examined and the number  $n_p$  of parents (or families) used to form the next generation. A low  $n_p$  increases inbreeding and thus not only invites inbreeding depression, but also reduces the eventual long term response (Chapter 24). For fixed  $n_p$  and  $N$ , the goal is to find the number of sibs  $n$  per family that maximizes response. Noting that  $\sigma_z h^2$  is fixed, while  $n_p = mp$  and that  $m = N/n$ , Equation 21.15a shows that the single-generation response is maximized by maximizing

$$\Theta \frac{\bar{l}_{(N/n, n_p)}}{\sqrt{t_n}}$$

with respect to  $n$ . With the exception of family selection (where  $\Theta = r_n$ )  $\Theta$  is a fixed constant (with respect on  $n$ ). Maximizing the long term response (or more generally the expected response after  $k > 1$  generations) also needs to consider differences in the effective populations sizes. This is examined in Chapter 24.

### Within-Family Selection

Within-family selection chooses individuals based on their relative performance within families. Under family-deviations (FD) selection, individuals with the largest family deviations are chosen, independent of which family they come from. In contrast, strict within-family (WF) selection chooses the largest individuals from each family, independent of how much they actually deviate from their family means. Suppose in family one the deviations are 4, 3, and  $-7$ , while the deviations in family two are 1, 0,  $-1$ . If we select the upper 1/3, then under WF selection, the top individual from each family are chosen, while under FD selection, two individuals from family one, and none from family two, are chosen. The result of this rather subtle difference is that FD selection is influenced by the observed mean  $\bar{z}_i$ , while WF selection is not. Family deviations and strict within-family selection have been confused in the literature, with the correct expression for WF selection due to Dempfle (1975, 1990) and Hill et al. (1996). Since WF selection ensures equal representation of families, while FD selection does not, WF selection has a larger effective population size and hence an expected larger long-term response (Chapter 24).

Under family-deviations (FD) selection, the selection unit is the value of an individual's within-family deviation,  $z_{ij} - \bar{z}_i$ . Using the results from Table 21.3 and 21.4, Equation 21.1 gives the expected response as

$$\begin{aligned} R_{FD} &= \frac{\sigma(z_{ij} - \bar{z}_i, y | \mathcal{R}_m)}{\sigma(z_{ij} - \bar{z}_i)} \bar{t}_{x_m} + \frac{\sigma(z_{ij} - \bar{z}_i, y | \mathcal{R}_f)}{\sigma(z_{ij} - \bar{z}_i)} \bar{t}_{x_f} \\ &= \frac{1 - r_n}{\sqrt{1 - t_n}} \sigma_A h \left( \frac{\bar{t}_{x_m} + \bar{t}_{x_f}}{2} \right) \end{aligned} \quad (21.19)$$

The last equality follows from  $\sigma_A^2 / \sigma_z = \sigma_A h$ .

Under strict within-family (WF) selection, individuals are chosen entirely on their rank within each family, resulting in the observed mean  $\bar{z}_i$  being replaced by the true (and unobserved) mean  $\mu_i$ . The response becomes

$$\begin{aligned} R_{WF} &= \frac{\sigma(z_{ij} - \mu_i, y | \mathcal{R}_m)}{\sigma(z_{ij} - \mu_i)} \bar{t}_{x_m} + \frac{\sigma(z_{ij} - \mu_i, y | \mathcal{R}_f)}{\sigma(z_{ij} - \mu_i)} \bar{t}_{x_f} \\ &= \frac{1 - r}{\sqrt{1 - t}} \sigma_A h \left( \frac{\bar{t}_{x_m} + \bar{t}_{x_f}}{2} \right) \end{aligned} \quad (21.20)$$

Noting that

$$\frac{1 - r_n}{\sqrt{1 - t_n}} = \frac{(1 - 1/n)(1 - r)}{\sqrt{(1 - 1/n)(1 - t)}} = \frac{1 - r}{\sqrt{1 - t}} \sqrt{1 - \frac{1}{n}}$$

it follows that

$$R_{FD} = R_{WF} \frac{\bar{t}_{FD}}{\bar{t}_{WF}} \sqrt{1 - \frac{1}{n}} \quad (21.21)$$

Thus, when the number of measured sibs in each family is modest to large (so that the selection intensities are essentially equal,  $\bar{t}_{FD} \simeq \bar{t}_{WF}$ ), the difference between the expected responses under WF vs. FD selection is very small. Equation 21.20 gives resulting response for strict within-family selection using half- and full-sib families as

$$R_{WF} = \begin{cases} \frac{(3/8) \sigma_A^2}{\sqrt{1 - t(HS)}} h (\bar{t}_{x_m} + \bar{t}_{x_f}) & \text{half-sibs} \\ \frac{(1/4) \sigma_A^2}{\sqrt{1 - t(FS)}} h (\bar{t}_{x_m} + \bar{t}_{x_f}) & \text{full-sibs} \end{cases} \quad (21.22a)$$



Expressed in terms of variance components,

$$R_{WF} = \begin{cases} \frac{(3/8)\sigma_A^2}{\sqrt{(3/4)\sigma_A^2 + \sigma_D^2 + \sigma_{E_s}^2}} (\bar{t}_{x_m} + \bar{t}_{x_f}) & \text{half-sibs} \\ \frac{(1/4)\sigma_A^2}{\sqrt{\sigma_A^2/2 + (3/4)\sigma_D^2 + \sigma_{E_s}^2}} (\bar{t}_{x_m} + \bar{t}_{x_f}) & \text{full-sibs} \end{cases} \quad (21.22b)$$

For half-sibs, the within-family additive variance is  $(3/4)\sigma_A^2$ , only half of which is passed from parent to offspring, giving the  $(3/8)\sigma_A^2$  term. For full-sibs, the within-family additive variance is  $(1/2)\sigma_A^2$ , again only half of which is passed onto offspring, giving  $\sigma_A^2/4$ .

The **within-family heritability**,  $h_w^2$ , is the same under both FD and WF within-family selection as

$$\frac{1 - r_n}{1 - t_n} = \frac{(1 - 1/n)(1 - r)}{(1 - 1/n)(1 - t)} = \frac{(1 - r)}{(1 - t)}$$

Hence,

$$h_w^2 = \frac{2\sigma(z_{ij} - \bar{z}_i, y | \mathcal{R}_1)}{\sigma^2(z_{ij} - \bar{z}_i)} = \frac{(1 - r_n)\sigma_A^2}{(1 - t_n)\sigma_z^2} = \frac{(1 - r)}{(1 - t)} h^2 \quad (21.23)$$

**Example 21.5.** Using the data of Clayton et al. (1957) from Example 21.4, what is the expected response under the two within-family selection schemes? Suppose strict within-family (WF) selection was performed on the full-sib families, with the upper 20 percent chosen from each family (the top four of the 20 measured sibs). Correcting for finite population size, the expected selection intensity is  $\bar{t}_{(20,4)} = 1.33$  and from Equation 21.20 the predicted response is

$$R_{WF} = \bar{t} \cdot (\sigma_A h) \frac{1 - r}{\sqrt{1 - t}} = 1.33 \cdot 1.70 \frac{1 - 0.275}{\sqrt{1 - 0.121}} = 1.75$$

Using within-family deviations (FD), selecting the uppermost 20 percent gives a corrected selection intensity of 1.39, and Equation 21.19 gives the predicted response as

$$R_{FD} = \bar{t} \cdot (\sigma_A h) \frac{1 - r_n}{\sqrt{1 - t_n}} = 1.39 \cdot 1.70 \frac{1 - 0.311}{\sqrt{1 - 0.165}} = 1.78$$

The selection intensity values used here can be further corrected to account to correlations among sibs and we do so later in the chapter.

### Realized Heritabilities

By analogy with individual selection, one can estimate the realized heritability (Chapter 16) associated with a particular family-based scheme from the ratio of observed response to selection differential,

$$\hat{h}_{R,x}^2 = \frac{R_x}{S_x} \quad (21.24a)$$

Falconer and Latyszewski (1952) used this approach to estimate a realized within-family heritability for response to selection on body size in mice. These authors computed the standard error of this estimate by noting that

$$\sigma^2 \left( \hat{h}_{R,wf}^2 \right) = \sigma^2 \left( \frac{R_{wf}}{S_{wf}} \right) = \frac{\sigma^2(R_{wf})}{S_{wf}^2} \quad (21.24b)$$

The last equality follows by assuming that the variance in measuring  $S$  can be ignored.

One is faced with several options for estimating the realized heritability when presented with several generations worth of data (Chapters 16–17). The simplest is to use the ratio of total response to total selection differential. Alternatively, cumulative response can be regressed on cumulative selection differential, with the slope being the estimate of the realized heritability. As with mass selection, an appropriately weighted GLS regression must be used to obtain an unbiased estimate of the standard error. Unweighted (OLS) regressions severely underestimate the standard error due to correlations among the residuals (Chapters 16–17). Another concern is that the underlying genetic parameters can change each generation. The general assumption is that these changes are small enough to be ignored, but this may not be a valid assumption in many cases (Chapters 15, 22). Provided the infinitesimal model applies, mixed-model (MM) analysis (Chapter 17) deals with these concerns. MM analysis is completely general and can easily accommodate family-based selection, as the relationship matrix  $\mathbf{A}$  accounts for all appropriate covariances between relatives.

Since between- and within-family heritabilities can be expressed as a function of the individual heritability  $h^2$  (Equations 21.15 and 21.23), we can similarly translate a realized heritability estimate for a particular family-based design into a realized individual heritability. With between-family selection,

$$\hat{h}_R^2 = \left( \frac{t_n}{\Theta} \right) \hat{h}_{R,b}^2 \quad (21.25a)$$

while for within-family selection,

$$\hat{h}_R^2 = \left( \frac{1-t}{1-r} \right) \hat{h}_{R,wf}^2 \quad (21.25b)$$

These expressions apply for a single generation of selection. Additional uncertainty is introduced into the estimate if the sib phenotypic correlation  $t$  is unknown and must itself be estimated. Equations 21.25a,b should be used only with extreme caution when multiple cycles of selection have taken place, as the sib additive-genetic correlation  $r$  increases each successive generation due to inbreeding. These changes in genetic variances also change the phenotypic correlation  $t$ .

### Accounting for Epistasis

The response to within- and between-family selection in the presence of epistasis has been briefly developed by Nyquist (1991), and we expand upon his results here. As with individual selection, additive epistasis contributes to the initial response, but its contribution to the ultimate response rapidly decays with time as recombination breaks up favorable combinations of alleles at different loci (Chapter 13). We first consider the single-generation response and then briefly examine the transient dynamics.

The between-family genetic variance  $\sigma_{GF}^2$  with arbitrary epistasis immediately follows from the genetic covariance between sibs (LW Table 7.2),

$$\sigma_{GF}^2 = \begin{cases} \frac{1}{4}\sigma_A^2 + \frac{1}{16}\sigma_{AA}^2 + \frac{1}{64}\sigma_{AAA}^2 + \cdots & \text{half-sibs} \\ \frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2 + \frac{1}{4}\sigma_{AA}^2 + \frac{1}{8}\sigma_{AD}^2 + \frac{1}{16}\sigma_{DD}^2 + \frac{1}{8}\sigma_{AAA}^2 + \cdots & \text{full-sibs} \end{cases} \quad (21.26a)$$

Likewise, the within-family genetic variance,  $\sigma_{Gw}^2 = \sigma_G^2 - \sigma_{GF}^2$ , becomes

$$\sigma_{Gw}^2 = \begin{cases} \frac{3}{4}\sigma_A^2 + \sigma_D^2 + \frac{15}{16}\sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2 + \frac{63}{64}\sigma_{AAA}^2 + \cdots & \text{half-sibs} \\ \frac{1}{2}\sigma_A^2 + \frac{3}{4}\sigma_D^2 + \frac{3}{4}\sigma_{AA}^2 + \frac{7}{8}\sigma_{AD}^2 + \frac{15}{16}\sigma_{DD}^2 + \frac{7}{8}\sigma_{AAA}^2 + \cdots & \text{full-sibs} \end{cases} \quad (21.26b)$$

The between- and within-family variances,  $\sigma^2(\bar{z}_i)$  and  $\sigma^2(z_{ij} - \bar{z}_i)$ , immediately follow by substituting Equation 21.26 into Equations 21.9a and 21.10a.

The genetic covariance between an individual  $x$  from the selection unit and the offspring  $y$  under epistasis follow using LW Equation 7.12,

$$\sigma_G(x, y) = (2\Theta_{xy}) \sigma_A^2 + (2\Theta_{xy})^2 \sigma_{AA}^2 + \cdots = \sum_{u=1} (2\Theta_{xy})^u \sigma_{A^u}^2$$

We assume  $\Delta_{xy} = 0$ , so that terms involving dominance are not included. Using the values of  $\Theta_{xy}$  from Table 21.2, the parent-offspring covariance is

$$\sigma(\mathcal{R}_1, y) = \frac{\sigma_A^2}{2} + \frac{\sigma_{AA}^2}{4} + \frac{\sigma_{AAA}^2}{8} + \cdots = \sum_{u=1} \left(\frac{1}{2}\right)^u \sigma_{A^u}^2 \quad (21.27a)$$

Table 21.2 shows that  $\Theta_{xy} = 1/16$  and  $1/8$  when  $x$  is a half- or full-sib (respectively) of  $\mathcal{R}$ . Expressed in terms of Wright's coefficient of relationship,

$$\sigma(x_1, y | \mathcal{R}_1) = (r/2)\sigma_A^2 + (r/2)^2\sigma_{AA}^2 + (r/2)^3\sigma_{AAA}^2 + \cdots = \sum_{u=1} \left(\frac{r}{2}\right)^u \sigma_{A^u}^2 \quad (21.27b)$$

Substituting Equation 21.27a and 21.27b into Equation 21.6 gives the covariance for family selection as

$$\begin{aligned} \sigma(\bar{z}, y | \mathcal{R}_1) &= \frac{1}{n} \sum_{u=1} \left(\frac{1}{2}\right)^u \sigma_{A^u}^2 + \left(1 - \frac{1}{n}\right) \sum_{u=1} \left(\frac{r}{2}\right)^u \sigma_{A^u}^2 \\ &= \sum_{u=1} \left(\frac{1}{2}\right)^u (r^u)_n \sigma_{A^u}^2 \end{aligned} \quad (21.28)$$

where  $(r^u)_n = r^u + (1 - r^u)/n$ . For large family size, the coefficient for  $u$ -fold additive epistasis approaches  $r^u/2^u$ , which is the value under sib selection. Taking  $r = 1/2$ , these are also the coefficients for parental and  $S_1$  seed selection. Equation 21.28 gives the single-parent covariance for half-sib family selection as

$$\sigma(\bar{z}_{HS}, y | \mathcal{R}_1) = \left(1 + \frac{3}{n}\right) \frac{\sigma_A^2}{8} + \left(1 + \frac{15}{n}\right) \frac{\sigma_{AA}^2}{64} + \left(1 + \frac{63}{n}\right) \frac{\sigma_{AAA}^2}{512} + \cdots \quad (21.29a)$$

Likewise, the single-parent covariance for full-sib family selection is

$$\sigma(\bar{z}_{FS}, y | \mathcal{R}_1) = \left(1 + \frac{1}{n}\right) \frac{\sigma_A^2}{4} + \left(1 + \frac{3}{n}\right) \frac{\sigma_{AA}^2}{16} + \left(1 + \frac{7}{n}\right) \frac{\sigma_{AAA}^2}{64} + \cdots \quad (21.29b)$$

For sib-selection,  $\sigma(\bar{z}_1, y | \mathcal{R}_1)$  is given directly from Equation 21.27b, and Equations 21.29a,b apply if terms of order  $1/n$  are ignored. For between-family selection using parental selection or  $S_1$  seed, the covariance is the same as that for full sibs under sib selection (as all three have the same  $\Theta_{xy}$  value).

The covariance for within-family deviations (again considering the contribution through a single parent of  $y$ ) becomes

$$\begin{aligned} \sigma(z_{ij} - \bar{z}_i, y | \mathcal{R}_1) &= \sigma(\mathcal{R}_1, y) - \sigma(\bar{z}_i, y | \mathcal{R}_1) \\ &= \sum_{u=1} \left(\frac{1}{2}\right)^u \left(1 - (r^u)_n\right) \sigma_{A^u}^2 \\ &= \left(1 - \frac{1}{n}\right) \sum_{u=1} \left(\frac{1}{2}\right)^u (1 - r^u) \sigma_{A^u}^2 \end{aligned} \quad (21.30)$$

where we have used the identity  $(1 - r_n) = (1 - 1/n)(1 - r)$ . Ignoring the common  $(1 - 1/n)$  factor found in all terms, for half-sibs we have

$$\sigma(z_{ij} - \bar{z}_{HS}, y | \mathcal{R}_1) = \left(\frac{3}{8}\right) \sigma_A^2 + \left(\frac{15}{64}\right) \sigma_{AA}^2 + \left(\frac{63}{512}\right) \sigma_{AAA}^2 + \cdots \quad (21.31a)$$

while for full-sibs,

$$\sigma(z_{ij} - \bar{z}_{FS}, y | \mathcal{R}_1) = \left(\frac{1}{4}\right) \sigma_A^2 + \left(\frac{3}{16}\right) \sigma_{AA}^2 + \left(\frac{7}{64}\right) \sigma_{AAA}^2 + \cdots \quad (21.31b)$$

Equations 21.29 and 21.31 show that additive epistasis contributes to the short-term response. However, as with individual selection, this contribution is transient and decays over time as recombination breaks up linkage groups of favorable alleles (Chapter 13). For  $u$ -locus additive epistasis ( $\sigma_{A^u}^2$ ), the per generation decay rate for unlinked loci is  $(1/2)^{u-1}$ , the probability that a parental gamete containing specific alleles at  $u$  unlinked loci is passed onto an offspring. The probability that such a gamete remains unchanged after  $\tau$  generations is  $2^{-\tau(u-1)}$ , which rapidly converges to zero. Thus, if  $R_{A^u}$  is the contribution due to  $u$ -locus additive epistasis, after  $\tau$  generations the contribution from a single generation of selection becomes  $2^{-\tau(u-1)} R_{A^u}$ .

### Response with Autotetraploids

Recall from Chapter 13 that selection response with autotetraploids (which are not uncommon among crop plants) has similar features to selection with epistasis — there is a transient component to the response that is contributed by non-additive gene action. In the case of autotetraploids, this is the dominance variance, which occurs because autotetraploid parents pass along two alleles at each locus to their offspring. As with epistasis, the contribution for non-additive variance entirely arises because the genotypes are not in Hardy-Weinberg equilibrium. After several generations of random-mating, the selection-induced allele frequencies remain unchanged (and hence additive contributions are permanent), but the non-additive contributions, which arise entirely from departures from Hardy-Weinberg, decay away.

This section is a bit technical, so here are the key results. Except in the case of selfing (using  $S_1$  seed), all of the permanent response to selection are the same as with a diploid. The transient contribution due to dominance is generally small (indeed, smaller than the contribution under individual selection, Chapter 13) and is only significant when the dominance variance is significantly larger than the additive variance. Further, this (generally small) transient contribution quickly decays under random mating. The offspring from a cross of two (unrelated) autotetraploid parents from  $S_1$  seed are inbred, as the two alleles from each parent can be identical by descent (and hence inbred). This is not the case when both parents are unrelated diploids. Chapter 20 examines this further.

---

**Example 21.6.** We will need a bit more detailed treatment of resemblance between tetraploid relatives than in given in LW Chapter 7. We start here and conclude in Example 21.7. Label the four alleles in a tetraploid by  $B_1, B_2, B_3, B_4$ . There are six possible gametes from this parent,  $(B_1, B_2), (B_1, B_3), (B_1, B_4), (B_2, B_3), (B_2, B_4)$ , and  $(B_3, B_4)$ . Allowing for non-additive interaction between alleles, the genotypic value can be decomposed into four additive (single allele) terms, six dominance (two allele) terms, four three-way interactions, and one four-way interaction:

$$G_{1234} = a_1 + a_2 + a_3 + a_4 + d_{12} + d_{13} + d_{14} + d_{23} + d_{24} + d_{34} + t_{123} + t_{124} + t_{134} + t_{234} + q_{1234}$$

The resulting total genetic variation can be partitioned as

$$\sigma_G^2 = 4\sigma_a^2 + 6\sigma_d^2 + 4\sigma_t^2 + \sigma_q^2 = \sigma_A^2 + \sigma_D^2 + \sigma_T^2 + \sigma_Q^2$$

If two relatives share only one allele IDB, then their genetic covariance is just  $\sigma_a^2 = (1/4)\sigma_A^2$ . If they share exactly two IDB alleles, then the genetic covariance is  $2\sigma_a^2 + \sigma_d^2 = (1/2)\sigma_A^2 + (1/6)\sigma_D^2$ . Filling out the rest of these covariances and letting  $\pi_i$  denote the probability that two relatives share exactly  $i$  IDB alleles, we have

IBD alleles	Prob	$\sigma_A^2$	$\sigma_D^2$	$\sigma_T^2$	$\sigma_Q^2$
1	$\pi_1$	1/4	0	0	0
2	$\pi_2$	1/2	1/6	0	0
3	$\pi_3$	3/4	1/2	1/4	0
4	$\pi_4$	1	1	1	1

Using these results, we see that the genetic covariance between any two relatives is a function of their  $\pi_1$  through  $\pi_4$  values,

$$\left( \frac{\pi_1 + 2\pi_2 + 3\pi_3 + 4\pi_4}{4} \right) \sigma_A^2 + \left( \frac{\pi_2 + 3\pi_3 + 6\pi_4}{6} \right) \sigma_D^2 + \left( \frac{\pi_3 + 4\pi_4}{4} \right) \sigma_T^2 + \pi_4 \sigma_Q^2$$

With a parent-offspring, exactly 2 alleles are IDB, so that  $\pi_2 = 1$ . With half-sibs, by looking at the  $6 \times 6$  table of pairs of gametes from the same parent, we see that (in the table) 6 entires share two alleles, 24 share one, and six share zero. Hence  $\pi_1 = 24/36 = 2/3$ ,  $\pi_2 = 6/36 = 1/6$ . Turning to full sibs, can can use the half-sib results. Let  $P_1$  and  $P_2$  denote the shared parents. The probability that four alleles are shared is the probability the sibs share two alleles from  $P_1$  times the probability they share two alleles from  $P_2$ , or  $\pi_4 = (1/6)(1/6)$ , assuming the parents are unrelated. Now consider the case of sharing exactly two alleles IDB. This can happen three different ways. Share one from each parent (probability  $[2/3][2/3]$ ), share two from  $P_1$  and zero from  $P_2$  (probability  $[1/6][1/6]$ ), or share two from  $P_2$  and none from  $P_1$  (probability  $[1/6][1/6]$ ), giving

$$\pi_2 = (2/3)^2 + (1/6)^2 + (1/6)^2 = 18/36 = 1/2$$

Similar logic gives  $\pi_1 = 2/9$  and  $\pi_3 = 2/9$ . Summarizing,

Relative pair	$\pi_0$	$\pi_1$	$\pi_2$	$\pi_3$	$\pi_4$
Parent-offspring	0	0	1	0	0
Half-sibs	1/6	2/3	1/6	0	0
Full-sibs	1/36	2/9	1/2	2/9	1/36

Substitution of the above results for the  $\pi_i$  into the general expression for the covariance gives Equation 21.32.

Using the results from Example 21.6, we find that

$$\sigma_G(x, y) = \begin{cases} \frac{1}{2}\sigma_A^2 + \frac{1}{6}\sigma_D^2 & \text{parent, offspring} \\ \frac{1}{4}\sigma_A^2 + \frac{1}{36}\sigma_D^2 & \text{half sibs} \\ \frac{1}{2}\sigma_A^2 + \frac{6}{27}\sigma_D^2 + \frac{1}{12}\sigma_T^2 + \frac{1}{36}\sigma_Q^2 & \text{full sibs} \end{cases} \quad (21.32)$$

where  $\sigma_T^2$  and  $\sigma_Q^2$  are the variances of third- and fourth-order interactions (see LW Chapters 5, 7 for details). Using these covariances and following same logic leading to Equations 21.26a/b gives the between-family genetic variance as

$$\sigma_{GF}^2 = \begin{cases} \frac{1}{4}\sigma_A^2 + \frac{1}{36}\sigma_D^2 & \text{half-sibs} \\ \frac{1}{2}\sigma_A^2 + \frac{6}{27}\sigma_D^2 + \frac{1}{12}\sigma_T^2 + \frac{1}{36}\sigma_Q^2 & \text{full-sibs} \end{cases}, \quad (21.33a)$$

while the within-family genetic variances following from  $\sigma_{Gw}^2 = \sigma_G^2 - \sigma_{GF}^2$ , giving

$$\sigma_{Gw}^2 = \begin{cases} \frac{3}{4}\sigma_A^2 + \frac{35}{36}\sigma_D^2 + \sigma_T^2 + \sigma_Q^2 & \text{half-sibs} \\ \frac{1}{2}\sigma_A^2 + \frac{19}{27}\sigma_D^2 + \frac{11}{12}\sigma_T^2 + \frac{35}{36}\sigma_Q^2 & \text{full-sibs} \end{cases} \quad (21.33b)$$

The between- and within-family variances,  $\sigma^2(\bar{z}_i)$  and  $\sigma^2(z_{ij} - \bar{z}_i)$ , immediately follow by substituting Equation 21.26 into Equations 21.9a and 21.10a.

To proceed further, we need to compute the expected genetic covariance between  $y$  (an offspring of  $\mathcal{R}$ ) and  $x$ , a sib upon which selection decisions are made. This requires us to compute two additional genetic covariances, namely between a half-uncle and a nephew ( $x$  and  $\mathcal{R}$  are half-sibs) and an uncle and nephew ( $x$  and  $\mathcal{R}$  are full-sibs). Example 21.7 carries out the bookkeeping.

**Example 21.7.** Example 21.6 computes the  $\pi_1$  through  $\pi_4$  values between  $x$  and  $\mathcal{R}$  when they are either half- or full-sibs. As we have seen above, expressions for response under various family-based selection schemes requires the covariance between  $x$  and an offspring  $y$  of  $\mathcal{R}$ . We do this by first conditioning on the number of IBD alleles shared by  $x$  and  $\mathcal{R}$ , and then compute the probability that  $\mathcal{R}$  passes along one or two of these IBD alleles to  $y$ . Or example, if  $x$  and  $\mathcal{R}$  share exactly 1 IBD allele, then with probability 1/2 it is also transmitted to  $y$ , in which case  $x$  and  $y$  share one IBD allele. The 1/2 comes from considering the six possible gametes that  $\mathcal{R}$  can generate. Letting  $B_1$  denote the IBD allele for  $x$  and  $\mathcal{R}$ , simply by computing the gametes we see that three of six contain  $B_1$ , while the other three do not. Similar counting fills out the table below. For example, if  $x$  and  $\mathcal{R}$  share 2 IBD alleles,  $B_1$  and  $B_2$ . Again counting the gametes of  $\mathcal{R}$ , only one contains both (1/6 have 2 IBD), while 4 of 6 contain either  $B_1$  or  $B_2$  (but not both). The values when  $x$  and  $\mathcal{R}$  share 3 and 4 IBD are given below:

IBD shared by $x$ and $\mathcal{R}$	Prob(IBD shared by $x$ and $y$ )	
	1	2
1	1/2	0
2	2/3	1/6
3	1/2	1/2
4	0	1

When  $x$  and  $\mathcal{R}$  are half-sibs, Example 21.6 shows that 2/3 of sibs share 1 IBD alleles, while 1/6 share two. Using the table above, we then see the probability that  $y$  and  $x$  share one IBD allele is just

$$\begin{aligned} \pi_1 &= \Pr(x, y \text{ 1IBD} | x, \mathcal{R} \text{ 1 IBD}) \Pr(x, \mathcal{R} \text{ 1 IBD}) + \Pr(x, y \text{ 1IBD} | x, \mathcal{R} \text{ 2 IBD}) \Pr(x, \mathcal{R} \text{ 2 IBD}) \\ &= (1/2)(2/3) + (2/3)(1/6) = 4/9 \end{aligned}$$

$$\pi_2 = \Pr(x, y \text{ 2 IBD} | x, \mathcal{R} \text{ 2 IBD}) \Pr(x, \mathcal{R} \text{ 2 IBD}) = (1/6)(1/6) = 1/36$$

giving

$$\sigma_G(x, y) = \left( \frac{1(4/9) + 2(1/36)}{4} \right) \sigma_A^2 + \frac{1/36}{6} \sigma_D^2 = \frac{1}{8} \sigma_A^2 + \frac{1}{216} \sigma_D^2$$

as the genetic covariance between  $x$  and  $y$  when the relationship is that of half-uncle ( $x$  and  $\mathcal{R}$  half-sibs) and nephew.

When  $x$  and  $\mathcal{R}$  are full-sibs, the probability that they share 1, 2, 3, and 4 IBD alleles is 2/9, 1/2, 2/9, and 1/36. Following the same logic and using the above tables,

$$\pi_1 = (1/2)(2/9) + (2/3)(1/2) + (1/2)(2/9) + (0)(1/36) = 20/36 = 5/9$$

$$\pi_2 = (0)(2/9) + (1/6)(1/2) + (1/2)(2/9) + (1)(1/36) = 2/9$$

giving

$$\sigma_G(x, y) = \left( \frac{1(5/9) + 2(2/9)}{4} \right) \sigma_A^2 + \frac{2/9}{6} \sigma_D^2 = \frac{1}{4} \sigma_A^2 + \frac{1}{27} \sigma_D^2$$

as the genetic covariance between  $x$  and  $y$  when the relationship is that of uncle ( $x$  and  $\mathcal{R}$  full-sibs) and nephew.

Using these results, Equation 21.6 (the covariance between the family mean and an offspring under family selection) again has two terms: a parent-offspring contribution ( $1/n$ ) and the covariance between  $x$  and  $y$  when  $\mathcal{R}$  is an unmeasured sib of  $x$ , giving

$$\begin{aligned} \sigma(\bar{z}, y | \mathcal{R}_1) &= \frac{1}{n} \sigma(y | \mathcal{R}_1) + \left( 1 - \frac{1}{n} \right) \sigma(x, y | \mathcal{R}_1) \\ &= \frac{1}{n} \left( \frac{1}{2} \sigma_A^2 + \frac{1}{6} \sigma_D^2 \right) + \left( 1 - \frac{1}{n} \right) \sigma(x, y | \mathcal{R}_1) \end{aligned} \quad (21.34a)$$

Using the results from Example 21.7, when  $x$  and  $\mathcal{R}$  are half-sibs,

$$\begin{aligned} \sigma(\bar{z}, y | \mathcal{R}_1) &= \frac{1}{n} \left( \frac{1}{2} \sigma_A^2 + \frac{1}{6} \sigma_D^2 \right) + \left( 1 - \frac{1}{n} \right) \left( \frac{1}{8} \sigma_A^2 + \frac{1}{216} \sigma_D^2 \right) \\ &= \frac{1}{8} \sigma_A^2 \left( 1 + \frac{3}{n} \right) + \frac{1}{216} \sigma_D^2 \left( 1 + \frac{35}{n} \right) \end{aligned} \quad (21.34b)$$

The additive-genetic contribution is the same as with diploids, while the contribution from dominance is extremely small. Turning to full-sibs,

$$\begin{aligned} \sigma(\bar{z}, y | \mathcal{R}_1) &= \frac{1}{n} \left( \frac{1}{2} \sigma_A^2 + \frac{1}{6} \sigma_D^2 \right) + \left( 1 - \frac{1}{n} \right) \left( \frac{1}{4} \sigma_A^2 + \frac{1}{27} \sigma_D^2 \right) \\ &= \frac{1}{4} \sigma_A^2 \left( 1 + \frac{1}{n} \right) + \frac{1}{27} \sigma_D^2 \left( 1 + \frac{7}{n} \right) \end{aligned} \quad (21.34c)$$

Again, additive-genetic contribution is the same as with diploids, and the contribution from dominance is small. In both cases, the contribution from dominance is transient, decaying under random mating by 2/3 each generation (Chapter 13, refs.). Using the same logic, similar expressions can be found for within-family selection. Again, the contribution from additive variance is the same as diploids, while the contribution from dominance is small to very small and decays by 67% each generation of random mating.

## EFFICIENCY OF FAMILY-BASED vs. INDIVIDUAL SELECTION

Intuition suggests that individual selection is better than either within- or between-family selection when  $h^2$  is modest to large. When  $h^2$  is small, we expect within-family selection to be more efficient if there is a large common family environmental effect ( $\sigma_{E_c} \simeq \sigma_z^2$ ) and between-family selection to be more efficient if the individual-specific environmental effects are large ( $\sigma_{E_s} \simeq \sigma_z^2$ ). To more formally develop these points, recall that the expected response under mass (individual) selection is  $R_m = \bar{i}_m \sigma_A h$ . Equation 21.14a implies that the ratio of response of between-family vs. individual selection is thus:

$$\frac{R_b}{R_m} = \left( \frac{\bar{i}_b}{\bar{i}_m} \right) \left( \frac{\Theta}{\sqrt{t_n}} \right) \quad (21.35a)$$

where  $\Theta$  is a function of the type of between-family selection (see Equation 21.14b) and  $\bar{i} = (\bar{i}_m + \bar{i}_f)/2$  is the average selection intensity. Likewise, for family deviations selection, Equation 21.19 gives

$$\frac{R_{FD}}{R_m} = \left( \frac{\bar{i}_{FD}}{\bar{i}_m} \right) \left( \frac{1 - r_n}{\sqrt{1 - t_n}} \right) \quad (21.35b)$$

Finally, Equation 21.21 gives the response ratio for strict within-family selection as

$$\frac{R_{WF}}{R_m} = \left( \frac{\bar{i}_{WF}}{\bar{i}_{FD}} \right) \left( \frac{R_{FD}}{R_m} \right) \sqrt{\frac{n}{n-1}} \quad (21.35c)$$

Equations 21.35a-c shows that the relative efficiency of any particular family-based scheme is the product of the ratio of selection intensities (the first term) and the **accuracy of selection** relative to individual selection (the second term). The accuracy ratio measures how well (relative to individual selection) the selection criteria predicts the breeding values of the parents. We focus first on the accuracy ratio as the selection intensity ratio is generally close to one unless sample sizes are very small (Example 21.1, Table 21.6).

## The Relative Accuracies of Family-based vs. Individual Selection

Relative accuracies are typically expressed in terms of the phenotypic correlation  $t$  between sibs and their coefficient of relatedness  $r$ . Since under the simple environmental model ( $E = E_c + E_s$ ), the variance of family means is  $\sigma^2(\mu_i) = t\sigma_z^2 = \sigma_{GF}^2 + \sigma_{Ec}^2$ . Hence,

$$t = \frac{\sigma_{GF}^2}{\sigma_z^2} + \frac{\sigma_{Ec}^2}{\sigma_z^2} = \frac{r\sigma_A^2}{\sigma_z^2} + \frac{(\sigma_{GF}^2 - r\sigma_A^2) + \sigma_{Ec}^2}{\sigma_z^2} = rh^2 + \frac{(\sigma_{GF}^2 - r\sigma_A^2) + \sigma_{Ec}^2}{\sigma_z^2} \quad (21.36a)$$

In the absence of epistasis, Equation 21.11a gives

$$t = rh^2 + c^2, \quad \text{where} \quad \sigma_z^2 c^2 = \begin{cases} \sigma_{Ec(HS)}^2 & \text{half-sibs} \\ \frac{1}{4}\sigma_D^2 + \sigma_{Ec(FS)}^2 & \text{full-sibs} \end{cases} \quad (21.36b)$$

where  $c^2$  is the residual between-family variance (upon removal of any additive variance). Figures 21.4 and 21.5 plot the relative accuracies and relative responses under between-family selection (family selection) and within-family selection (family deviations).

What are the exact conditions for a particular method to be more accurate than individual selection? Equation 21.35a shows that between-family selection is more accurate when  $\Theta/\sqrt{t_n} > 1$ , or  $\Theta^2 > t_n$ , or

$$t_n = t + \frac{1-t}{n} = (rh^2 + c^2) \left( 1 - \frac{1}{n} \right) + \frac{1}{n} < \Theta^2 \quad (21.37a)$$



For  $n$  moderate to large, between-family selection is more accurate than mass selection when the fraction  $c^2$  of total variance due to residual between-family effects is sufficiently small. Substituting  $\Theta$  (Equation 21.14b) into Equation 21.37a (for moderate to large  $n$ ) yields

$$c^2 < \begin{cases} \frac{1}{16} (1 - 4h^2) & \text{half-sibs (for family, sib selection)} \\ \frac{1}{4} (1 - h^2) & \text{half-sibs (for parental and } S_1 \text{ seed selection)} \\ \frac{1}{4} (1 - 2h^2) & \text{full-sibs} \end{cases} \quad (21.37b)$$

If  $h^2 > 1/2$ , the condition becomes  $c^2 < 0$  and hence between-family selection using full-sibs is always less efficient than individual selection. With half-sibs, family- and sib-selection is always less efficient than individual selection when  $h^2 > 1/4$ . Thus between-family selection is only more effective than mass selection when heritability is small and common-family environmental variation is also small.

Turning to within-family selection, family deviations (FD) gives a larger response than individual selection when  $(1 - r_n)/\sqrt{1 - t_n} > 1$ . When families are large ( $n \gg 1$ ), this condition reduces to

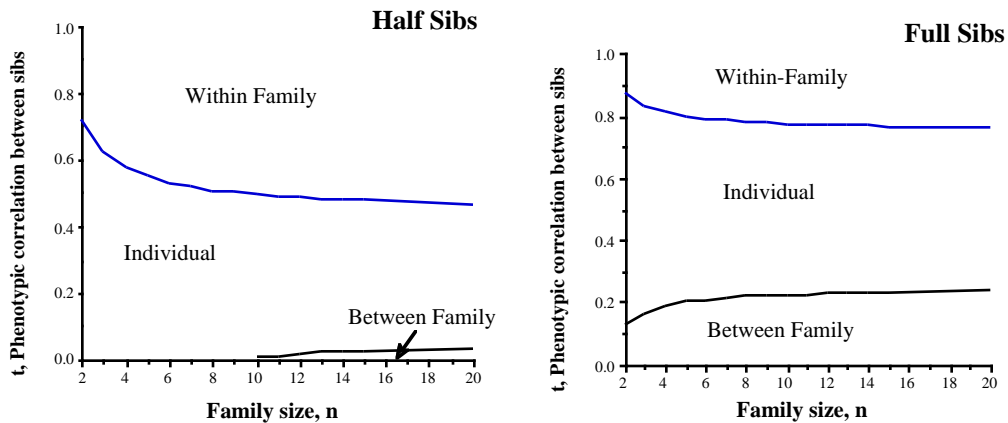
$$t = rh^2 + c^2 > 1 - (1 - r)^2 \quad (21.38a)$$

or

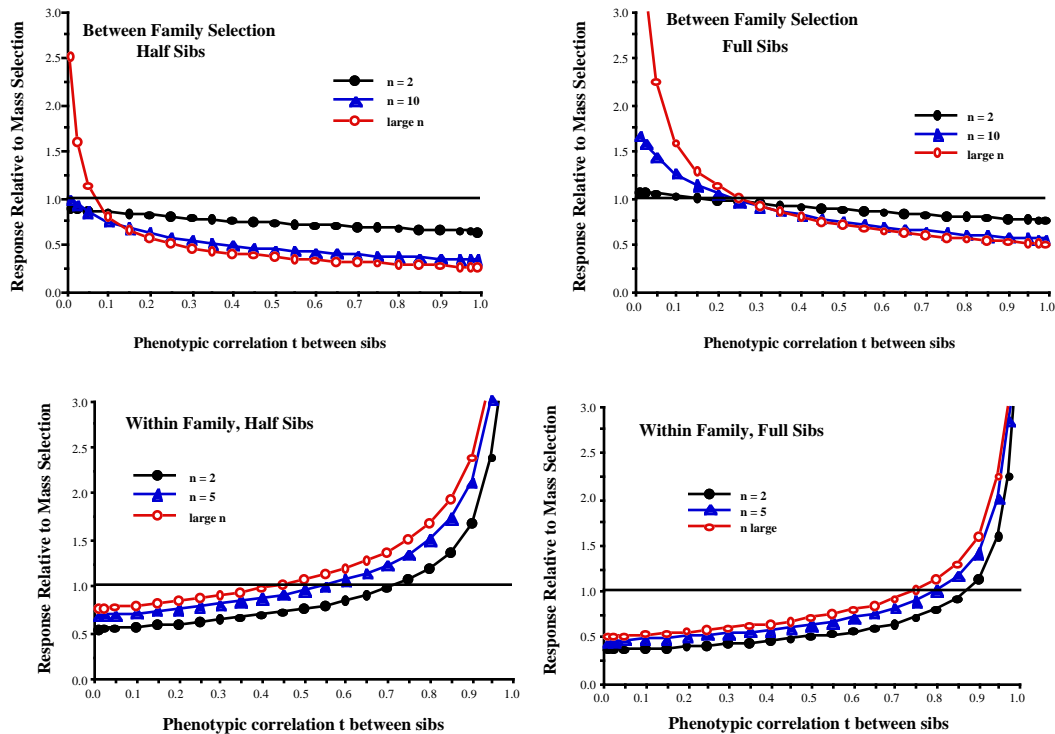
$$c^2 > 1 - (1 - r)^2 - rh^2 = \begin{cases} \frac{7}{16} - \frac{h^2}{4} & \text{half-sibs} \\ \frac{3}{4} - \frac{h^2}{2} & \text{full-sibs} \end{cases} \quad (21.38b)$$

Since  $h^2 + c^2 \leq 1$  (both being fractions of the total variance due to difference sources), there is an additional constraint that  $1 - h^2 \geq c^2$ . When  $h^2 > 0.75$  within-(half-sib) family selection is less efficient than individual selection, as  $c^2 \leq 1 - h^2 = 0.25$ , while the critical  $c^2$  value that must be exceeded is 0.25. For full-sibs, individual selection is more efficient than within-family selection whenever  $h^2 > 0.5$ . Because we assumed  $n$  is large,  $r$  and  $t$  replace  $r_n$  and  $t_n$  and hence Equations 21.38a and 21.38b are also the conditions for strict within-family (WF) selection. Within-family selection is thus more efficient than individual selection only when the heritability is low and the residual between-family variance ( $c^2\sigma_z^2$ ) accounts for a very significant fraction of the total variance, i.e., the individual-specific environmental effects account for much of the phenotypic variance.

Willeke (1982) suggested that an excellent candidate trait for within-family selection would be litter size in pigs. Recall (Chapter 13) that mice from large litters tend to have a negative environmental value for litter size. Given that heritabilities estimates for pig litter size from grandmother-granddaughter regression are higher than those based on mother-daughter regression, a similar situation is likely occurring in pigs. Thus, there is a large family contribution that obscures prediction of breeding value from phenotypic value, resulting in female's ranking within a family being a more informative predictor of breeding value than her phenotypic value.



**Figure 21.4.** Regions of the family size ( $n$ ) - sib correlation ( $t$ ) space where individual, between-family (family selection) and within-family (selection of family deviations, FD) are the most accurate. If  $t$  is sufficiently large, within-family selection gives the largest response (for large  $n$ ,  $t > 7/16 = 0.4375$  for half-sibs and  $t > 3/4$  for full-sibs), between-family selection is best when  $t$  is sufficiently small (for large  $n$ ,  $t < 1/16 = 0.0625$  for half-sibs,  $t < 1/4$  for full-sibs), and individual selection gives the largest response for intermediate values of  $t$ . For large  $n$  ( $t_n, r_n$  approaching  $t, n$ ), family selection equals sib selection, as does parental selection (using the curve for full-sibs), while family-deviations selection approaches strict within-family (WF) selection.



**Figure 21.5.** Accuracies of family selection (top row) and selection on family deviations FD (bottom row) relative to mass selection. Values exceeding one indicate an increased response relative to individual selection assuming equal selection intensities.

**Example 21.8.** Wilson (1974) examined family selection (using full-sibs) on larval and pupal weight in *Tribolium castaneum*. Correlations among full-sibs were estimated to be  $t = 0.16$  for larval weight and  $t = 0.20$  for pupal weight. Family size was  $n = 12$ . Under family selection  $\Theta = r_n$ . With full-sibs,  $r = 1/2$ , giving the relative accuracy of family to mass selection on larval weight as

$$\frac{\Theta}{\sqrt{t_n}} = \frac{r + (1-r)/n}{\sqrt{t + (1-t)/n}} = \frac{0.5 + 0.5/12}{\sqrt{0.16 + 0.84/12}} = 1.13$$

Likewise, the relative accuracy for pupal weight is

$$\frac{1/2 + 1/(2 \cdot 12)}{\sqrt{0.2 + 0.8/12}} = 1.05$$

Both characters are thus expected to show a larger response under family selection than under mass selection. However, Wilson observed that individual selection gave a larger (though not significant) response for both characters. Note from Equation 21.14b that the expected response for sib selection using full sibs is the same as for parental selection (both have  $\Theta = 1/2$ ). The relative accuracy of these methods on larval weight is

$$\frac{\Theta}{\sqrt{t_n}} = \frac{0.5}{\sqrt{0.16 + 0.84/12}} = 1.04,$$

while their relative accuracy for pupal weight is

$$\frac{\Theta}{\sqrt{t_n}} = \frac{0.5}{\sqrt{0.2 + 0.8/12}} = 0.97$$

Thus for pupal weight, family selection is slightly more accurate than mass selection, while sib and parental selection are slightly less accurate.

Several other studies have compared family and individual selection. Campo and Tagarro (1977) compared full-sib family and individual selection on *Tribolium* pupal weight, using experiments with family sizes of four and ten. For  $n = 4$ , the predicted relative accuracy is 0.87. In both experiments, family selection gave the larger single-generation response, while mass selection had the larger response after six generations. None of these differences were significant. Two other studies compared individual and between-family selection, both using half-sib family selection in chickens. Garwood et al. (1980) examined laying rate ( $h^2 = 0.22$ ) and egg weight ( $h^2 = 0.55$ ), finding that individual selection gave a greater single-generation response for both characters, but the difference for egg weight was not significant. Kinney et al. (1970) examined several characters, finding individual selection exceeded family selection, although again none of the differences were significant.

Finally, it is also informative to compare methods using the heritability version of response,  $R_x = h_x^2 S_x$ . From Equation 21.23, the within-family heritability exceeds the individual heritability when  $1 - r > 1 - t$ . Hence,

$$h_w^2 > h^2 \quad \text{when} \quad t > r \quad (21.39a)$$

Likewise, from Equation 21.15, the between-family heritability satisfies

$$h_b^2 > h^2 \quad \text{when} \quad \Theta > t_n \quad (21.39b)$$

These conditions are rather different from those given by Equations 21.37a and 21.38a. For example, Equation 21.37a implies that between-family selection gives a larger response than individual selection when  $\Theta > \sqrt{t_n}$ , Equation 21.39b implies that the between-family heritability is greater than  $h^2$  when  $\Theta > t_n$ . What is the discrepancy between these two approaches (accuracies vs. heritabilities)?

The key is that the variances of the groups being selected differ. Since  $\sigma_z^2 = \sigma_b^2 + \sigma_w^2$ , the between-family and within-family variances are each less than the phenotypic variance of a random individual. Since  $S_x = \bar{i}_x \sigma_x$ , larger selection intensities are required to give a family-based approach the same selection differential as individual selection. Since the within- and between-family variances are  $(1 - t_n)\sigma_z^2$  and  $t_n\sigma_z^2$ , respectively, it follows that

$$\frac{S_b}{S_m} = \frac{\bar{i}_b \sigma_b}{\bar{i}_m \sigma_z} = \frac{\bar{i}_b}{\bar{i}_m} \sqrt{t_n} \quad \text{and} \quad \frac{S_w}{S_m} = \frac{\bar{i}_w \sigma_w}{\bar{i}_m \sigma_z} = \frac{\bar{i}_w}{\bar{i}_m} \sqrt{1 - t_n}$$

Under identical selection intensities, the differentials for between- and within-family selection are  $\sqrt{t_n}$  and  $\sqrt{1 - t_n}$  of the differential under mass selection. Thus even when  $h_w^2$  or  $h_b^2$  exceeds  $h^2$ , this advantage is partially countered by smaller selection differentials due to smaller variances. The initial comparison assuming the same selection differential thus has a hidden assumption of more selection (a larger selection intensity) under family-based selection.

### Comparing Selection Intensities: Finite Size Corrections

While not nearly as dramatic as differences in the selection differentials, the selection intensities can differ across methods even if the same fraction  $p$  is saved (Example 21.1). These differences arise from the finite sample size correction of  $\bar{i}$  (Chapter 12). Suppose a total of nine individuals are measured, three from each of three families. If we select for the upper 1/3, we keep the best one of three families under between-family selection, and the best of the three individuals within each family under WF selection, giving an expected selection intensity of  $\bar{i}_{(3,1)} = 0.846$  (the expected value of the largest of the three order statistics). Under family deviations (FD) and mass selection, we chose the largest three of nine values, giving an expected selection intensity of  $\bar{i}_{(9,3)} = 0.996$ . Table 21.6 summarizes the selection intensities for the different methods, showing that  $\bar{i}_b, \bar{i}_{WF} \leq \bar{i}_m \leq \bar{i}_{FD}$ .

**Table 21.6.** Selection intensities for various forms of family-based selection corrected for finite sample size. The upper  $p$  of the population is saved and the population consists of  $m$  families each with  $n$  members, for a total of  $N = mn$  individuals. Tables of exact values for  $\bar{i}_{(N,K)}$  (the average value of the top  $K$  of the  $N$  standardized order statistics, see Chapter 12) are given by Becker (1992). Approximations for  $\bar{i}_{(N,K)}$  are given in Equations 10.17a-d.

Selection Type	Corrected Selection Intensity
Individual	$\bar{i}_m = \bar{i}_{(N,pN)}$
Between-family	$\bar{i}_b = \bar{i}_{(m,pm)}$
Family-deviations	$\bar{i}_{FD} = \bar{i}_{(N,pN)} \sqrt{1 + \frac{1}{N-1}}$
Within-family	$\bar{i}_{WF} = \bar{i}_{(n,pn)}$

An additional subtlety in adjusting the selection intensity was pointed out by Hill (1976, 1977). The expected selection intensity is computed by taking the expected value of the largest standardized order statistics (Chapter 12), under the assumption that the order statistics are *uncorrelated*. With family deviations (FD), family index, and even mass selection, there is the potential for correlations between order statistics. Here different families can contribute different numbers of individuals, resulting in correlations between those measures from the same family and hence correlations between some of the order statistics. The correction for mass selection is generally very small and will be ignored here (see Equation 21.57b). Within-family deviations are negatively correlated within a family, ( $\rho = -1/(n - 1)$  for a family of size  $n$ ), as they are deviations from a common family mean. Dempfle (1990) and Hill et al. (1996) show that the resulting selection intensity for within-family deviations is thus slightly larger than the intensity for mass selection  $\bar{i}_m$ ,

$$\bar{i}_{FD} = \bar{i}_m \sqrt{1 + \frac{1}{N - 1}} \quad (21.40)$$

where  $N$  is the total number of measured sibs. Again, this correction is only important when family size is very small. On the other hand, with selection on a family index, the correlations between index scores are positive and can be considerable even for large  $n$  (Equation 21.58). We consider the appropriate correction for  $\bar{i}$  in our treatment of family index selection at the end of this chapter.

## RESPONSE WHEN FAMILIES ARE REPLICATED OVER ENVIRONMENTS

Family members are often raised in multiple plots and/or environments. If the investigator is not aware of this underlying environmental structure, it inflates the between-family means relative to designs that take this structure into account. Carefully designed family replication offers two potential advantages. First, it allows for selection of families that perform best over a range of environments, even when extensive genotype-environment interactions are present. Second, replication within an environment reduces the effects of microenvironmental differences, increasing the predictability of a family's breeding value (and hence the response).

Since family replication is a hallmark of plant breeding, we examine several schemes used by crop breeders in detail in this section (Chapter 20 and Volume 2 examines related designs involving inbreeding and line crossing, while Chapters 33 and 34 more fully examine selection in the presence of  $G \times E$ ). Detailed reviews of plant breeding methodology are given by Namkoong (1979), Hallauer and Miranda (1981), Hallauer (1981, 1985), Nguyen and Sleper (1983), Wricke and Weber (1986), Mayo (1987), Hallauer et al. (1988), Gallais (1990), Nyquist (1991), Stoskopf et al. (1993), Bos and Caligari (1995), Allard (1999), Bernardo (2002), Gallais (2003), Sleper and Poehlman (2006), and Acquaah (2007).

### Between-family Variance Under Replication

The expected response to between family selection under replication follows from Equation 21.14a using the appropriate between-family variance  $\sigma^2(\bar{z})$  given the replication design.

In the simplest case, only a single macroenvironment (such as a growing region) is considered, and the family is replicated by raising  $n_s$  sibs in each of  $n_p$  separate plots (for a total of  $N = n_p n_s$  sibs per family). Here the total environmental value can be partitioned as  $E = E_c + E_p + E_{w(p)}$ , a common-family effect ( $E_c$ ), a plot-specific effect ( $E_p$ ), and individual within-plot effects ( $E_{w(p)}$ ). Following similar logic to that in Example 21.3, the resulting

variance is

$$\sigma^2(\bar{z}) = \sigma_F^2 + \frac{\sigma_{E_p}^2}{n_p} + \frac{\sigma_w^2}{N} \quad (21.41a)$$

where  $\sigma_{E_p}^2$  is the plot-to-plot variance (the environmental variance between plots in the same macroenvironment),  $\sigma_F^2 = \sigma_{GF}^2 + \sigma_{E_c}^2$  is the between-family variance and  $\sigma_w^2 = \sigma_{Gw}^2 + \sigma_{E_{w(p)}}^2$  is the within-plot variance of individuals from their family averages. Noting that  $\sigma_G^2 = \sigma_{GF}^2 + \sigma_{Gw}^2$ , we can express Equation 21.41a as

$$\sigma^2(\bar{z}) = \sigma_{GF}^2 + \frac{\sigma_{Gw}^2}{N} + \sigma_{E_c}^2 + \frac{\sigma_{E_p}^2}{n_p} + \frac{\sigma_{E_{w(p)}}^2}{N} \quad (21.41b)$$

showing that while the weighting on the within-family genetic component ( $\sigma_{Gw}^2$ ) depends only on the total number of sibs  $N$ , weighting of the environmental components is a function of the particular experimental design (the values of  $n_p$  and  $n_s = N/n_p$  used).

More generally, if the family is replicated (in  $n_p$  plots of size  $n_s$ ) over  $n_e$  distinct macroenvironments for a total of  $N = n_p n_s n_e$  individuals,

$$\sigma^2(\bar{z}) = \sigma_{GF}^2 + \frac{\sigma_{Gw}^2}{N} + \sigma_{E_c}^2 + \frac{\sigma_{F \times E}^2}{n_e} + \frac{\sigma_{E_p}^2}{n_e n_p} + \frac{\sigma_{E_{w(p)}}^2}{N} \quad (21.42a)$$

where  $\sigma_{GF}^2$  is the genetic variance among family means over this set of environments, and  $\sigma_{F \times E}^2$  is the variance from the family-environment interaction (LW Chapter 22). Note that if we measured the family in only a single environment,  $\sigma_{GF}^2$  also includes a  $\sigma_{GF \times E}^2$  term. Plant breeders often use an alternative partition of the environment into location  $L$  and year  $Y$  effects. If a family is replicated in  $n_\ell$  locations over  $n_y$  years, the resulting family variance is

$$\sigma^2(\bar{z}) = \sigma_F^2 + \frac{\sigma_{F \times L}^2}{n_\ell} + \frac{\sigma_{F \times Y}^2}{n_y} + \frac{\sigma_{F \times L \times Y}^2}{n_\ell n_y} + \frac{\sigma_{E_p}^2}{n_\ell n_y n_p} + \frac{\sigma_w^2}{n_\ell n_y n_p n_s} \quad (21.42b)$$

where  $\sigma_{F \times L}^2$ ,  $\sigma_{F \times Y}^2$ , and  $\sigma_{F \times L \times Y}^2$  are the family by environment (year, location, and year-location) interactions (Lonnquist 1964, Comstock and Moll 1973, Patterson et al. 1977, Brennan and Byth 1979, Thompson and Cunningham 1979).

Equation 21.42 shows the importance of replication and provides some guidance as to how one should allocate resources — for a fixed number of sibs per family ( $N$ ), how should one choose  $n_e$ ,  $n_p$ , and  $n_s$  to minimize  $\sigma^2(\bar{z})$ ? With  $N$  fixed, the relative weighting on the within-family genetic variance, total genetic, and within-plot individual environmental variance are fixed. When the genotype-environment interaction variance ( $\sigma_{GF \times E}^2$ ) is large, its effect on the selection response can be reduced by replication of families across more environments (increasing  $n_e$ ). When the between-plot variance ( $\sigma_{E_p}^2$ ) is large, its effect is reduced by increasing  $n_p$  and/or  $n_e$ . With preliminary estimates of the variance components in hand, one can numerically search for the optimal values of  $n_e$ ,  $n_p$ , and  $n_s$  that give the smallest  $\sigma^2(\bar{z})$  for a fixed value of  $N = n_e n_p n_s$ . Using replication can result in a considerable improvement over mass selection. For example, using variance components estimated for maize strains grown in several locations in Indian, Sanghi (1983) estimated that full-sib selection with replication would be three to six times more efficient than mass selection. Chapters 33 and 34 develop this further, and present more sophisticated approaches to dealing with selection under  $G \times E$ .

One consequence of replication is that the between-family heritability,  $h_b^2 = \Theta \sigma_A^2 / \sigma^2(\bar{z})$  is now a complex function of the design (i.e.,  $n_e$  and  $n_p$  in addition to the total number of sibs). Thus with replication, a between-family heritability does not simply translate into

an individual heritability (Hanson 1963, Nyquist 1991). Even with the same variance components, the heritability changes as a function of the replication design. Hanson suggests that the between-family heritability when replication is present needs to be defined with respect to a particular standard design, such as two years over two locations each with two replications for soybeans.

Finally, consider the between-family variance under a nested sib design with replication. Suppose (as before) that there are  $n_f$  females per male, but now that each full-sib family is replicated as  $n_s$  sibs over  $n_e$  environments. The resulting variance becomes

$$\sigma^2(\bar{z}) = \sigma_{GF(HS)}^2 + \frac{\sigma_{G(f|m)}^2}{n_f} + \frac{\sigma_{GF(HS) \times E}^2}{n_e} + \frac{\sigma_{G(f|m) \times E}^2}{n_f n_e} + \frac{\sigma_{Gw(FS)}^2 + \sigma_e^2}{N} \quad (21.43a)$$

where  $N = n_f n_e n_s$  is the total number of half-sibs per male (Robertson et al. 1955, Webel and Lonnquist 1967, da Silva and Lonnquist 1968). Assuming no epistasis, we can express this between-family variance as

$$\frac{\sigma_A^2}{4} + \frac{\sigma_A^2 + \sigma_D^2}{4 n_f} + \frac{\sigma_{A \times E}^2}{4 n_e} + \frac{\sigma_{A \times E}^2 + \sigma_{D \times E}^2}{4 n_f n_e} + \frac{(1/2)\sigma_A^2 + (3/4)\sigma_D^2 + \sigma_e^2}{N} \quad (21.43b)$$

**Example 21.9.** Eberhart et al. (1966) estimated genetic variance components for seven characters in two different open-pollinated maize varieties. Using individuals grown in two locations in North Carolina, they obtained the following estimates for yield in the variety Jarvis:

$$\sigma_A^2 = 120, \quad \sigma_{A \times L}^2 = 114, \quad \sigma_D^2 = 270, \quad \sigma_{D \times L}^2 = 98, \quad \sigma_e^2 = 508$$

Estimates of epistatic variances were not significantly different from zero. Consider the expected response under a design with 25 half-sib families, each with a total of 50 offspring scored over five environments ( $n_e = 5$ ). The top five families are selected, using  $S_1$  seed to form the next generation (allowing for selection on both sexes). The expected response is

$$R = \frac{2\bar{z}_{25,5}(\sigma_A^2/4)}{\sigma(\bar{z})} = \frac{2 \cdot 1.345 \cdot 30}{\sigma(\bar{z})} = \frac{80.7}{\sigma(\bar{z})}$$

Using the above variance estimates, Equation 21.26 gives  $\sigma_{GF}^2 = \sigma_A^2/4 = 30$ ,  $\sigma_{Gw}^2 = (3/4)\sigma_A^2 + \sigma_D^2 = 360$ , and  $\sigma_{GF \times E}^2 = \sigma_{A \times L}^2/4 = 217.5$ . If the families being scored are strict half-sibs (all offspring from a pollen parent each have a different seed parent,  $n_f = N = 50$ ), Equation 21.42a gives

$$\begin{aligned} \sigma^2(\bar{z}_{HS}) &= \sigma_{GF}^2 + \frac{\sigma_{Gw}^2 + \sigma_e^2}{N} + \frac{\sigma_{GF \times L}^2}{n_e} \\ &= 30 + \frac{360 + 508}{50} + \frac{28.5}{5} = 53.06 \end{aligned}$$

The expected response is thus  $80.7/\sqrt{53.0646} = 11.08$ .

Now suppose that the sibs are from a nested design with each male pollinating five seed parents, with each cross producing 10 offspring ( $n_f = 5$ ,  $N = 50$ ). Using the above variance components, Equation 21.43b gives  $\sigma^2(\bar{z})$  as

$$\begin{aligned} &\frac{\sigma_A^2}{4} + \frac{\sigma_A^2 + \sigma_D^2}{4 n_f} + \frac{\sigma_{A \times L}^2}{4 n_e} + \frac{\sigma_{A \times L}^2 + \sigma_{D \times L}^2}{4 n_f n_e} + \frac{(1/2)\sigma_A^2 + (3/4)\sigma_D^2 + \sigma_e^2}{N} \\ &= \frac{120}{4} + \frac{120 + 270}{20} + \frac{114}{20} + \frac{114 + 98}{100} + \frac{(1/2)120 + (3/4)270 + 508}{50} \\ &= 72.73 \end{aligned}$$

giving the expected response as  $80.7/\sqrt{72.73} = 9.47$ . The strict half-sib design has a smaller between-family variance, and hence a larger expected response, being 117% of that expected under a nested design.

---

Another example of family selection with replication is provided by selection for increased grain yield in maize by the International Maize and Wheat Improvement Center (CIMMYT), summarized by Pandey et al. (1986, 1987) and Crossa and Gardner (1989). The goal of the CIMMYT selection schemes was to develop varieties of maize that yield well over a wide range of environments. Starting in 1974, 250 full sib families along with six local **checks** were evaluated at six lowland tropical locations (with two replications per location) in the northern and southern hemispheres. A total of 28 countries were used during the course of five cycles of selection. Selection (initially) was strictly between-families with the international field trails conducted on full-sib families, while the recombination unit consisted of  $S_1$  seed from the superior families. Selection was later modified to allow for within-family selection as well. Roughly 50% of the families were selected based on the international trails, about 20% of which were subsequently rejected given their poor performance in disease and insect resistance trials in separate nurseries. The average gain in yield per cycle was around 2%.

### Ear-to-Row Selection

One of the earliest examples of family-based selection in plants is **ear-to-row** selection in maize, first used by Hopkins (1899) to start his classic long-term selection experiments (Chapter 23). Here the seeds from each ear are planted in a single row (so that a row corresponds to a family), with individuals from the best rows chosen as seed parents for the next generation. Plants in rows being scored are detassled (or tassles bagged), removing their ability to produce pollen. As a result, these plants can neither self nor pollinate other members of their (or any other) row. Pollen is provided by rows planted with bulk of all seeds (a **polycross** mating design). Assuming open pollination, the seeds on a single ear are half-sibs (with a common mother), so that the ear-to-row method is an example of half-sib family selection, with selection on only one sex (the seed parent). Suppose a total of  $N = n_e n_p n_s$  sibs per family are scored, by growing  $n_p$  rows of  $n_s$  sibs replicated over  $n_e$  distinct environments. The expected response when choosing the top  $K$  of  $M$  families is

$$R_{ER} = \bar{i}_{(M,K)} \frac{(1 + 3/N) (\sigma_A^2/8)}{\sigma(\bar{z}_{HS})} \simeq \bar{i}_{(M,K)} \frac{\sigma_A^2/8}{\sigma(\bar{z}_{HS})} \quad (21.44)$$

where  $\sigma^2(\bar{z}_{HS})$  is given by Equation 21.42. For large  $N$  (in the absence of epistasis) we have

$$R_{ER} = \bar{i}_{(M,K)} \frac{\sigma_A^2/8}{\sqrt{\frac{\sigma_A^2}{2} + \frac{\sigma_{F \times E}^2}{n_e} + \frac{\sigma_{E_p}^2}{n_e n_p} + \frac{\sigma_{E_{w(p)}}^2}{N}}} \quad (21.45)$$

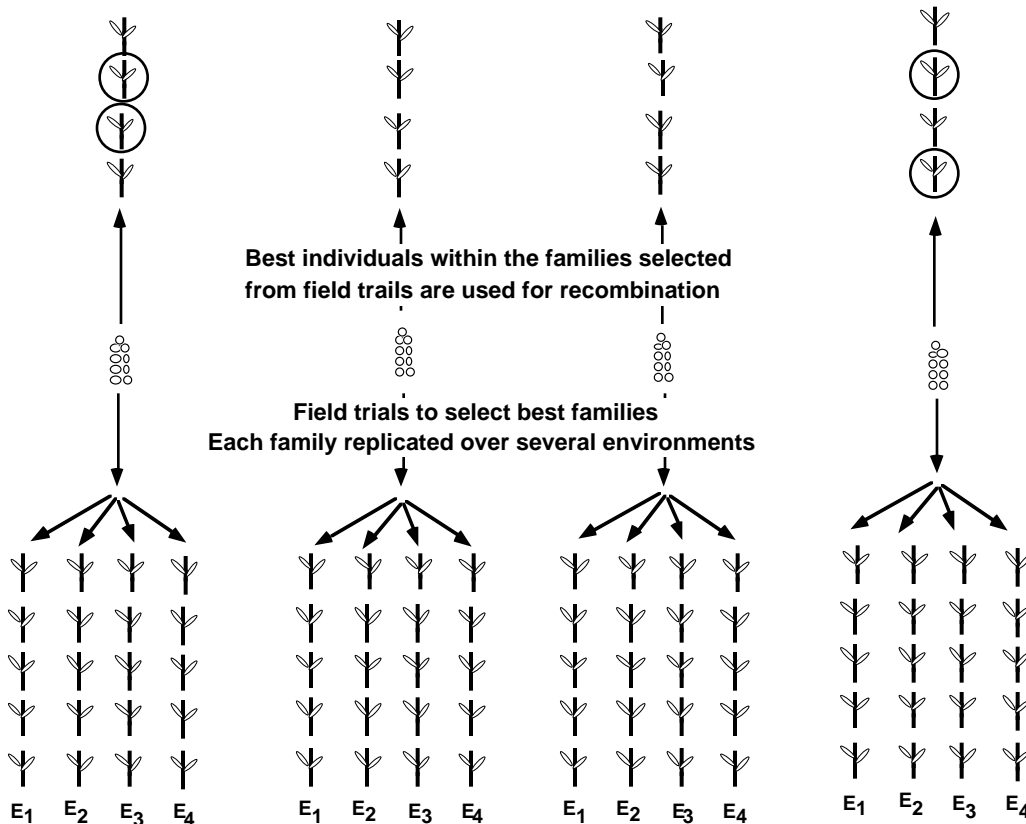
In rice, **panicle-to-row** selection has been used (e.g., Ntanos and Roupakias 2001), when the **panicle** is essentially the equivalent of the maize ear, and again a row equals a family.

### Modified Ear-to-Row Selection

The ear-to-row method has the advantages of being fairly easily to implement and of testing a family (and hence controlling the environmental variance), coupled with the same cycle time as mass selection (one generation). As a result, this method was commonly used by early



maize breeders, e.g., Hopkins (1899), Smith (1908, 1909), Montgomery (1909), Williams and Walton (1915), Kiesselbach (1916), and Hume (1919). While it proved effective at modifying highly heritable traits (such as kernel protein and oil content), ear-to-row selection was generally not successful in improving yield (Kiesselbach 1922, Richey 1922, Smith and Bruson 1925), and it was not regarded as a practical scheme for yield improvement. Sprague (1955) suggested that the failure for yield improvement was largely the result of insufficient control over environmental variance, resulting in  $\sigma_E^2$  largely obscuring the additive variance. (For this same reason, mass selection was also regarded as being impractical for improving maize yield.) An alternative hypothesis was suggested by Hull (1945, 1952), who thought that the lack of response in yield was a result of most genetic variance being nonadditive. The finding of considerable additive variance in yield by a number of maize geneticists motivated the development by Lonnquist (1964) of the **modified ear-to-row** selection scheme, a combined selection scheme involving both between-family (ear-to-row) and within-family (within row) selection (Figure 21.6).



**Figure 21.6.** Lonnquist's modified ear-to-row selection scheme. Half-sib families (represented here by the maize ears in the middle of the figure) are planted both as rows in multiple environments (the yield trials here over environments  $E_1$  to  $E_4$  at the bottom of the figure) and as a single row in yet another location, the so-called **crossing block** (the rows at the top of the figure). For the best families in the yield trials (here families one and four), one then chooses the best individuals (indicated by the circled plants) from their sibs in the crossing block to form the next generation.

Under Lonnquist's design, seed from each family is planted as rows in several environments. These form the **yield** or **performance trails** for selecting the best-performing families averaged over these environments. On a separate plot (the **crossing block**), seed for each family is also planted as a single row. Again, each row is detassled and pollen is provided from rows consisting of a bulk of all families. Within the crossing block, the best individuals from the rows corresponding to the families with the best performance in the yield trails are used as the seed parents for the next generation. One advantage of this scheme is that one can use bulk measures over rows in the yield trails and more detailed (and labor-intensive) individual plant measures in the smaller crossing block.

Under Lonnquist's original design, the replicated field trials and the crossing block are grown contemporaneously (planting of the crossing block may be delayed slightly to ensure that all field information from the yield trials can be gathered). Thus, one cycle of modified ear-to-row can be carried out in a single generation. Plants in the crossing block are open pollinated from a random bulk of all the initially planted families. The expected total response is the sum of the expected gains at each step in the cycle,  $R_{ER(m)} = R_{ER} + R_{ER(w)}$ . The response under the first step (choosing the best families) is the same as for ear-to-row (Equations 21.44-21.45). Since plants in the crossing block are open pollinated using a bulk of all families, selection is only on females within each row. If one chooses the best  $k$  of  $m$  plants within each selected row (i.e., strict within-family, FW, selection), Equation 21.20 gives the expected response to within-row selection as

$$R_{ER(w)} = \bar{i}_{(m,k)} \frac{(3/8) \sigma_A^2}{\sigma_{w(HS)}} \quad (21.46)$$

Since families are not replicated within the crossing block, we have

$$\sigma_{w(HS)}^2 = \sigma_{GW(HS)}^2 + \sigma_{E_w(p)}^2$$

so that (in the absence of epistasis), the component of response from within-row selection becomes

$$R_{ER(w)} = \bar{i}_{(m,k)} \frac{(3/8) \sigma_A^2}{\sqrt{(3/4) \sigma_A^2 + \sigma_D^2 + \sigma_{E_w(p)}^2}} \quad (21.47)$$

Ignoring any potential changes in  $\sigma_A^2$  due to the first step of selection, the expected response becomes

$$\begin{aligned} R_{ER(m)} &= R_{ER} + R_{ER(w)} \\ &= \bar{i}_{(M,K)} \frac{\sigma_A^2/8}{\sigma(\bar{z}_{HS})} + \bar{i}_{(m,k)} \frac{(3/8) \sigma_A^2}{\sigma_{w(HS)}} \end{aligned} \quad (21.48a)$$

where we have chosen the best  $K$  of  $M$  families in the yield trails and the best  $k$  of  $m$  within each selected family in the crossing block. With a large number of sibs/row ( $n$  large) and a roughly equal selection within- and between rows, the expected response to modified ear-to-row selection is

$$R_{ER(m)} = \frac{\bar{i} \sigma_A^2/8}{\sqrt{\sigma_{GF}^2 + \frac{\sigma_{F \times E}^2}{n_e} + \frac{\sigma_{E_p}^2}{n_e n_p} + \frac{\sigma_{E_{w(p)}}^2}{N}}} + \frac{\bar{i} (3/8) \sigma_A^2}{\sqrt{\sigma_{GW(HS)}^2 + \sigma_{E_w(p)}^2}} \quad (21.48b)$$

Inspection of Equation 21.48b shows that it is not obvious which component (within- vs. between-family) contributes more to the total response. The three-fold increase in usable additive variance in the within-family component in the numerator can be partly or fully offset

by the fact that  $\sigma_{Gw}^2 > \sigma_{GF}^2$  (the within-family genetic variance is greater than the between-family variance, see Equations 21.26a,b). Likewise, it is not clear whether the between- or the within-family environmental variance is expected to be larger. Some fine-tuning is possible on the between-family component, as if estimates of the appropriate environmental variances are available, changing the experimental design (the values of  $n_p$ ,  $n_s$ , and  $n_e$ ) can reduce the between-family variance.

---

**Example 21.10.** Webel and Lonnquist (1967) used modified ear-to-row selection for yield in the Hays Golden open-pollinated variety of maize. Performance of each family was evaluated using single rows grown in three different locations. Based on these yield trails, the best 44 of roughly 220 families were identified. In the crossing block the best five of the 25 (or so) plants are chosen in each of the 44 rows corresponding to the selected families. The resulting expected selection intensities for the between- and within-family components are  $\bar{i}_{220,44} = 1.40$  and  $\bar{i}_{25,5} = 1.35$ , respectively. Over the first four cycles of selection, Webel and Lonnquist observed a 9.4 % increase in yield per cycle, compared with the 3% per cycle observed under mass selection (Gardner 1973). The predicted response was 8.4%, with expected contributions of 4.6% from among-families (54% of predicted response) and 3.8% from within-families. The results for ten cycles of selection are summarized by Compton and Bahadur (1977).

Paterniani (1967) also used modified ear-to-row for three cycles of selection for yield in Brazilian maize populations. The average yield increased by 42% over the course of the experiment.

---

Compton and Comstock (1976) suggested a variant of Lonnquist's design. Families are again planted ear-to-row in performance trails, but remnant seed from each family is stored. The best families are chosen and the remnant seed for these families is planted to form the crossing block. The pollen plants in the crossing block are a bulk of the selected families. Hence, both parents in the crossing block were subjected to half-sib selection, doubling the response from the between-family component, giving

$$R_{ER(m)} = \bar{i}_{(M,K)} \frac{\sigma_A^2/4}{\sigma(\bar{z}_{HS})} + \bar{i}_{(m,k)} \frac{(3/8)\sigma_A^2}{\sigma_{W(HS)}} \quad (21.49)$$

The Compton-Comstock modified ear-to-row scheme requires two generations per cycle, but offers increased response (per cycle) as the pollen is also from selected parents. Using the predicted values of Webel and Lonnquist (Example 21.10), the expected response per cycle under the Compton-Comstock design would be  $2 \cdot 4.6 + 3.8 = 13$ , for an expected 155% increase per cycle over the Lonnquist design (which had a predicted response 8.4). However, the Compton-Comstock design also requires two generations per cycle, so that the response per generation is 6.5, 77% of that expected under the Lonnquist design. The use of **off-season** (or winter) **nurseries**, where seeds are grown in either the opposite hemisphere or in the tropics (such as the Hawaiian island of Moloka'i) can allow for two generations in the same calendar year, but this may require more resources than the breeder has available.

## SELECTION ON A FAMILY INDEX

While our focus to this point has been on schemes that use either within- or between-family selection, the modified ear-to-row approach points out the advantage of using selection

schemes containing both within- and between-family components. Modified ear-to-row is an example of combined selection where the components are sequentially selected in different generations (and/or plots), and several such schemes are used by plant breeders. Alternatively, one can use both within- and between-family information to select individuals within a single generation. The most general way to do this is to select on a **family index**,

$$I = b_1 (z_{ij} - \bar{z}_i) + b_2 \bar{z}_i \quad (21.50a)$$

where the index value  $I$  is for an individual  $j$  from family  $i$ . Individuals with the largest index scores are mated (avoiding within-family crosses) to form the next generation. Note that individual ( $I = z_{ij}$ ), family ( $I = \bar{z}_i$ ) and family-deviation ( $I = z_{ij} - \bar{z}_i$ ) selections are all special cases of this general family index. The family index is often written as

$$I = z_{ij} + B \bar{z}_i \quad (21.50b)$$

where  $B$  is the relative weight on family mean compared to an individual's phenotype. As the reader can easily verify with a little algebra, this is equivalent to the index given by Equation 21.50a, with

$$B = \frac{b_2}{b_1} - 1 \quad (21.50c)$$

### Response to Selection on a Family Index

Once again, Equations 21.1a/21.4a can be used to predict the short-term (single-generation) response to selection. Taking  $x = I$  gives

$$R_I = \frac{\sigma(I, y | \mathcal{R}_1)}{\sigma_I^2} (S_{I_m} + S_{I_f}) = \bar{r}_I \sigma_z \rho(I, y) \quad (21.51)$$

The variances and covariances in Equation 21.51 are obtained as follows. Using the covariances summarized in Table 21.3,

$$\begin{aligned} \sigma(I, y | \mathcal{R}_1) &= b_1 \sigma(z_{ij} - \bar{z}_i, y | \mathcal{R}_1 = x_{ij}) + b_2 \sigma(\bar{z}_i, y | \mathcal{R}_1 = x_{ij}) \\ &= b_1(1 - r_n)(\sigma_A^2/2) + b_2 r_n(\sigma_A^2/2) \\ &= [b_1 + r_n(b_2 - b_1)] (\sigma_A^2/2) \end{aligned} \quad (21.52a)$$

Likewise, recalling that  $\sigma^2(x + y) = \sigma_x^2 + \sigma_y^2 + 2\sigma_{x,y}$  (LW Equation 3.11a), the variances summarized in Table 21.4 give

$$\begin{aligned} \sigma^2(I) &= b_1^2 \sigma^2(z_{ij} - \bar{z}_i) + b_2^2 \sigma^2(\bar{z}_i) + 2b_1 b_2 \sigma(z_{ij} - \bar{z}_i, \bar{z}_i) \\ &= b_1^2 (1 - t_n) \sigma_z^2 + b_2^2 t_n \sigma_z^2 + 2b_1 b_2 \sigma(z_{ij}, \bar{z}_i) - 2b_1 b_2 \sigma^2(\bar{z}_i) \\ &= (b_1^2 (1 - t_n) + b_2^2 t_n + 2b_1 b_2 t_n - 2b_1 b_2 t_n) \sigma_z^2 \\ &= [b_1^2 + t_n(b_2^2 - b_1^2)] \sigma_z^2 \end{aligned} \quad (21.52b)$$

The heritability of the index thus becomes

$$h_I^2 = \frac{2\sigma(I, y | \mathcal{R}_1)}{\sigma^2(I)} = h^2 \left[ \frac{b_1 + r_n(b_2 - b_1)}{b_1^2 + t_n(b_2^2 - b_1^2)} \right] \quad (21.53a)$$

Finally, the correlation between an individual's index score ( $I$ ) and the expected value of its offspring ( $y$ ) is

$$\rho(I, y) = \frac{2\sigma(I, y | \mathcal{R}_1)}{\sigma(I) \sigma(y)} = h^2 \left[ \frac{b_1 + r_n(b_2 - b_1)}{\sqrt{b_1^2 + t_n(b_2^2 - b_1^2)}} \right] \quad (21.53b)$$

The term in the brackets represents the accuracy of the index relative to mass selection. Substituting Equation 21.53b into Equation 21.51 (and recalling that  $\sigma_z h^2 = h\sigma_A$ ) gives the expected response as

$$R_I = \bar{r}_I h \sigma_A \frac{b_1 + r_n(b_2 - b_1)}{\sqrt{b_1^2 + t_n(b_2^2 - b_1^2)}} \quad (21.53c)$$

---

**Example 21.11.** Again consider the selection experiment of Clayton et al. (1957) on abdominal bristle number in *Drosophila* (Examples 21.4, 21.5). Here  $r_n = 0.542$ ,  $t_n = 0.326$ , and  $\sigma_A h = 1.70$ . Suppose individuals with index scores in the upper twenty percent are chosen. What is the expected response if we place three times the weight on within-family deviations as we do on family means ( $b_1 = 3$ ,  $b_2 = 1$ )? Since 20 families each with 12 sibs are scored, the expected selection intensity is  $\bar{r}_{240,48} = 1.39$ , and Equation 21.53c gives the expected response as

$$R_I = 1.39 \cdot 1.70 \left( \frac{3 + 0.542(1 - 3)}{\sqrt{3^2 + 0.326(1^2 - 3^2)}} \right) = 1.79$$

This is not as efficient as between-family selection (where  $R_b = 2.15$ , see Example 21.4). Likewise, the response under individual (mass) selection is  $R_m = \bar{r}_m \sigma_A h = 2.36$ . Since individual selection is a special case of the general index, we can always choose the index weights to give at least as large an expected response as individual selection. For example, placing twice the weight on family means as is placed on within-family deviation ( $b_1 = 1$ ,  $b_2 = 2$ ), gives an expected response of  $R = 2.59$ , 110% of the expected response under individual selection.

---

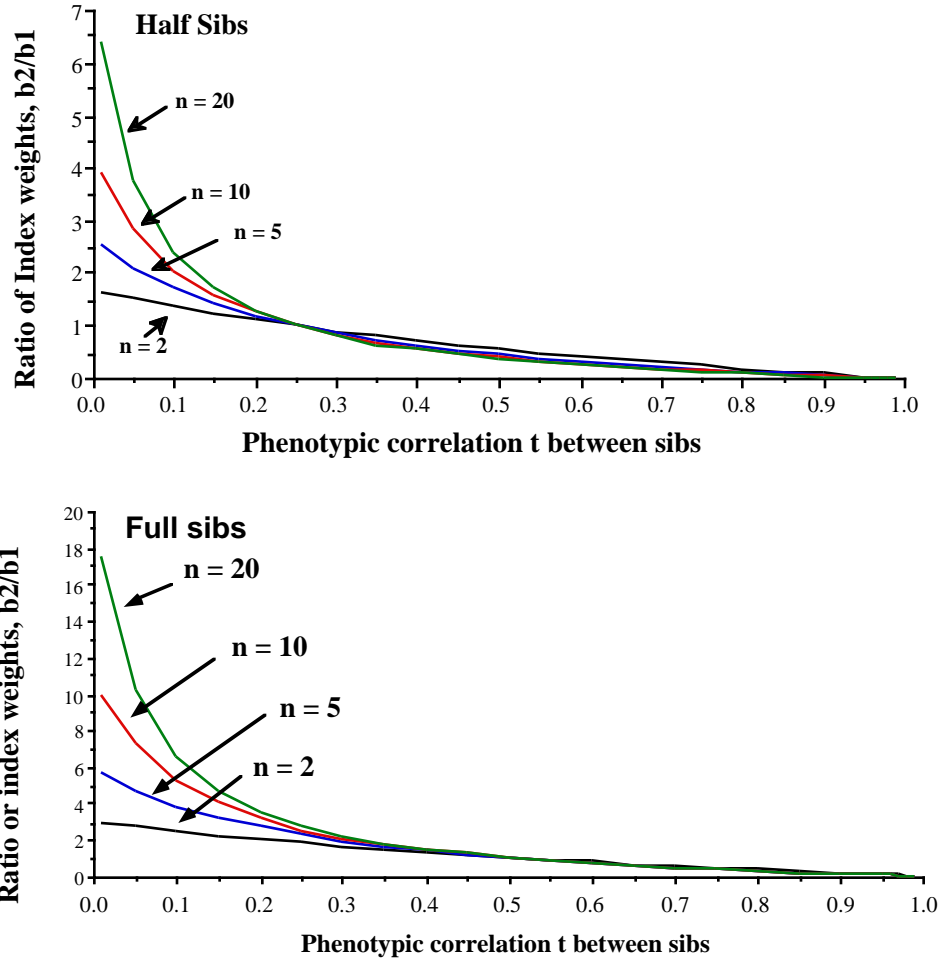
### Lush's Optimal Index

As the previous example shows, by the appropriate choice of index weights we can always obtain a response at least as large as under mass selection. Note from Equation 21.51 that  $\sigma_A$  and  $\bar{r}_I$  remain constant under different index weights, so that the maximal response occurs by choosing the weights that maximize the correlation  $\rho(I, y)$  between the index and offspring value (Equation 21.5 shows that this is equivalent to maximizing the correlation  $\rho(I, A)$  between the index and breeding values of an individual). Lush (1947) showed that the corresponding index weights are

$$b_1 = \frac{1 - r}{1 - t}, \quad b_2 = \frac{1 + (n - 1)r}{1 + (n - 1)t} \quad (21.54)$$

The formal derivation is given in Chapter 32. We refer to the family index using these weights as the **Lush Index**. Note that the weight  $b_1$  on family deviations is independent of the family size  $n$ , while the weight on the family mean  $b_2$  depends on  $n$ , approaching  $r/t$  for large families. Figure 21.7 plots the ratio of between- to within-family weights ( $b_2/b_1$ ) for the Lush index as a function of  $t$  and  $n$ . For small between-sib correlations, more weight is placed on family mean, while more weight is placed on within-family deviation when the sib correlation is large. We can rearrange the Lush index as  $I_L = z_{ij} + B_L \bar{z}_i$ , where Equation 21.50c gives

$$B_L = \frac{(r - t)n}{(1 - r)[1 + (n - 1)t]} \quad (21.55)$$



**Figure 21.7.** Ratio  $b_2/b_1$  of the weights placed on the between ( $b_2$ ) relative to within ( $b_1$ ) component under the optimal Lush index. These optimal weights are a function of the phenotypic correlation  $t$  between sibs and number  $n$  of sibs per family.

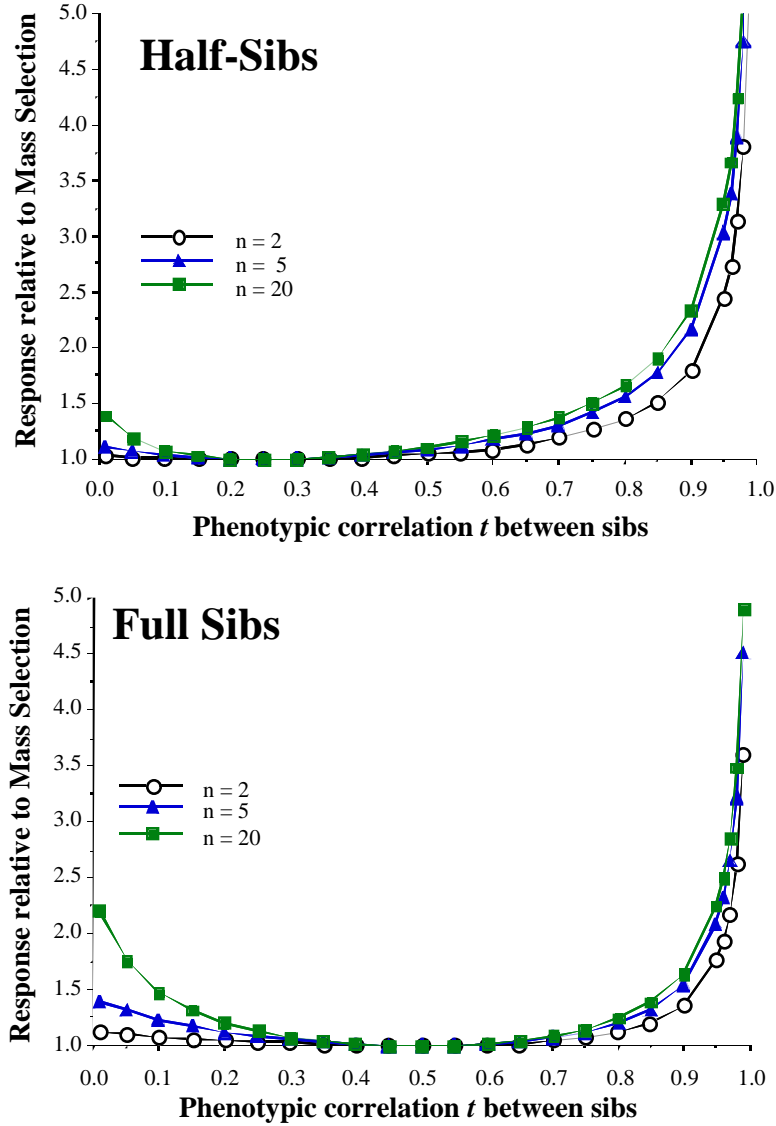
Using the optimal weights, Equation 21.51 simplifies to give the response under Lush's index as

$$R_{LI} = \bar{i} \sigma_A h \sqrt{1 + \frac{(r - t)^2(n - 1)}{(1 - t)[1 + (n - 1)t]}} \quad (21.56a)$$

The increase in response over that expected under individual selection is thus

$$\frac{R_{LI}}{R_m} = \sqrt{1 + \frac{(r - t)^2(n - 1)}{(1 - t)[1 + (n - 1)t]}} \quad (21.56b)$$

Figure 21.8 plots this ratio as a function of  $t$  and  $n$  for half- and full-sibs. Since the quantity in the square root exceeds one, the expected response under Lush's index exceeds the response under individual selection, except at  $r = t$  (i.e.,  $t = 0.25$  for half-sibs,  $t = 0.5$  for full-sibs) in which case the expected responses are equal.



**Figure 21.8.** Response of Lush's index relative to individual selection, as a function of the number of sibs  $n$ . Except at  $r = t$  (where the expected response equals that of individual selection), Lush's index gives a larger expected response than individual selection.

**Example 21.12.** Recalling (Example 21.4) that  $t = 0.265$  and  $r = 0.5$  for full-sibs in Clayton et al.'s bristle experiments, the resulting Lush weight on family deviations is

$$b_1 = \frac{1 - r}{1 - t} = \frac{1 - 0.5}{1 - 0.265} = 0.680$$

Further recalling that the family size was  $n = 12$ , the Lush weight on family means becomes

$$b_2 = \frac{1 + (n - 1)r}{1 + (n - 1)t} = \frac{1 + (12 - 1)0.5}{1 + (12 - 1)0.265} = 1.66$$

We can rescale the weights as  $b_1 = 1$  and  $b_2 = 1.66/0.680 = 2.44$ , giving a resulting response of

$$R_{LI} = \bar{v}_{LI} \sigma_A h \rho(I, y) = 1.39 \cdot 1.70 \cdot \left( \frac{1 + 0.542(2.44 - 1)}{\sqrt{1^2 + 0.326(2.44^2 - 1^2)}} \right) = 2.60$$

The expected response under individual selection is  $R_m = 2.36$  (Example 21.10), so that the expected response under the Lush index is 110% that of mass selection.

The Lush index weights hold for a single generation, and potentially need to be adjusted each generation as  $t$  and  $r$  change (due to potential changes in the genetic variances). Both drift and gametic-phase disequilibrium can be important when several generations of selection are considered. As selection proceeds both these forces increase the importance of within- versus between-family selection (Chapters 15, 22), so that individual value becomes weighted more and family mean less. Wray and Hill (1989) note that while the relative efficiency of index selection over individual selection may be greatly diminished by gametic-phase disequilibrium, the relative rankings of the methods still hold.

A second concern is that, as with any index, population parameters have to be correctly estimated or the index constructed from these estimates has incorrect weights and is less than optimal. For the Lush index only the intraclass correlation  $t$  must be estimated, and Sales and Hill (1976) have shown that the efficiency of index selection is quite robust to estimation errors in  $t$  (as initially suggested by Lush 1947).

Based on these concerns, it is not surprising that experimental verification of the advantage of the Lush index over individual or family selection is mixed. Further, the common problem of low statistical power in most selection experiments due to small sample sizes makes negative results difficult to interpret. McBride and Robertson (1963) and Avalos and Hill (1981) found that index selection gave a larger response than individual selection for abdominal bristles in *Drosophila melanogaster*. More conclusive results for selection on the same character were those of James (cited in Frankham 1982), who found that the observed increase in response under index selection (relative to mass selection) was  $133 \pm 9.7\%$  and  $111 \pm 7\%$  in two replicates, very consistent with the expected increase of 121%. Results for selection for egg production in poultry were less conclusive, with Kinney et al. (1970) finding that individual selection gave a larger (but not significant) response than index selection, while Garwood and Lowe (1981) found that index selection gave a larger response (again not significant) than family selection. Larval and pupal weight in *Tribolium* showed similar mixed results, with Wilson (1974) finding that individual selection gave the largest response, while Campo and Tagarro (1977) did not find any significant differences (index selection gave a larger response in a replicate with large family size, while individual selection showed the larger response in a replicate with small family size).

We note in passing that more a general family index was considered by Osborne (1957a, b) for the nested sib design which separately weights information from full- and half-sib families. If  $z_{ijk}$  denotes the  $k$ th full sib from dam  $j$  and sire  $i$ , an index weighting both half- and full-sib information is

$$I = b_1 (z_{ijk} - \bar{z}_{ij}) + b_2 (\bar{z}_{ij} - \bar{z}_{i\cdot}) + b_3 \bar{z}_{i\cdot}$$

where  $b_1$  is the weight on within full-sib family deviations,  $b_2$  the weight on deviations among dams within a sire, and  $b_3$  the sire weights (half-sib means). Chapter 32 examines this (and other indices) in some detail.

### Correcting the Selection Intensity for Correlated Variables



As mentioned previously, expressions for the selection intensity in finite populations make the assumption that the order statistics are uncorrelated. Selection of multiple individuals from the same family results in correlations among the order statistics due to the correlation between sibs. Our treatment follows that of Hill (1976, 1977).

Suppose the population from which individuals are drawn consists of  $m$  families each with  $n$  sibs, for a total of  $N = mn$  individuals. If phenotypic values are uncorrelated between all members of the sample (the sib correlation  $t$  is zero), Burrow's correction (Equation 10.17c) gives the finite-population size adjusted selection intensity as

$$\bar{i}_{N,K} = \bar{i}_p - \frac{1-p}{2\bar{i}_p p(N+1)}$$

where a fraction  $p = K/N$  of the population is saved and  $\bar{i}_p$  is the infinite-population selection intensity associated with the fraction  $p$  saved. When some members are correlated, the net result is to reduce the effective number of independent variables, so that the correct value is somewhat below  $N$ . This ranges from  $mn = N$  with no correlation between sibs ( $t = 0$ ) to  $m$  with a perfect correlation between sibs ( $t = 1$ ). Using this observation, Hill (1976) suggested the linear approximation for the effective number  $N_e$  of independent variables as

$$N_e = N(1-t) + mt \quad (21.57a)$$

Substituting into Burrow's correction gives the expected selection intensity adjusted for correlations as approximately

$$\bar{i}_{N,K}(t) = \bar{i}_p - \frac{1-p}{2\bar{i}_p p[N(1-t) + mt + 1]} \quad (21.57b)$$

Simulation studies by Hill shows that this is a reasonable approximation, and Hill (1976) provides tables of exact values (over a limited set of  $n$  and  $t$  values). Note that  $\bar{i}$  decreases as  $t$  increases, being most extreme when only a few families are chosen (here  $N_e$  approaches  $m$ ). An alternative approximation is offered by Rawlings (1976), while Tong (1982) considers unequal family size.

The effect of sib-correlations on the selection intensity for individual selection is generally small, as  $t$  is typically less than 0.5, which has only a modest effect on reducing  $\bar{i}$ . In contrast, the presence of the family mean  $\bar{z}_i$  in the index scores greatly inflates the correlation between the sib index values  $I$  over the correlation among phenotypic values  $z$ . Hill (1976) showed that if selection occurs on the index  $I = z_{ij} + B\bar{z}_i$  (Equation 21.50b), the intra-class correlation  $T$  among the index values of sibs is given by

$$T = 1 - \frac{n(1-t)}{n + B(2+B)[1 + (n-1)t]} \quad (21.58a)$$

where  $t$  is the intra-class correlation of individual phenotypic values among sibs. Note that for large  $B$ ,  $T$  approaches one. Hence, for schemes that place considerable weight on family means, the index scores for sibs within a family are almost perfectly corrected, and the effective number of independent order statistics approaches the number of families  $m$  chosen. This is very reasonable, as  $I$  approaches  $\bar{z}_i$  for large  $B$ , which is simply between-family selection. In this case the number of independent order statistics is simply the number of families.

Using the value of  $B$  (from Equation 21.55) under Lush index weights, Hill (1976) showed (for large  $n$ ) that

$$T_{Lush} \simeq \begin{cases} 1-t & \text{full-sibs} \\ \frac{1-t}{1+8t} \simeq \frac{1}{1+2h^2} & \text{half-sibs} \end{cases} \quad (21.58b)$$

---

**Example 21.13.** Once again, consider Clayton et al's experiment on *Drosopholia* bristle number. From Example 21.10, the Lush index weights are  $b_2/b_1 = 2.44$ , giving  $B = 1.44$ . Recalling that  $t = 0.265$  and  $n = 12$ , Equation 21.58a gives the correlation  $T$  among index values of sibs as

$$T = 1 - \frac{12(1 - 0.265)}{12 + 1.44(2 + 1.44)[1 + (12 - 1)0.265]} = 0.72$$

The correlation among index values of sibs is thus 2.7 times the correlation among sib phenotypic values. Note that under strict family selection ( $T = 1$ ), the correlation among the index value increases to 3.8 times the sib phenotypic correlation. Suppose we select on a family index using four families ( $m = 4$ ). The resulting total number of individuals becomes  $N = 12 \cdot 4 = 48$ , while the effective number of independent variables is

$$N_e = 48(1 - 0.72) + 4 \cdot 0.72 = 16.3$$

Thus the effective number is only 34% of the actual number of total individuals. Here  $p = 0.2$ , with  $\bar{i}_p = 1.40$ , giving the corrected selection intensity as

$$\bar{i} = 1.40 - \frac{1 - 0.2}{2 \cdot 1.40 \cdot 0.2 \cdot (16.3 + 1)} = 1.32$$


---