Genetic models with reduced penetrance related to the Y chromosome

R.W.R. Darling[1] and Tim Holt[2]

Running Head

Genetic models for hypertension

Abstract

Classical statistical genetics models of a quantitative trait depending on an autosomal gene indicate that father-to-daughter and mother-to-son correlations should be the same. If phenotypes are not sex-dependent, father-to-son and mother-to-daughter correlations also share this common value. On the other hand, if the gene is sex-linked, then the father-to-son correlation is zero.

Such models do not explain genetic variation in pulmonary artery pressure (PAP) of cattle – important because cattle with high PAP are known to develop Brisket Disease, pulmonary heart disease and congestive heart failure when taken to high altitudes. Data on 966 calves at a ranch in Colorado showed positive correlation (0.2) between sire PAP and male calf PAP, but slightly negative correlation (-0.01) between sire PAP and female calf PAP; the dam to male calf and dam to female calf correlations are both about 0.1.

The model presented here postulates an autosomal gene with reduced penetrance (i.e. the trait may remain at a normal level even when the genotype suggests abnormality), and that, in males, the rate of penetrance is related to an abnormality in the *Y* chromosome, and is therefore passed on from father to son. Then under plausible selective breeding assumptions, the pairwise correlation between fathers and daughters can become zero or negative. Explicit formulas are computed for the model covariances, and numerical computations indicate that plausible parameter values can be chosen for the model.

AMS (1991) Subject Classification

Primary: 92D10. Secondary: 62P10

Key Words

familial correlations, quantitative character, brisket disease, pulmonary hypertension

1. Mathematics Department, University of South Florida PHY 114, Tampa, Florida 33620-5700
2. (Corresponding author) Town and Country Animal Hospital, 1525 Highway 135, Gunnison, Colorado 81230

1 Pulmonary Artery Pressure in Cattle: Introduction to the Data

1.1 A Puzzling Parent-to-Offspring Correlation Structure

Holt (1997) describes a procedure for measuring pulmonary artery pressure (PAP) in cattle at high altitudes (i.e. 5,000 feet or more). It has been known for some time that animals with elevated PAP at one year of age are prone to develop brisket disease (see Alexander and Will (1963)), pulmonary heart disease and congestive heart failure when taken to high altitudes.

Schimmel's (1981) study of 667 high altitude calves in the years 1974 - 1980 demonstrated the genetic component of PAP, using analysis of variance and regression techniques. Holt, Darling, Bittle, Miller, and Ramirez (1997) consider a data set consisting of 481 female and 485 male or castrated calves, living on a ranch in Colorado at an elevation of 7,000 feet. The PAP scores of these calves, and of their parents, were measured at approximately one year of age. It was found that the genetic component in high PAP scores is highly statistically significant (p <0.0001).

Schimmel (1981) did not present a genetic model. Weir et al. (1974) propose an autosomal dominant gene, but our statistical results are inconsistent with that hypothesis. The present paper presents a genetic model which seems to be capable of explaining certain unusual statistical characteristics of the data set of Holt et al (1997).

When a quantitative character depends upon an autosomal gene whose genotypes are in Hardy-Weinberg equilibrium, standard theory dictates that father-to-daughter and mother-to-son correlations take the same value, ρ say; indeed if the levels of the trait depend only on the genotype, and not on sex, then father-to-son and mother-to-daughter correlations are also equal to ρ; see Ewens (1979), Section 1.3. Moreover Hogben (1932) showed that, for a sex-linked gene, the father-to-son correlation is zero, while the father-to-daughter and mother-to-son correlations are equal. Even in more elaborate sex-linked models such as that of Leach and Mayo (1967), father-to-son correlation is still zero.]

TABLE 1 PAP Correlations[a]

	Male Calf (N = 485)	*Female Calf (N = 481)*
Sire (N = 67)	0.199	-0.011[b]
Dam (N = 266)	0.106	0.108

a. Computed pairwise; see Eliasziw and Donner (1990), and Vogler et al (1995) for a discussion of statistical issues.

b. The significance level of the column difference is hard to compute because of the dependence between pairs sharing a common parent.

However in this case, although sire PAP is positively correlated with male calf PAP, it has zero or negative correlation with female calf PAP, as shown in TABLE 1, in contrast

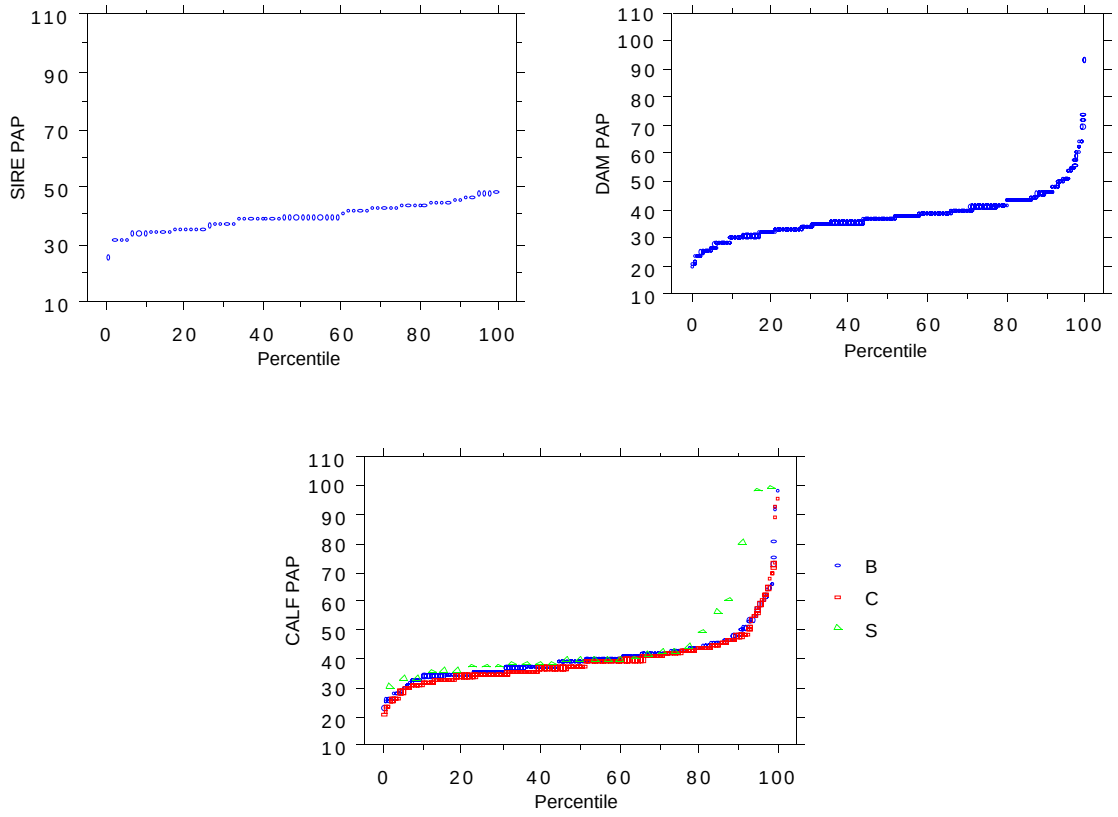


FIGURE 1 Empirical cumulative distribution functions of PAP scores for dams, sires, and calves (B = male; C = female; S = steer)

to the traditional models above. [Mayo et al (1990) present a model for a balanced polymorphism where such negative correlations are possible, but which has other features which appear unnatural here, as discussed further below.]

It is noteworthy that Schimmel (1981) did not observe a different regression coefficient for male calf PAP with respect to sire PAP to the one for female calf PAP with respect to sire PAP. We conjecture that the reason for this is that, twenty years ago, high PAP sires were still used for breeding, and the selective mating policies mentioned in Sections 2.2 and 3.2 were not in force; this would lead to correlation and regression coefficients which were less sex-dependent than the ones that we observed.

1.2 First Clue: PAP Scores of Sires are Atypical

In order to frame a suitable genetic hypothesis, some statistical characteristics of the PAP data need to be mentioned. FIGURE 1 presents the empirical cumulative distribution functions of PAP scores for dams, sires, and calves (B denotes bull, or male calf; C denotes cow, or female calf; S denotes steer, i.e. castrated male), respectively. A comparison of the empirical distribution functions of PAP scores for male calves and for

female calves showed no significant difference between the two (bottom graph), except that very high values seem less common among steers (which hereafter are pooled with male calves). Likewise PAP scores for dams and for calves show similar distributions. However the PAP distribution for the 67 sires is evidently truncated so as to exclude sires with PAPs over 49, or less than 32.

1.3 Environmental Factors Which Might Have Affected PAP

Holt (1997) has found that PAP scores taken at 11- 13 months of age are very reliable and highly repeatable. Animals used in this study are range cattle and treated as such. These cattle were not kept or raised in a feedlot or feed test situation. This issue is relevant because Dr. Joe Schimmel has shown that diet and ration type can indeed affect the PAP measurement. All cattle in this group, male and female alike, received virtually the same treatment.

2 First Modeling Attempt - A Selective Mating Hypothesis

We shall first attempt to explain the data using a relatively simple model; it turns out to be wrong, but the calculations involved are extremely helpful in preparing for the subsequent model.

Let us postulate, following Weir et al. (1974), that PAP scores are related to an autosomal gene, where we denote the normal gene by R^+, and the gene associated with high risk of brisket disease (and elevated PAP) by *R*. We further suppose that the genotypes R^+R^+, R^+R, and *RR* are in Hardy-Weinberg equilibrium (or, at least, were in equilibrium until the last generation or so), with relative frequencies $(1-p)^2$, $2p(1-p)$, and p^2, respectively, for some $0 < p < 1$. We do not assume that *R* is either recessive or dominant, but that the mean PAP score is different for each genotype, and possibly for each gender; thus mean PAP scores are as shown in the following table, where we assume $\mu_0 \le \mu_1 \le \mu_2$ and $\phi_0 \le \phi_1 \le \phi_2$:

TABLE 2 Genotypes and Phenotypes for the First Model

		Male Phenotypes	Female Phenotypes
Genotype	Relative Frequency	Mean PAP	Mean PAP
R^+R^+	$(1-p)^2$	μ_0	ϕ_0
R^+R	$2p(1-p)$	μ_1	ϕ_1
RR	p^2	μ_2	ϕ_2

In TABLE 3 we introduce some integer-valued random variables to assist in our calculations. It is assumed that these variables are independent, apart from the dependencies

TABLE 3 Genetic Random Variable Definitions

Variable Name	Meaning	Possible Values	Probability Distribution
N_S, N_D	Number of *R* genes in the sire and dam, respectively	0, 1, 2	Binomial(2, *p*)
U_S	Number of *R* genes transmitted from sire to male calf	0, 1	Bernoulli($N_S/2$), conditional on N_S
U_D	Number of *R* genes transmitted from dam to male calf	0, 1	Bernoulli($N_D/2$), conditional on N_D
V_S	Number of *R* genes transmitted from sire to female calf	0, 1	Bernoulli($N_S/2$), conditional on N_S
V_D	Number of *R* genes transmitted from dam to female calf	0, 1	Bernoulli($N_D/2$), conditional on N_D

noted in the right column. We assume the usual system of inheritance (see Kempthorne (1957)), in which a male calf inherits $U_S + U_D$ *R* genes from his parents, and a female calf inherits $V_S + V_D$ *R* genes from her parents.

2.1 Additive Genetic and Environmental Factors for the Simplistic Model

See Kempthorne (1957) for a general discussion of the modeling of quantitative characters. We postulate that the PAP score Y_S of the sire can be written as

$$Y_S = \mu_0 1_{\{N_S = 0\}} + \mu_1 1_{\{N_S = 1\}} + \mu_2 1_{\{N_S = 2\}} + \varepsilon_S,$$

where the "residual" ε_S is a random variable with mean zero, representing environmental and other factors affecting the PAP score. This can be abbreviated to

$$Y_S = \mu_0 + \Delta_1 1_{\{N_S = 1\}} + \Delta_2 1_{\{N_S = 2\}} + \varepsilon_S, \quad (1)$$

where $\Delta_i \equiv \mu_i - \mu_0 \geq 0$. Likewise the PAP score Y_D of the dam can be written as

$$Y_D = \phi_0 + \Lambda_1 1_{\{N_D = 1\}} + \Lambda_2 1_{\{N_D = 2\}} + \varepsilon_D, \quad (2)$$

where $\Lambda_i \equiv \phi_i - \phi_0 \geq 0$. The PAP scores Y_B of a bull calf and Y_C of a female calf are

$$Y_B = \mu_0 + \Delta_1 1_{\{U_S + U_D = 1\}} + \Delta_2 1_{\{U_S + U_D = 2\}} + \varepsilon_B. \quad (3)$$

$$Y_C = \phi_0 + \Lambda_1 1_{\{V_S + V_D = 1\}} + \Lambda_2 1_{\{V_S + V_D = 2\}} + \varepsilon_C. \quad (4)$$

The residuals $\{\varepsilon_S, \varepsilon_D, \varepsilon_B, \varepsilon_C\}$ are assumed to be independent of each other, and of all genetic random variables. Their variances are assumed to be finite, but possibly unequal. Note that no further distributional assumptions are made about the residuals; hence we are not in a position to maximize a likelihood to estimate parameters.

Genetic models with reduced penetrance related to the Y chromosome

2.2 Selective Mating Assumption

As noted in Section 1.2, comparison of the empirical probability distributions of PAP scores of sires, dams, male calves, and female calves, at the ranch showed evidence that bulls with PAP scores over 49 were not used for breeding in this ranch, although PAP scores above this level occur in roughly ten per cent of male (or female) calves. We shall model this phenomenon as follows.

Let *A* denote the event that an adult male (which factors other than PAP would render eligible for breeding) is considered acceptable for breeding. Our probability model includes the following **selective mating assumptions**:

$$P(A|N_S = 2) = 0; \tag{5}$$

$$P(A|N_S = 1) = \gamma; \tag{6}$$

$$P(A|N_S = 0) = 1. \tag{7}$$

Assumption (5) says, neglecting the effect of the "residuals", that very high PAP adult males are not selected for breeding; (6) says that adult males with somewhat elevated PAP, and otherwise having desirable characteristics, have probability γ of being selected; (7) says that adult males with low PAP, and otherwise having desirable characteristics, have probability 1 of being selected. [For the sake of simplicity we are neglecting the fact that males with PAP below 32 appear not to be selected either, since their numbers are rather small.]

2.3 Theorem: Parent to Offspring Covariances – Simplistic Model

Under the assumptions described above, the conditional covariances of PAP scores are:

$$\frac{\mathrm{Cov}(Y_S, Y_B|A)}{\Delta_1^2} = \frac{\alpha\theta\gamma}{q+2p\gamma}; \frac{\mathrm{Cov}(Y_S, Y_C|A)}{\Delta_1\Lambda_1} = \frac{\alpha'\theta\gamma}{q+2p\gamma}; \tag{8}$$

$$\frac{\mathrm{Cov}(Y_D, Y_B|A)}{\Delta_1\Lambda_1} = \theta\alpha'\beta; \frac{\mathrm{Cov}(Y_D, Y_C|A)}{\Lambda_1^2} = \theta\alpha'\beta' \tag{9}$$

where $q = 1 - p$ *and*

$$\alpha \equiv 1 - 2p + \frac{\Delta_2}{\Delta_1}p,\ \alpha' \equiv 1 - 2p + \frac{\Lambda_2}{\Lambda_1}p, \tag{10}$$

$$\beta \equiv q + \frac{\Delta_2}{\Delta_1}p\gamma,\ \beta' \equiv q + \frac{\Lambda_2}{\Lambda_1}p\gamma, \tag{11}$$

$$\theta \equiv \frac{pq}{q+2p\gamma}. \tag{12}$$

2.3.1 Remarks

Our statistical data indicates that $\Delta_i \approx \Lambda_i$, $i = 1, 2$, and $\mathrm{Var}(Y_B|A) \approx \mathrm{Var}(Y_C|A)$. Hence $\alpha \approx \alpha'$, and (8) would imply that the correlation between sire PAP and male calf PAP is about the same as the correlation between sire PAP and female calf PAP, and that both are positive, which is inconsistent with the correlation data in TABLE 1. Hence we are forced to dismiss this model, based on an autosomal gene and selective mating alone, as inadequate.

A possible variation on this model, along the lines of Mayo et al. (1990), would be to allow $\mu_1 < \mu_0$ or $\mu_1 > \mu_2$, and likewise for the $\{\phi_i\}$. The selective mating hypothesis would no longer make sense as it stands, and would have to be rephrased. This may permit the correlation between sire PAP and female calf PAP to become zero or negative, but would still preserve approximate equality between the correlation between sire PAP and male calf PAP and the correlation between sire PAP and female calf PAP, which is inconsistent with the data.

Proof of Theorem 2.3: These formulas follow from Theorem 3.4 below, on taking $t = \eta = \pi = 1$. However they were in fact computed separately from the computations for Theorem 3.4, and serve as an independent check on the validity of the formulas therein.

3 A Model with Reduced Penetrance, Depending on Y

We now add two new features to the model of Section 2, which produce a model capable of explaining TABLE 1.

3.1 New Features in the Model

3.1.1 Reduced penetrance in the female:

This means that a female with genotype R^+R or *RR* will manifest a medium or high PAP score, respectively, with a probability π, and a low PAP score with a probability 1 - π, independently of all other variables in the model. The idea that reduced penetrance was present is due to geneticist Wilmer J. Miller; for other examples, see Hollander and Miller (1978).

3.1.2 Reduced penetrance in the male, depending on abnormality in the *Y* chromosome:

We postulate that a proportion 1 - *t* of males have a normal *Y* chromosome, and a proportion *t* have an abnormal version denoted Y^*. Normal males always have low PAP, regardless of whether they are R^+R^+, R^+R, or *RR*. Males with the Y^* chromosome and with R^+R or *RR* have conditional probability η of manifesting medium or high PAP score, respectively. In effect the penetrance parameter for males is $t\eta$ (which in practical calculations we take to be equal to π), but male penetrance becomes a characteristic partly inherited from the sire, with no dependence on the dam at all.

Genetic models with reduced penetrance related to the Y chromosome

In TABLE 4 we add new random variables to those already introduced in TABLE 3:

TABLE 4 More Genetic Random Variable Definitions

Variable Name	Meaning	Possible Values	Probability Distribution
T	Number of Y^* genes in the sire	0, 1	Bernoulli(t)
H_S, H_B,	Indicator of penetration of the R gene in the sire and male calf, respectively	0, 1	Bernoulli(η)
H_D, H_C	Indicator of penetration of the R gene in the dam and female calf, respectively	0, 1	Bernoulli(π)

3.2 Revised Selective Mating Hypothesis

Let $W_i \equiv \{T = 1, H_S = 1, N_S = i\}$, for $i = 1, 2$. Thus $W_1 \cup W_2$ is the event that the sire has elevated PAP. Let A denote the event that an adult male (which factors other than PAP would render eligible for breeding) is accepted for breeding. We replace (5), (6), and (7) by the following **selective mating assumptions**:

$$P(A|W_2) = 0; \tag{13}$$

$$P(A|W_1) = \gamma; \tag{14}$$

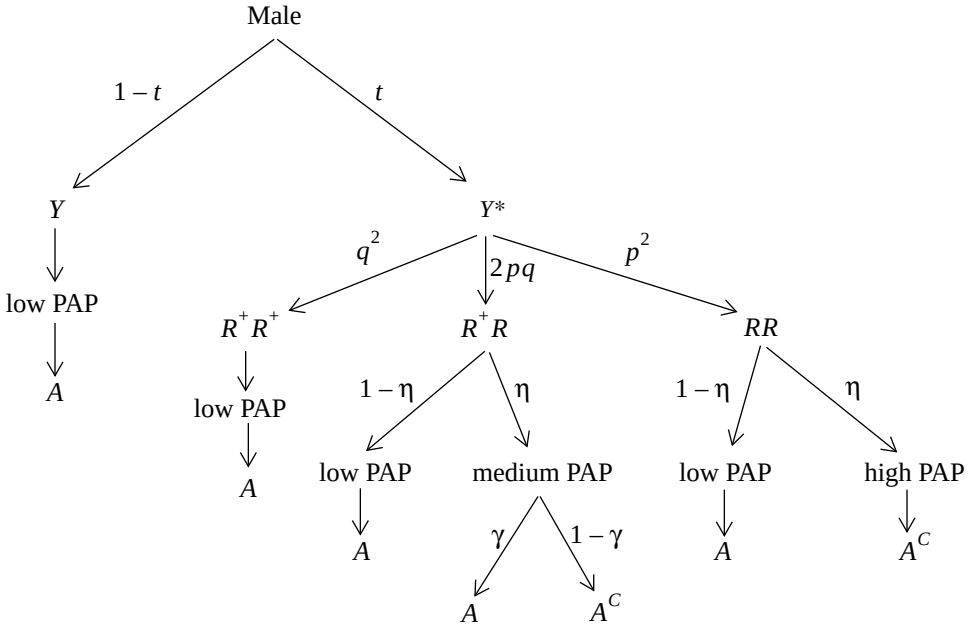
$$P(A|(W_1 \cup W_2)^c) = 1. \tag{15}$$

Assumption (13) says that high PAP adult males are not selected for breeding; (14) says that adult males with somewhat elevated PAP have probability γ of being selected; (15) says that low PAP adult males are always selected.

Let us note that the random variables N_D, H_D, U_D, V_D, H_C, H_B are independent of A. However T, H_S, N_S are clearly dependent on A, and hence are so are U_S, V_S. The formulas for the conditional probabilities will be derived in Section 5.

The model for males can now be summarized in a tree diagram as in FIGURE 2; the model for females is similar, without the Y versus Y^* dichotomy.

FIGURE 2 Tree diagram showing genotypes, phenotypes, and breeding acceptability (*A*) for males



3.3 Additive Genetic and Environmental Factors for the Full Model

In the notation of TABLE 3 and TABLE 4, the PAP score Y_S of the sire can be written:

$$Y_S = \mu_0 + \Delta_1 1_{\{N_S = 1,\, T = 1,\, H_S = 1\}} + \Delta_2 1_{\{N_S = 2,\, T = 1,\, H_S = 1\}} + \varepsilon_S ,$$

where the "residual" ε_S is as in (1), and $\Delta_i \equiv \mu_i - \mu_0$. This can be abbreviated to

$$Y_S = \mu_0 + TH_S [\Delta_1 1_{\{N_S = 1\}} + \Delta_2 1_{\{N_S = 2\}}] + \varepsilon_S . \qquad \textbf{(16)}$$

Likewise the PAP score Y_D of the dam can be written as

$$Y_D = \phi_0 + H_D [\Lambda_1 1_{\{N_D = 1\}} + \Lambda_2 1_{\{N_D = 2\}}] + \varepsilon_D , \qquad \textbf{(17)}$$

where $\Lambda_i \equiv \phi_i - \phi_0$. The PAP scores Y_B of a male calf and Y_C of a female calf are

$$Y_B = \mu_0 + TH_B [\Delta_1 1_{\{U_S + U_D = 1\}} + \Delta_2 1_{\{U_S + U_D = 2\}}] + \varepsilon_B . \qquad \textbf{(18)}$$

$$Y_C = \phi_0 + H_C [\Lambda_1 1_{\{V_S + V_D = 1\}} + \Lambda_2 1_{\{V_S + V_D = 2\}}] + \varepsilon_C . \qquad \textbf{(19)}$$

The residuals $\{\varepsilon_S, \varepsilon_D, \varepsilon_B, \varepsilon_C\}$ have the properties described in Section 2.1.

3.4 Theorem: Parent to Offspring Covariances – Full Model

Under the assumptions described above, the conditional covariances of PAP scores are:

$$\frac{\mathrm{Cov}\,(Y_S, Y_B | A)}{\Delta_1^2} = \frac{t\eta^2 pq\gamma}{P(A)^2}\left[t\alpha\zeta + (1-t)\left(1 + \frac{\Delta_2}{\Delta_1} p\right)\right]; \qquad \textbf{(20)}$$

$$\frac{\mathrm{Cov}\,(Y_S, Y_C|A)}{\Delta_1\Lambda_1} = \frac{\pi t\eta pq\gamma\xi\alpha'}{P\,(A)^2}; \tag{21}$$

$$\frac{\mathrm{Cov}\,(Y_D, Y_B|A)}{\Delta_1\Lambda_1} = \frac{\pi t\eta pq\alpha'}{P\,(A)}\{\zeta + \frac{\Delta_2}{\Delta_1}p\,[1+\eta\,(q\gamma-1)\,]\}\,; \tag{22}$$

$$\frac{\mathrm{Cov}\,(Y_D, Y_C|A)}{\Lambda_1^2} = \frac{\pi^2 pq\alpha'}{P\,(A)}\{\xi + \frac{\Lambda_2}{\Lambda_1}p\,[1+t\eta\,(q\gamma-1)\,]\}\,, \tag{23}$$

where $q = 1-p$ *and*

$$\alpha \equiv 1-2p+\frac{\Delta_2}{\Delta_1}p\,,\ \alpha' \equiv 1-2p+\frac{\Lambda_2}{\Lambda_1}p\,, \tag{24}$$

$$\zeta \equiv 1-2p+\eta p^2\,,\ \xi \equiv 1-2p+t\eta p^2\,, \tag{25}$$

$$P\,(A) = 1+t\eta p\,(p+2q\gamma-2)\,. \tag{26}$$

3.4.1 Remarks

(a) The central point of this paper, mathematically speaking, is that ξ is negative when $p>1/2$ and $t\eta$ is less than $(2p-1)/p^2$; this fact allows the sire-to-daughter covariance in (21) to be negative.

(b) Note that, when $t=\eta=\pi=1$, $P\,(A)$ becomes $q\,(q+2p\gamma)$, and $\zeta=\xi=q^2$. It is straightforward to check that formulas (20) through (23) reduce to formulas (8) and (9) of Theorem 2.3.

(c) The proof of Theorem 3.4 will be deferred to Section 5.

Now we derive a dimension-free version of the formulas above, not by considering correlations, which has the unfortunate effect of introducing the unknown variances of the residuals, but by taking ratios of the covariances under certain simplifying assumptions.

3.5 Corollary: Ratios of the Covariances

Assume that $\pi = t\eta$, *and* $\Lambda_2=\Delta_2=\rho\Delta_1=\rho\Lambda_1$ *for some* $\rho>1$. *Then* $\alpha=\alpha'$, *and*

$$\frac{\mathrm{Cov}\,(Y_S, Y_C|A)}{\mathrm{Cov}\,(Y_S, Y_B|A)} = \frac{t\alpha\xi}{t\alpha\zeta+(1-t)\,(1+\rho p)}; \tag{27}$$

$$\frac{\mathrm{Cov}\,(Y_D, Y_C|A)}{\mathrm{Cov}\,(Y_D, Y_B|A)} = \frac{\xi+\rho p\,[1+\pi\,(q\gamma-1)\,]}{\zeta+\rho p\,[1+\eta\,(q\gamma-1)\,]}; \tag{28}$$

$$\frac{\mathrm{Cov}\,(Y_D, Y_B|A)}{\mathrm{Cov}\,(Y_S, Y_B|A)} = \frac{t\alpha P\,(A)\,\{\zeta+\rho p\,[1+\eta\,(q\gamma-1)\,]\}}{\gamma\,[t\alpha\zeta+(1-t)\,(1+\rho p)\,]}. \tag{29}$$

4 Numerical Results

We performed the following computational experiment in MATLAB®. Let R_1, R_2, R_3 denote the numerical values, obtained from the dataset described in Section 1, for the three covariance ratios represented in equations (27), (28), and (29) respectively. For $i = 1, 2, 3$, let $f_i(p, t, \eta, \gamma, \rho)$ denote the function represented by the right side of equations (27), (28), and (29) respectively. Our numerical routine first selected at random some initial values for the variables p, t, η, γ, ρ; the first four were selected uniformly on $(0.5, 1)$, and subsequently constrained to $(0, 1)$, and ρ was selected uniformly on (1.5, 3.5), and subsequently constrained to $(1.5, 10)$. Next a straightforward steepest descent algorithm was used to minimize the objective function

$$\sum_{i=1}^{3} \left| f_i(p, t, \eta, \gamma, \rho) - R_i \right| . \qquad (30)$$

We ran the program dozens of times from different randomly selected initial values, to avoid the problem of being trapped in a local minimum. The results indicate that there are indeed many local minima of (30), and there is absolutely no guarantee that we have located the global minimum. Moreover there is a range of parameter values in which the objective function does not appear to vary a great deal. For these reasons we cannot attribute any reliability to our computations of the unknown parameters. The word “estimate” is inappropriate, because we derive no confidence intervals. All that can be claimed is that the model is capable of producing covariance ratios which match reasonably well the ratios seen in the data set – see TABLE 5 – for parameter values which seem to be statistically plausible, namely $p = 0.65$, $t = 0.83$, $\eta = 0.77$, and $\rho = 2.00$:

TABLE 5 Fit between observed and modeled covariance ratios

Observed	Computed from the Model
$R_1 = -0.0544$	$f_1 = -0.0539$
$R_2 = 0.9258$	$f_2 = 1.0824$
$R_3 = 0.9811$	$f_3 = 0.9811$

Here we are still assuming that $\pi = t\eta$, so $\pi = 0.64$.

4.1 Phenotype Frequencies Under the Computed Model

The TABLE 6 is based on the figures above, and the hypothesis that R^+R, and RR cattle will only manifest elevated PAP if the R gene penetrates, which in females has probability π, and in males has probability $t\eta$ which is also supposed equal to π. Our figure $\rho = 2$ means that the difference between High PAP and Low PAP is twice the difference between Medium PAP and Low PAP.

TABLE 6 Phenotype Frequencies

Genotype	Phenotypes	Theoretical Relative Frequency	Computed Value
R^+R^+	Low PAP	$q^2 + (1-q^2)(1-\pi)$	0.46
R^+R	Medium PAP	$2pq\pi$	0.29
RR	High PAP	$p^2\pi$	0.27

4.2 Comparison with the Data

In the data set, 23% of male calves and 21% of female calves had PAP scores of 44 or more, versus medians of 39 and 38 respectively. Our model, by giving only three possible values to PAP (before adding the residual) does not adequately explain the 5% or so of calves with PAP scores over 50, unless these extremely high values are considered to have environmental, not genetic causes. Another flaw in the model is that it only explains the upper cutoff (approximately 49) for sire PAP, but not the lower cutoff (only one sire below 32). These defects could be resolved by introducing a more complicated model. However the model is successful in giving a theoretical explanation for the correlation figures shown in TABLE 1.

5 Proof of Theorem 3.4

To keep this article to a reasonable length, only the proof of (20) will be given in full detail; the proofs of (21), (22), and (23) will be give in outline only. The proof demands a series of related steps.

5.1 Preliminary Calculations

5.1.1 Probability of Acceptance

Let $W_i \equiv \{T = 1, H_S = 1, N_S = i\}$, for $i = 1, 2$. If $q \equiv 1 - p$,

$$P(W_1) = P(T=1)\,P(H_S=1)\,P(N_S=1) = 2pqt\eta; \tag{31}$$

$$P(W_2) = p^2 t\eta;$$

$$P(A) = P(A|(W_1 \cup W_2)^c)\,P((W_1 \cup W_2)^c) + P(A|W_1)\,P(W_1)$$

$$= P((W_1 \cup W_2)^c) + \gamma P(W_1)$$

$$= 1 - (2pqt\eta + p^2 t\eta) + 2pqt\eta\gamma.$$

Hence

$$P(A) = 1 + t\eta p(p + 2q\gamma - 2) . \quad (32)$$

5.1.2 Transmission of *R* Genes from the Sire

$$P(A, T = 1, H_S = 1, N_S = 2) = 0;$$

$$P(A, T = 1, H_S = 1, N_S = 1) = P(A|W_1) P(W_1) = 2pqt\eta\gamma;$$

$$P(A, T = i, H_S = j, N_S = 0) = P(T = i) P(H_S = j) q^2;$$

$$P(A, T = i, H_S = 0, N_S = k) = P(T = i) (1 - \eta) P(N_S = k);$$

$$P(A, T = 0, H_S = j, N_S = k) = (1 - t) P(H_S = j) P(N_S = k) .$$

Hence

$$P(V_S = 1, A) = \sum_i \sum_j \sum_k P(A, V_S = 1, T = i, H_S = j, N_S = k)$$

$$= \frac{1}{2} \sum_i \sum_j P(A, T = i, H_S = j, N_S = 1) + \sum_i \sum_j P(A, T = i, H_S = j, N_S = 2)$$

$$= \frac{1}{2} \{ \sum_j P(A, T = 1, H_S = j, N_S = 1) + \sum_j P(A, T = 0, H_S = j, N_S = 1) \}$$

$$+ \sum_j P(A, T = 1, H_S = j, N_S = 2) + \sum_j P(A, T = 0, H_S = j, N_S = 2)$$

$$= \frac{\{2pqt\eta\gamma + 2pqt(1 - \eta) + 2pq(1 - t)\}}{2} + p^2 t(1 - \eta) + p^2 (1 - t)$$

Hence

$$P(V_S = 1, A) = p[1 + t\eta(\gamma q - 1)] = P(U_S = 1, A); \quad (33)$$

$$\mathrm{E}[U_S | A] = \mathrm{E}[V_S | A] = \frac{p[1 - t\eta(1 - \gamma q)]}{P(A)} . \quad (34)$$

5.1.3 Transmission of *R* Genes from the Dam

$$\mathrm{E}[U_D | A] = \mathrm{E}[D/2] = p = \mathrm{E}[V_D | A] . \quad (35)$$

5.2 Conditional Expectations of Random Variables Involved in PAP Scores

5.2.1 Conditional Expectations for the Sire

$$\mathrm{E}[TH_S 1_{\{N_S = 1\}} | A] = \frac{P(A \cap W_1)}{P(A)} = \frac{P(A|W_1) P(W_1)}{P(A)} = \frac{2pqt\eta\gamma}{P(A)}; \quad (36)$$

$$\mathrm{E}[TH_S 1_{\{N_S = 2\}} | A] = \frac{P(A \cap W_2)}{P(A)} = 0 . \quad (37)$$

5.2.2 Conditional Expectations for the Dam

$$\mathrm{E}[H_D 1_{\{N_D = 1\}} | A] = \mathrm{E}[H_D 1_{\{N_D = 1\}}] = 2pq\pi; \quad (38)$$

$$\mathrm{E}\left[H_D 1_{\{N_D=2\}} \middle| A\right] = \mathrm{E}\left[H_D 1_{\{N_D=2\}}\right] = p^2\pi . \tag{39}$$

5.2.3 Conditional Expectations for the Male Calf

$$\mathrm{E}\left[TH_B 1_{\{U_S+U_D=1\}} \middle| A\right] = \frac{P(A, T=1, H_B=1, U_S+U_D=1)}{P(A)}$$

$$= \frac{\eta}{P(A)}\{P(A, T=1, U_S=1, U_D=0) + P(A, T=1, U_S=0, U_D=1)\}$$

$$= \frac{\eta}{P(A)}\{qP(A, T=1, U_S=1) + pP(A, T=1, U_S=0)\} . \tag{40}$$

Now

$$P(A, T=1, U_S=1) = \sum_j \sum_k P(A, U_S=1, T=1, H_S=j, N_S=k)$$

$$= \sum_{k=1}^{2} P(U_S=1, T=1, H_S=0, N_S=k) + \gamma P(U_S=1, T=1, H_S=1, N_S=1) ;$$

$$P(A, T=1, U_S=1) = t(1-\eta)p + \gamma t\eta pq .$$

Also

$$P(A, T=1, U_S=0) = \sum_j \sum_k P(A, U_S=0, T=1, H_S=j, N_S=k)$$

$$= \sum_{k=0}^{1} P(U_S=0, T=1, H_S=0, N_S=k) + \gamma P(U_S=0, T=1, H_S=1, N_S=1)$$

$$+ P(U_S=0, T=1, H_S=1, N_S=0) ;$$

$$P(A, T=1, U_S=0) = t(1-\eta)q + \gamma t\eta pq + t\eta q^2 .$$

Thus (40) becomes

$$\frac{t\eta}{P(A)}\{q[(1-\eta)p + \gamma\eta pq] + p[(1-\eta)q + \gamma\eta pq + \eta q^2]\} ;$$

$$\mathrm{E}\left[TH_B 1_{\{U_S+U_D=1\}} \middle| A\right] = \frac{t\eta pq}{P(A)}\{2 - \eta(2-\gamma-q)\} . \tag{41}$$

Finally

$$\mathrm{E}\left[TH_B 1_{\{U_S+U_D=2\}} \middle| A\right] = \frac{p\eta P(A, T=1, U_S=1)}{P(A)}$$

$$= \frac{p\eta[t(1-\eta)p + \gamma t\eta pq]}{P(A)} ;$$

$$\mathrm{E}\left[TH_B 1_{\{U_S+U_D=2\}} \middle| A\right] = \frac{\eta t p^2}{P(A)}[1 - \eta + \eta\gamma q] . \tag{42}$$

5.2.4 Conditional Expectations for the Female Calf

Omitting the calculations,

$$\mathrm{E}\left[H_C 1_{\{V_S+V_D=1\}}\middle|A\right] = \frac{\pi pq}{P(A)}\left[2-t\eta(1+p-\gamma)\right]; \tag{43}$$

$$\mathrm{E}\left[H_C 1_{\{V_S+V_D=2\}}\middle|A\right] = p\pi\mathrm{E}\left[V_S|A\right] = \frac{p^2\pi\left[1-t\eta(1-\gamma q)\right]}{P(A)}. \tag{44}$$

5.3 Covariance Calculations

5.3.1 Sire to Male Calf

First note that

$$\begin{aligned}\mathrm{E}\left[TH_S 1_{\{N_S=1\}} H_B 1_{\{U_S+U_D=1\}}\middle|A\right] &= \frac{\eta P(A, T=1, N_S=1, U_S+U_D=1, H_S=1)}{P(A)}\\ &= \frac{\eta\gamma}{P(A)}P(T=1, N_S=1, U_S+U_D=1, H_S=1)\\ &= \frac{t\eta^2\gamma}{P(A)}P(U_S+U_D=1|N_S=1)\,2pq\\ &= \frac{2pqt\eta^2\gamma}{P(A)}\{qP(U_S=1|N_S=1)+pP(U_S=0|N_S=1)\}\\ &= \frac{pqt\eta^2\gamma}{P(A)}.\end{aligned} \tag{45}$$

Also

$$\begin{aligned}\mathrm{E}\left[TH_S 1_{\{N_S=1\}} H_B 1_{\{U_S+U_D=2\}}\middle|A\right] &= \frac{\eta pP(A, T=1, N_S=1, U_S=1, H_S=1)}{P(A)}\\ &= \frac{\eta p\gamma}{P(A)}P(T=1, N_S=1, U_S=1, H_S=1)\\ &= \frac{t\eta^2 p\gamma}{P(A)}P(N_S=1, U_S=1)\\ &= \frac{p^2qt\eta^2\gamma}{P(A)}.\end{aligned} \tag{46}$$

With reference to (16) and (18),

$$\begin{aligned}\mathrm{Cov}(Y_S, Y_B|A) = {}& \Delta_1^2\mathrm{Cov}\left(TH_S 1_{\{N_S=1\}}, TH_B 1_{\{U_S+U_D=1\}}\middle|A\right) +\\ & \Delta_1\Delta_2\mathrm{Cov}\left(TH_S 1_{\{N_S=1\}}, TH_B 1_{\{U_S+U_D=2\}}\middle|A\right).\end{aligned} \tag{47}$$

Using (45), (36), and (41), the first covariance is

$$\mathrm{E}\left[TH_S 1_{\{N_S=1\}} H_B 1_{\{U_S+U_D=1\}}\middle|A\right] - \mathrm{E}\left[TH_S 1_{\{N_S=1\}}\middle|A\right]\mathrm{E}\left[TH_B 1_{\{U_S+U_D=1\}}\middle|A\right]$$

$$= \frac{pqt\eta^2\gamma}{P(A)} - \left(\frac{2pqt\eta\gamma}{P(A)}\right)\frac{\eta tpq}{P(A)}\left[2-\eta(2-\gamma-q)\right]$$

$$= \frac{t\eta^2 pq\gamma\left[1-t+t(1-2p)\ (1-2p+\eta p^2)\right]}{P(A)^2}, \tag{48}$$

after simplification. Using (46), (36), and (42), the second covariance is

$$\mathrm{E}\left[TH_S 1_{\{N_S=1\}} H_B 1_{\{U_S+U_D=2\}} \middle| A\right] - \mathrm{E}\left[TH_S 1_{\{N_S=1\}} \middle| A\right] \mathrm{E}\left[TH_B 1_{\{U_S+U_D=2\}} \middle| A\right]$$

$$= \frac{p^2 qt\eta^2\gamma}{P(A)} - \left(\frac{2pqt\eta\gamma}{P(A)}\right)\frac{\eta tp^2}{P(A)}\left[1-\eta+\eta\gamma q\right]$$

$$= \frac{t\eta^2 p^2 q\gamma}{P(A)^2}\left[1-t+t(1-2p+\eta p^2)\right]. \tag{49}$$

Combining (47), (48), and (49), $\mathrm{Cov}(Y_S, Y_B|A)$ is given by

$$\mathrm{Cov}(Y_S, Y_B|A) =$$

$$\frac{\Delta_1^2 t\eta^2 pq\gamma}{P(A)^2}\left[t(1-2p+\eta p^2)\left(1-2p+\frac{\Delta_2}{\Delta_1}p\right)+(1-t)\left(1+\frac{\Delta_2}{\Delta_1}p\right)\right]. \tag{50}$$

5.3.2 Sire to Female Calf

From here on, only an outline of the calculations will be given.

$$\mathrm{E}\left[TH_S 1_{\{N_S=1\}} H_C 1_{\{V_S+V_D=1\}} \middle| A\right] = \frac{pqt\eta\pi\gamma}{P(A)};$$

$$\mathrm{E}\left[TH_S 1_{\{N_S=1\}} H_C 1_{\{V_S+V_D=2\}} \middle| A\right] = \frac{p^2 qt\eta\pi\gamma}{P(A)};$$

$$\mathrm{Cov}(Y_S, Y_C|A) = \Delta_1\Lambda_1\mathrm{Cov}\left(TH_S 1_{\{N_S=1\}}, H_C 1_{\{V_S+V_D=1\}} \middle| A\right) + ;$$

$$\Delta_1\Lambda_2\mathrm{Cov}\left(TH_S 1_{\{N_S=1\}}, H_C V_S V_D \middle| A\right). \tag{51}$$

The first covariance in (51) is

$$\frac{pqt\eta\pi\gamma}{P(A)^2}(1-2p)\ (1-2p+t\eta p^2).$$

The second covariance is

$$\frac{p^2 qt\eta\pi\gamma}{P(A)^2}\left\{1-2p+t\eta p^2\right\}.$$

In conclusion

$$\mathrm{Cov}(Y_S, Y_C|A) = \frac{\Delta_1\Lambda_1 pqt\eta\pi\gamma}{P(A)^2}(1-2p+t\eta p^2)\left(1-2p+\frac{\Lambda_2}{\Lambda_1}p\right).$$

5.3.3 Dam to Male Calf

$$\mathrm{Cov}\,(Y_D, Y_B|A) \;=\; \sum_{i=1}^{2}\sum_{j=1}^{2} \Lambda_i \Delta_j S_{ij}\ ,$$

where

$$S_{ij} \equiv \mathrm{Cov}\,(H_D 1_{\{N_D = i\}}, TH_B 1_{\{U_S + U_D = j\}}|A)\ . \tag{52}$$

To compute these quantities, we use:

$$\mathrm{E}\,[H_D T H_B 1_{\{N_D=1\}} 1_{\{U_S+U_D=1\}}|A] \;=\; \pi t \eta p q \left(\frac{1-\eta+2\gamma\eta p q+\eta q^2}{P(A)}\right);$$

$$\mathrm{E}\,[H_D T H_B 1_{\{N_D=1\}} 1_{\{U_S+U_D=2\}}|A] \;=\; \frac{\pi t \eta p^2 q\,(1-\eta+\gamma\eta q)}{P(A)};$$

$$\mathrm{E}\,[H_D T H_B 1_{\{N_D=2\}} 1_{\{U_S+U_D=1\}}|A] \;=\; \pi t \eta p^2 q \left(\frac{1-\eta+\gamma\eta p+\eta q}{P(A)}\right);$$

$$\mathrm{E}\,[H_D T H_B 1_{\{N_D=2\}} 1_{\{U_S+U_D=2\}}|A] \;=\; \pi t \eta p^3 \left(\frac{1-\eta+\gamma\eta q}{P(A)}\right).$$

From these and the formulas in 5.2.2 and 5.2.3,

$$S_{11} \;=\; \frac{\pi t \eta p q\,(1-2p)}{P(A)}\,\{1-2p+\eta p^2\}\ ;$$

$$S_{12} \;=\; \frac{\pi t \eta p^2 q\,(1-2p)}{P(A)}\,\{1+\eta\,[q\gamma-1]\}\ ;$$

$$S_{21} \;=\; \frac{\pi t \eta p^2 q}{P(A)}\,\{1-2p+\eta p^2\}\ ;$$

$$S_{22} \;=\; \frac{\pi t \eta p^3 q}{P(A)}\,\{1+\eta\,[q\gamma-1]\}\ .$$

Thus $\mathrm{Cov}\,(Y_D, Y_B|A)$ is given by:

$$\frac{\pi t \eta p q}{P(A)}\,\{\Lambda_1(1-2p)+\Lambda_2 p\}\,\{\Delta_1[1-2p+\eta p^2]+\Delta_2 p\,[1+\eta\,(q\gamma-1)]\}\ . \tag{53}$$

5.3.4 Dam to Female Calf

$$\mathrm{Cov}\,(Y_D, Y_C|A) \;=\; \sum_{i=1}^{2}\sum_{j=1}^{2} \Lambda_i \Lambda_j S_{ij}'\ ,$$

where

$$S_{ij}' \equiv \mathrm{Cov}\,(1_{\{N_D=i\}}, 1_{\{V_S+V_D=j\}}|A)\ .$$

To compute these quantities, we use:

Genetic models with reduced penetrance related to the Y chromosome

$$\mathrm{E}\left[H_D 1_{\{N_D=1\}} H_C 1_{\{V_S+V_D=1\}} \middle| A\right] = \pi^2 pq;$$

$$\mathrm{E}\left[H_D 1_{\{N_D=1\}} H_C 1_{\{V_S+V_D=2\}} \middle| A\right] = \frac{\pi^2 p^2 q\,[1+t\eta\,(\gamma q-1)\,]}{P(A)};$$

$$\mathrm{E}\left[H_D 1_{\{N_D=2\}} H_C 1_{\{V_S+V_D=1\}} \middle| A\right] = \frac{\pi^2 p^2 q\,\{1+t\eta p\,(\gamma-1)\,\}}{P(A)};$$

$$\mathrm{E}\left[H_D 1_{\{N_D=2\}} H_C 1_{\{V_S+V_D=2\}} \middle| A\right] = \frac{\pi^2 p^3\,[1+t\eta\,(\gamma q-1)\,]}{P(A)}.$$

Hence

$$S_{11}' = \frac{\pi^2 pq\,(1-2p)}{P(A)}\,\{1-2p+t\eta p^2\};$$

$$S_{12}' = \frac{\pi^2 p^2 q\,(1-2p)}{P(A)}\,\{1+t\eta\,(\gamma q-1)\,\};$$

$$S_{21}' = \frac{\pi^2 p^2 q}{P(A)}\,\{1-2p+t\eta p^2\};$$

$$S_{22}' = \frac{\pi^2 p^3 q}{P(A)}\,\{1+t\eta\,(\gamma q-1)\,\}.$$

Thus $\mathrm{Cov}\,(Y_D, Y_C|A)$ is given by:

$$\frac{\pi^2 pq}{P(A)}\,(\Lambda_1(1-2p)+\Lambda_2 p)\Big(\Lambda_1\{1-2p+t\eta p^2\}+\Lambda_2 p\,\{1+t\eta\,(\gamma q-1)\,\}\Big).$$

6 Conclusions

The model presented here, in which reduced penetrance of an autosomal gene is partially dependent on abnormality in the *Y* chromosome, seems to be successful as a first attempt to explain the covariances between pulmonary artery pressures of cattle and those of their parents. Examination of further datasets, and deeper genealogical studies, will help to support or invalidate this model. If our model is correct, and if the purported abnormal *Y* chromosome could be identified, then it would possible to develop a breed of cattle in which males would no longer be susceptible to Brisket Disease, although about half of females would still be susceptible.

Acknowledgments: The authors thank an anonymous Colorado rancher for allowing data on his cattle to be used for this study. German Ramirez brought us together and provided valuable resources and encouragement. This work was possible thanks to many hours of data preparation by Polly Bittle. The authors thank geneticist Wilmer J. Miller for much helpful advice. Anonymous referees provided valuable references to the literature.

7 References

[1] A. F. Alexander and D. H. Will (1963): The stress of high altitude environment upon cattle. Technical Report, Colorado State University, Fort Collins, Colorado.

[2] W. J. Ewens (1979): *Mathematical Population Genetics*. Springer-Verlag, New York.

[3] M. Eliasziw and D. Donner (1990): Comparison of recent estimators of interclass correlation from familial data. Biometrics 46, 391-398.

[4] O. Kempthorne (1957): An *Introduction to Genetic Statistics*. Wiley, New York.

[5] J. B. S. Haldane (1949): Paternal and fraternal correlations for fitness. Ann. Eugen. 14, 288-292.

[6] L. Hogben (1932): Filial and fraternal correlations in sex-linked inheritance. Proc. R. Soc. Edinburgh 52, 331-336.

[7] W. F. Hollander and W. J. Miller (1978): Amputated, frayed and sideburns: three new mutants in the pigeon. Iowa state Journal of Research, 53, No. 2, 109-118.

[8] Tim Holt, R. W. R. Darling, P. Bittle, W. J. Miller, and G. Ramirez (1997): Genetic adaptation model for susceptibility of cattle to the development of brisket disease at high altitudes. University of South Florida College of Medicine Report.

[9] Tim Holt (1997): Pulmonary arterial hypertension resulting in congestive heart failure. Preprint, Town and Country Animal Hospital, 1525 Hwy 135, Gunnison Co 81230, USA.

[10] C. R. Leach and O. Mayo (1967): On the effects of natural selection on correlations between relatives. Ann. Hum. Genet. 30, 379-385.

[11] O. Mayo, R. Burger, and C. R. Leach (1990): The heritability of fitness: some single gene models. Theor. Appl. Genet. 79, 278-284.

[12] J. G. Schimmel (1981): Genetic aspects of high mountain disease in beef cattle. Ph.D. Dissertation, Colorado State University.

[13] G. P. Vogler, R. Wette, M. K. McGue, D. C. Rao (1995): Properties of alternative estimators of familial correlations under variable sibship size. Biometrics 51, 276-283.

[14] E. K. Weir, A. Tucker, J. T. Reeves, D. H. Will, R. F. Grover (1974): The genetic factor influencing pulmonary hypertension in cattle at high altitude. Cardiovasc. Res. 8, 745.