

The Effect of Sample Attrition on the Frequency Distribution of Blood Pressure and Genetic Marker Phenotypes Representing a Natural Unselected Community: Tecumseh, Michigan

DAVID G. SMITH¹ AND CHARLES F. SING²

Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48104

KEY WORDS Nonrandom sample · Blood pressure · Sample attrition · Phenotype by subsample interaction.

ABSTRACT Highly significant differences in mean age, blood pressure and phenotype frequency distributions between the non-migrants and "emigrants" of a total unselected community sample were discovered. Use of the mean of BP scores collected from epidemiologic surveys over a period of time as an individual score allows sample attrition to produce both a genetically and demographically biased sample of a population intended to represent an unselected community of people.

Multiple regression analyses estimated the contribution of an individual's age, genotype and mobility out of the sample to predicting blood pressure variation. Variation in blood pressure means among certain marker phenotype classes was greater in those who leave than in those who stay, but only the upper portion of the pressure distribution contributed to this relationship. A genetic-environment interaction is suggested.

The greater emphasis more recently given to the environmental role in the controversy over the genetic contribution to blood pressure (BP) variation has, in large part, resulted from difficulties in the proper collection and statistical treatment of data for valid genetic analyses (Feldman et al., '73). The traditional strategy to determine the role of genes has been based on correlations of BP values between first degree relatives and the comparison of the properties of frequency distributions (adjusted for non-genetic variables) collected from families with the properties of the distribution representing the population at large. Such studies may have failed to indicate a role of genetic factors because they involved small samples or an inappropriate sampling design so that the statistical power to estimate the relative roles of genes and environments has been lacking. While the family set procedure (Schull et al., '70) may offer an improved strategy, the properties of estimates using the procedure have yet to be established in a broad range of sampling situations.

Only with very large samples can single-occasion BP measurements produce valid, representative BP distributions of a population (McKeown et al., '63; Lowe and McKeown, '62). Even with large samples, the use of single measurements results in an overestimate of the true between-subject standard deviations by as much as one-third, resulting in the underestimation of familial resemblances (Armitage and Rose, '66; Armitage et al., '66). The mean of readings acquired on different occasions has been found to yield a considerably greater reduction in the between-subject standard deviation, and is therefore considered the better estimate of an individual's BP, than if the mean is based on multiple readings taken on the same occasion (Armitage and Rose, '66).

The use of the mean of individual scores collected by a multiple occasion epidemiologic survey, however, would seem advis-

¹ Postdoctoral Fellow, Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48104.

² Associate Professor, Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48104.

able only if (1) measurement techniques do not vary among examination periods and (2) those individuals available for the required number of consecutive samplings are representative of the larger populace about which inferences are to be drawn. The turnover in technical personnel typical of most large epidemiologic surveys makes the first criterion difficult to meet. As for the latter, there are reasons to believe the required time lag between examination occasions may allow sample attrition due to mortality and emigration to non-randomly select BP values from the frequency distribution. First, high BP is known to correlate with coronary heart disease and related disorders (Kannel et al., '61; Epstein, '64). Secondly, excessive environmental stress upon highly mobile individuals can lead to permanently elevated BP (Scotch and Geiger, '63). These conditions might assure that one's sample constitutes a selected survivor population whose mean scores for multiple occasion examinations are biased. Furthermore, the higher mortality of higher pressures removes from the sample observations on older individuals and the rate of increase in BP with age is underestimated (Record and Whitfield, '64). Within the survivor population, young people with high BP of genetic origin are, then, more likely to have lost relatives with high BP through mortality than are youths with lower pressures, thus reducing the correlation between BP scores of *propositi* and their first degree relatives (Miall et al., '67). Nonrandom emigration of individuals with higher BP who are related to members who stay or emigration of entire nuclear family units containing members with higher BP could also reduce familial correlations and hence lead to an underestimate of the true role of genes. The present analysis measures the biases in a large epidemiologic study and extracts information about the role of genes and environment in a way which exploits these difficulties of analysis.

To estimate the effects of emigration and mortality on the study of BP we ask three fundamental questions: (1) how does the observed BP distribution available for a total community population change as mortality and emigration reduce the sample, (2) are those who leave the sample a

random sample of ages and genotypes and (3) can a typical multiple-occasion epidemiologic survey give an unbiased sample of individual mean BP scores as an alternative to the use of single occasion measurements? The failure of BP differences among genetic marker phenotypes to be homogeneous between those who are lost from the cohort and those who remain will be utilized to test the null hypothesis that an individual's genotype is independent of his BP level.

POPULATION SAMPLES AND THE METHODS OF ANALYSIS

The analyses reported here are based on age, BP and genetic marker phenotypes for polymorphic blood and serum loci measured for members of a 90% sample of a population of nearly ten thousand individuals residing in Tecumseh, Michigan between the years 1959 and 1960. Beginning in 1960, data were collected by the Tecumseh Community Health Study during "rounds" of examinations made at roughly four year intervals. Those individuals examined in round 1 and their sex, age, and BP define the cohort and data analyzed in this study. The study design called for re-examination in a second "round" of everyone identified in round 1. Of these less than 10% refused reexamination. Unless a doctor inadvertently failed to take a reading all individuals reporting for the round 2 examination had at least a systolic BP measurement taken. The two measures of BP, then, define separate populations of inference and require separate analyses. For the analysis of genetic markers, the individual defined by this sampling procedure must have been typed for at least the particular locus under consideration. Analyses of data on all individuals ever typed are given by Shreffler et al., '71; Sing et al., '71 and Sinnock and Sing, '72a,b. Those typed include about 98% of those reexamined in the second round and 17% of those not reexamined. By using a large number of genetic markers one can observe variation for a large portion of chromosomal material and identify regions linked to the marker lost at higher rates from the sample. Either a correlation of such a frequency change with mean BP change for the marker phenotypes or a

change in the mean BP of a marker phenotype without a change in its relative frequency suggests the *possible* role of genetic factors in the region marked.

The sample of data collected during the first round was subdivided into five subsamples distinguished by whether the individual was examined (BP measured) in the subsequent rounds. The subsamples are designated and defined as follows:

S_1 = all individuals measured for age and BP at round 1.

S_2 = all individuals measured at round 1 and examined in round 2.

S_3 = S_1 less S_2 .

S_4 = S_3 whose round 1 BP score fell below the S_1 mean.

S_5 = S_3 whose round 1 BP score fell above the S_1 mean.

Subsample S_3 and its subsamples, S_4 and S_5 , represent attrition from the round 1 sample of measurements. Because all subsamples represent round 1 data, the change in an individual's BP over time cannot affect the contrasts among subsamples.

Each subsample of data identified by sex, subsample contrast, and BP measure (diastolic or systolic) will hereafter be termed a case (e.g., S_2 - S_3 male diastolic). When a case is further identified by a specific marker locus, the sample will be termed a marker case (e.g., ABO S_2 - S_3 female diastolic). Estimates of parameters which define the distributions of age and BP for each of the five subsamples were first calculated (mean, standard deviation, skew and kurtosis for both systolic and diastolic BP and mean and standard deviation for age) and students *t* test (Snedecor and Cochran, '71: p. 104) was employed to detect statistically significant mean differences between the BP and age of the two subsamples represented by each case. The significance of skew (g_1) and kurtosis (g_2) was also tested using the *t* statistic (Sokal and Rohlf, '69: p. 137).

To determine whether those who leave the sample are genetically atypical of the original cohort, the hypothesis of homogeneity of frequencies of the marker phenotypes for each locus between subsamples was tested using the contingency chi-square test. Of the 14 markers seven (Hp, Gc, MN, Ss, Kell, Rh-C and Rh-E) are codominant and seven (ABO, P, Duffy, ABH

secretor, Lewis secretor, Rh-D and Kidd) are dominant systems. Since one Kell phenotype is rare only the two most frequent phenotypes were considered.

The effects of age, sample subdivision, and single locus marker phenotype on BP were next considered. An analysis of the BP sum of the squares by multiple linear regression was conducted to estimate the interaction between the subsample and marker phenotype effects in predicting BP. The goodness-of-fit to the data of the linear regression model based on a complete set of indicators was compared to the model based on a reduced set (excluding this interaction term). In this regression model the two subsamples were assigned numerical values of 1 or 2. The marker phenotypes for each codominant system were assigned 1, 2, and 3 for AA, Aa, and aa, respectively, while values of 1 and 2 were assigned to the two phenotypes of each dominant system. An additional interaction term, age \times subsample, was also included as an independent variable in the regressions, since, if significant, the difference in BP between subsamples could be biased by the non-random distribution of ages between the subsamples of the marker case. While higher order polynomials for age regression were significant, linear regressions removed over ninety percent of the BP sum of squares removed by cubic regressions. The complete regression model may be represented as

$$Y_{ijk} = a + \alpha A_i + \beta S_j + \Delta AS_{ij} + \delta P_k + LSP_{jk} + E_{ijk}$$

and the reduced model as

$$Y_{ijk} = a + \alpha A_i + \beta S_j + \Delta AS_{ij} + \delta P_k + E_{ijk}$$

where *a* is the Y-intercept and α , β , Δ , δ , and *L* are the linear regression coefficients for BP on age, on subsample given age, on the product of the age and subsample after fitting age and subsample, on marker phenotype after fitting the previous terms and on marker phenotype by subsample after fitting all other terms. *E* is the error term for the complete and reduced models. The variables *A*, *S*, *AS*, *P* and *SP* represent the numerical assignments, respectively for age, subsample, age \times subsample, marker phenotype and subsample \times marker pheno-

type which characterize an individual. A variance ratio tests the statistical significance of the marker phenotype by subsample interaction effect. It is a function of the difference in sum of squares removed from the data by the complete (SSR_c) and the reduced (SSR_r) models (Bancroft, '68). The variance ratio is

$$F = \frac{SSR_c - SSR_r}{MSE_c}$$

where

MSE_c is the residual mean square error for the complete model.

The chi-square test for heterogeneity of phenotype frequencies between subsamples and the regression analysis of sum of squares of BP provide a basis for evaluation of causation of the mean difference in BP between those individuals who stay and leave after round 1 for each of the genetic markers. When marker phenotype frequencies were homogeneous between subsamples and mean change is independent of marker phenotype the marker case was denoted as a type I outcome. Involvement of a marker phenotype effect may be of two sorts. First, the relative frequencies of the marker phenotypes fail to change significantly, but the difference among BP means of the marker phenotypes of the subsample which leaves are significantly different from the differences among the same marker phenotypes for those who stay (type II). Assuming a difference in the spectrum of environments for individuals who leave the sample and those who stay, this may be interpreted as a marker phenotype by environment interaction effect on BP. Secondly, the relative frequencies of the marker phenotypes may change and a significant interaction of subclass and phenotype class may or may not occur. If BP means do not change significantly among marker phenotype classes, either the effect of attrition is on the phenotype frequencies but not on BP or a sampling error of the first kind has occurred in the test of frequencies (type III). If there is a subsample by marker phenotype interaction and a significant change in mean BP of a marker phenotype is associated with a frequency change, either the locus identifies a chromosome region which can be used to predict BP scores (type IV) or sampling errors of the first kind are responsible for the outcome.

RESULTS

Males (but not females) who left S_1 were significantly older than those who stayed and both males and females of S_3 had statistically significant higher systolic (but not diastolic) BP than S_2 (table 1). Age alone may be responsible for the significant BP difference in males since BP is highly correlated with age in the Tecumseh data (Johnson et al., '64). The skew (g_1) and kurtosis (g_2) of all five subsamples differ significantly from zero. The loss of S_3 , with higher variance, skew and kurtosis (except for the female diastolic case) reduces these statistical measures of the S_2 subsample by removing a large portion of the upper tail and mid-section of the S_1 BP distribution.

Table 2 lists those marker cases which gave statistically significant heterogeneity of phenotype arrays between subsamples for a 5% test. Since a fraction of those measured for systolic BP were also measured for diastolic BP, the subsamples identified by the two measures define two different but non-independent samples of the population of inference. Heterogeneity between subsamples for a marker was accepted as significant only when both measures (systolic and diastolic) gave a statistically significant test (0.05 level of probability). A significant chi-square for only one of the two marker cases was regarded as a sampling error of the first kind. Of the eight marker cases giving statistically significant chi-square values, six were thusly eliminated as probably chance rejections of homogeneity and assigned to type I. After accounting for such errors, only the male Duffy S_2 - S_3 contrast gave statistically significant heterogeneity of frequencies between subsamples, representing an excess of $Fy(a-)$ in S_3 males.

For those marker cases where $N < 15$ for any single phenotype of a marker system for a subsample the null hypothesis of no marker system by subsample interaction effect on BP level and of homogeneity of the frequency of marker phenotypes were accepted and the marker case was assigned to the type I outcome. This seemed reasonable since a number of those leaving could belong to a single family unit whose members are not genetically independent. Furthermore, for marker cases where $N < 30$ for any single phenotype of a marker system for a subsample, the null hypothesis

TABLE 1

Statistical definitions of the blood pressure and age ¹ for each of the five subsamples (defined in text)

Sub-samples	Male systolic							Female systolic						
	Blood pressure				Age			Blood pressure				Age		
	N	μ	σ	g ₁ ³	g ₂ ³	μ	σ	N	μ	σ	g ₁ ³	g ₂ ³	μ	σ
1	3954	127.2	21.5	0.80 ²	1.7 ²	27.2	20.3	4082	126.9	23.9	1.45 ²	3.9 ²	28.0	20.3
2	2955	126.0	20.6	0.63 ²	1.3 ²	26.3	19.4	3135	126.2	23.1	1.41 ²	3.7 ²	27.9	19.8
3	999	130.7	23.7	1.04 ²	2.7 ²	30.1	22.4	947	129.1	26.3	1.49 ²	4.0 ²	28.6	21.8
4	487	112.7	10.2	-0.94 ²	0.6 ²	18.7	17.7	531	112.0	10.3	-1.01 ²	1.3 ²	17.6	13.2
5	512	147.9	19.9	1.71 ²	3.4 ²	41.0	20.9	416	151.0	24.3	1.89 ²	5.2 ²	42.8	22.5

Sub-samples	Male diastolic							Female diastolic						
	Blood pressure				Age			Blood pressure				Age		
	N	μ	σ	g ₁ ³	g ₂ ³	μ	σ	N	μ	σ	g ₁ ³	g ₂ ³	μ	σ
1	3525	77.1	13.1	0.30 ²	1.6 ²	27.6	20.1	3518	76.7	13.1	0.70 ²	2.7 ²	28.2	20.1
2	2574	77.0	12.7	0.24 ²	1.4 ²	26.7	19.1	2654	76.6	12.0	0.70 ²	2.7 ²	28.1	19.5
3	951	77.3	14.2	0.44 ²	1.8 ²	29.9	22.0	864	77.1	13.7	0.65 ²	2.5 ²	28.4	21.6
4	480	66.4	8.3	-1.65 ²	4.5 ²	20.1	18.8	434	66.3	7.7	-1.38 ²	3.4 ²	19.3	16.7
5	471	88.4	10.8	1.63 ²	3.0 ²	39.9	21.2	430	88.0	10.8	1.97 ²	5.9 ²	37.6	22.1

Absolute differences in mean scores and significance of differences ²

Contrasts	Male systolic		Female systolic		Male diastolic		Female diastolic	
	Pressure	Age	Pressure	Age	Pressure	Age	Pressure	Age
2 vs 3	4.7 ²	3.8 ²	2.9 ²	0.7 N.S.	0.3 N.S.	3.2 ²	0.5 N.S.	0.3 N.S.
4 vs 5	—	22.3 ²	—	25.2 ²	—	19.8 ²	—	18.3 ²

¹ Age in years.

² Statistically significant at the 0.001 level of probability.

$$g_1 = \frac{m_3}{(\sqrt{m_2})^3} \quad g_2 = \frac{m_4}{m_2^2} \quad -3 \text{ where } m_r = \frac{\sum(X - \bar{X})^r}{N}$$

TABLE 2

Sample sizes (N) and Chi-square for eight cases yielding significant heterogeneity in the distribution of marker phenotypes ¹

Marker cases	Contrasts					
	S ₂ vs S ₃			S ₄ vs S ₅		
	S ₂ N values	S ₃	χ ²	S ₄ N values	S ₅	χ ²
ABO MS				61	142	8.0 ²
MN FS	3056	153	6.1 ²			
Ss FD	2590	86	6.3 ²			
Rh-C FS						
Rh-C FD						
Kell FS	3048	154	5.1 ²			
Duffy MS	2876	203	5.6 ²	61	142	4.1 ²
Duffy MD	2512	219	5.7 ²			
Kidd b FS						
Kidd b FD						
P FS						
Lewis MS						
Lewis MD						
Hp MS	2707	64	10.8 ³			
Hp MD						
Gc FS						

¹ M and F refer to Male and Female and S and D indicate the measure, i.e., systolic or diastolic.

² Statistically significant at the 0.05 level of probability.

³ Statistically significant at the 0.01 level of probability.

TABLE 3

Outcomes of the chi-square test of heterogeneity of phenotype frequencies and regression analysis of the subsample by marker phenotype interaction effect on BP

Measure	Type I (neither test significant)	Type II (significant interaction effect on BP)			Type III (significant heterogeneity of phenotype frequencies)		
		Marker system	Sex ¹	Contrast	Marker system	Sex ¹	Contrast
Systolic	all others	Kell	F	2-3	Duffy	M	2-3
Diastolic	all others	MN	F	2-3	Duffy	M	2-3
		Kell	F	2-3			
		Duffy	F	2-3			

¹ M. Male; F. Female.

TABLE 4

Estimated blood pressure means corrected for age and age × subsample effects for phenotypes involved in a significant marker system by subclass interaction effect

Cases ¹		Markers						
		Kell		Duffy		MN		
		kk	Kk	A +	A -	MM	MN	NN
FS	S ₂	127.7	125.0					
FS	S ₃	128.0	114.5					
	FS S ₄	132.0	132.8					
	FS S ₅	151.1	131.5					
FD	S ₂	77.2	76.0	77.1	77.2	76.7	77.2	77.7
FD	S ₃	76.9	69.4	74.7	78.8	78.2	75.9	73.6
	FD S ₄	74.6	75.4	75.3	73.8	74.8	74.8	74.8
	FD S ₅	88.1	80.1	85.7	90.2	89.5	87.3	85.0

¹ M and F refer to male and female and S and D indicate the measure; systolic and diastolic, respectively.

of no interaction as measured by the regression analysis was rejected only when the probability of observing a greater F value was less than 0.025. Otherwise, statistical significance at the 0.05 level of probability was accepted as evidence for an interaction effect between subsample and marker system on BP. Table 3 arranges the marker cases according to the types of outcome for the two tests of significance. No marker case was of type IV. Table 4 gives the phenotype by subsample BP means adjusted for the effects of age and age by subsample interaction for all type II outcomes. For females both the systolic and diastolic S₂-S₃ contrasts involved the Kell marker system. While the difference in mean BP between the two phenotypes of S₂ was in the same direction as that in S₃ the latter difference was much

greater. The S₄-S₅ contrast revealed that the greater differences between phenotypes in those who leave was reflected only in the upper (S₅) portion of the S₃ distribution. Both of the phenotypes in S₄ yielded almost identical mean BP scores whereas there was a 20 mm and 8 mm difference in S₅ for the systolic and diastolic measures respectively. Two additional marker cases, the diastolic S₂-S₃ contrasts for Duffy and MN, were assigned to type II. Again, the pressures of S₂ vary in the same direction as those of S₃ for the Duffy marker case; for the MN marker case pressures of S₂ and S₃ vary in opposite directions. The pressures vary much greater among marker phenotypes in S₃ than in S₂ and again the S₅ subsample identifies these differences. Although no male contrast was assigned to type II, the male systolic and diastolic

S_2 - S_3 Duffy contrasts represent the only type III cases found.

DISCUSSION

Three major findings comprise the thrust of this report:

(1) Blood pressures and ages of the S_1 sample are not randomly distributed between S_2 and S_3 .

(2) For females, the difference in mean BP between those leaving and staying is not homogeneous among phenotypes of certain marker systems. For these marker cases, a genotype by environment interaction is indicated.

(3) Genetic environmental interaction effects are non-linear with respect to BP scale, being greater at higher levels of BP.

There are at least three explanations for BP differences between subsamples which exist in our data: (1) age differences between subsamples, (2) environmental differences between subsamples, and (3) changes in frequency of genotypes which affect BP level. The first two explanations are relevant to our first major finding.

Epidemiologic investigations often require sample designs to measure variables in samples which are constantly changing demographically. It is, then, paramount to determine whether an outcome of such a study is the result of some independent variable under observation or simply an artifact of the demographic history of the cohort reflected by the sample design. If, in the fixed cohort, a few old people with higher BP die, younger members emigrate and the youngest members age but are not replaced by new births with very low BP, the ongoing sample ages both at the base and apex of the age distribution and represents a demographically unstable population. Because of the positive correlation between BP and age, distributions of both change due to loss from the cohort. Some of the differences in the distributional characteristics of BP of S_2 and S_3 may be attributable to such cohort effects since the S_1 age distribution is unstable. Figure 1, the Tecumseh age pyramid circa 1960 constructed from the S_1 sample, reflects the low net reproduction typical of the economic depression-war-era (1930-1945), the pinch in the age structure, common

in the United States population, occurring at ages 15-30 (those born between 1930 and 1945). This depression will ascend the pyramid generating a second depression, their offspring, already beginning to form at the 0-4 age group. In the wake of each depression will follow the type of hump represented by the 5-15 age group which, during round 2, begins to fill the war-time depression, but leaves in its wake a depression at the 5-15 age group.

Since the average age of S_3 males is about 30 and 10-15% of this subsample is absent because of mortality, the average age of death for the United States in 1960 being about 65 (Keyfitz and Fleiger, '71), the average male emigrant after round 1 fell well within this least represented age group, probably between the ages 15 and 25. If vital rates do not change, the changes in numbers of emigrants will be determined by the size of each age cohort as it enters the age at which risk at emigration is highest. Those at risk were few in number in 1960. The fairly low BP and age of emigrants could not balance the higher age and BP of those in S_3 who died causing the S_2 mean BP and age to fall slightly below the corresponding S_1 mean. That S_3 mean ages for both systolic and diastolic subsamples of males, but not females, are significantly higher than those of S_2 could result from the fact that age-specific mortality rates are typically higher for males in western populations than for females. It is unlikely that females emigrate in greater numbers, since the pinch in the age pyramid is less pronounced for females, or at earlier ages than males. Since the Tecumseh BP distribution becomes progressively skewed with age (Johnson et al., '65) and those who die have, on the average, higher pressures, which are distributed leptokurtically, removal of these older individuals is probably responsible for the lower variance and skew of S_2 . The lower kurtosis of S_2 could result from a tendency for emigration to occur within a restricted age group, one in which most pressures have not yet begun to rise. The final sample (S_2) examined at both occasions is, then, on the average, about 25 percent smaller, a year younger, more homogeneous in age, and has lower pressures which vary less about the mean than the original sam-

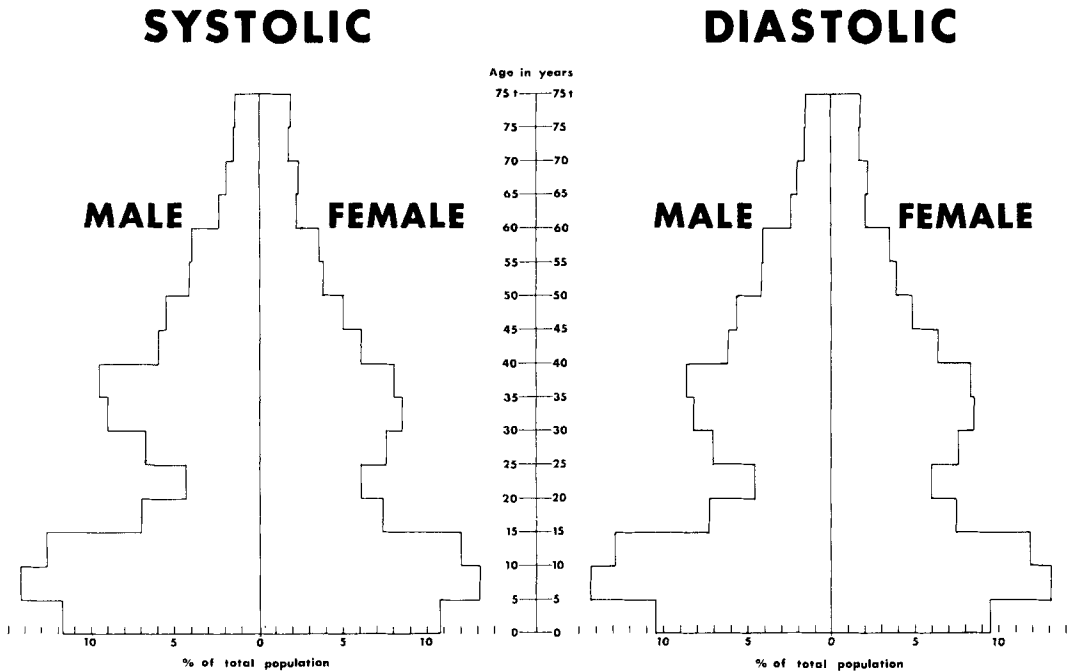


Fig. 1 Tecumseh S_1 age structure.

ple enumerated at S_1 . After approximately four years, the population of inference which S_2 is intended to represent is no longer the community from which the S_1 sample was acquired.

Those leaving after round 2 should, on the average, be younger and have lower pressures than S_3 . Emigrants should comprise a greater portion of the total subsample which leaves since between 1965 and 1969 those between the ages of 15 and 25 will proportionately comprise a greater portion of the total population than between 1960 and 1964. The result should be an increase in BP and age of non-migrants after the end of the third round of examination. Unfortunately, the manner in which data were collected during the third round of examinations prevented such an analysis. Irrespective of the length of time lag, but in proportion to it, any epidemiologic study will suffer attrition. Even if attrition can be replaced with new births, the sample will change as a result of the absence of demographic stability, a condition which characterizes most western populations.

Differences between the distributional characteristics of emigrants and non-mi-

grants might also result from environmental differences. Migrants are, it is often hypothesized, exposed to excessive amounts of environmental stress induced by maladaptation to new environmental situations (Stamler et al., '67; Scotch and Geiger, '63). Case studies have found cultural, social, economic and geographic mobility to be positively associated with elevated BP (Lowenstein, '61; Scotch, '63; Shaper, '67; Cruz-Coke, '60; Harburg et al., '70). Acute emotional stress induced experimentally by forced mental arithmetics has been found to raise temporarily systolic BP of otherwise healthy patients 25 mm Hg (Brod, '70; Levi, '70). Lability of BP induced by repetitive stress episodes resulting in repeated contraction of peripheral arterioles might cause hypertrophy and, eventually, permanent narrowing of vessel walls, a consequent sustained rise in peripheral resistance and, hence, in arterial pressure (Koster, '70; Weiner, '70). The inability to adjust peripheral resistance of constricted arterioles to maintain a normal BP under stressful environmental stimuli is usually the first observable physiological defect in labile hypertensives (Hoobler, '61). This labile stage usually precedes the onset of

permanent levels of high BP when it occurs (Pickering, '61).

Emigrants within the S_3 subsample may have had difficulty adapting culturally, socially, or economically before moving resulting in their experiencing greater emotional stress. If this inability to adapt is partially responsible for emigration, raised arterial pressure in genetically predisposed individuals and emigration could represent two effects of a common cause. Likewise, a lengthy illness preceding death could generate stress and raise BP of the remaining portion of S_3 .

Our second principal finding, that the BP difference between those who leave and those who stay is not random with respect to certain marker phenotypes, is measured by the interaction between subsample and marker phenotypes for the Kell, Duffy and MN systems for females. This effect detects greater differences between the BP means of marker phenotype classes for those who leave than for those who stay. Rank ordered by mean BP, the marker phenotype classes of S_2 and S_3 subsamples of most type II marker cases display patterns which differ in magnitude but not in direction. If BP of individuals of certain marker phenotypes is preferentially elevated under the highly stressful conditions of mobility or morbidity (which eventually leads to death) the lower stress to which living non-migrants are ordinarily exposed would be expected to elevate BP of those same phenotypes less dramatically. When the genetic contribution is more fully expressed under stressful environments, the use of samples which have suffered attrition due to stress will lead to an underestimate of familial resemblances. This could account, in part, for the failure of many genetic studies to implicate a genetic role in determining BP variation.

Our third principal finding concerns the non-linear nature of the interaction between subsample and marker phenotype for type II marker cases. The differences in mean BP among marker phenotypes of the S_3 subsamples were consistently greater in those with high BP than in those with low BP. Table 4 shows that when S_3 is subdivided into those falling below and above the mean of S_1 , the S_3 rank order of phenotypes by mean BP is reflected only in the upper portion of the BP distribution (S_5).

All differences between marker phenotypes in S_4 are small or negligible.

It seems likely that the effect of sample attrition upon BP is not linear with respect to the length, the intensity, or frequency of exposure to stimuli interpreted as stressful, nor are all phenotypes associated with some genetic markers equally successful in coping hemodynamically with increasing stress other than by allowing increased peripheral resistance, hence higher BP. It is well known that while BP rises rapidly in some individuals upon reaching middle age or mid-range of the BP distribution (140 mm Hg), others experience no further rise at all (Pickering, '61). If the ability to accommodate stress declines with age in a non-linear fashion, as some have claimed (Strehler, '62), the environmental effects upon BP may increase as length or intensity of exposure to stress increases. If genes linked to the markers studied here are involved in interactions with these effects, the greater variation in BP among the marker phenotypes in groups exposed to greater stress would be expected. Genotype differences could represent different thresholds of tolerance of peripheral arterioles to repeated contraction induced by stress above which hypertrophy, hence high BP, occurs.

CONCLUSIONS

The problems involved in acquiring unselected samples from a constantly changing population present the BP investigator with a dilemma. Small samples of individual mean scores can eliminate sample attrition but may be unrepresentative of the population of inference. Sampling schemes designed to represent large unselected populations produce random samples of populations of inference, but use of individual mean scores allows sample attrition to exclude subjects who may be both genetically and demographically atypical. The use of single-occasion scores prevents the elimination of much of the within-subject variation due to measurement error. Since the complete elimination of sample attrition may never become economically feasible, new approaches to the study of the causes of BP variation in the population at large must be sought.

Our approach revealed significant variation in BP (diastolic in most cases) between

phenotypes of females, but not males, who die and emigrate (but not living non-migrants) which was observed as a subsample by marker phenotype interaction at three loci. This result is consistent with a genetic by environment interaction affecting the rate and pattern of rise of BP. This interaction effect proved to be non-linear with respect to the level of BP already attained. It is concluded that a heritable component of BP is its tendency to rise rapidly near the middle region of the pressure distribution when and if exposed to excessive environmental stress.

ACKNOWLEDGMENTS

This study was supported by U. S. Atomic Energy Commission Contract AT(11-1)-1552 and U. S. P. H. S. Program Project Grant H-6378 to the Cardiovascular Research Center, University of Michigan, Ann Arbor, Michigan. The authors are indebted to Dr. L. D. Ostrander, Professor of Internal Medicine and Research Associate in Epidemiology, for his technical advice and information regarding the Tecumseh Community Health Study and to Drs. P. E. Smouse and J. V. Neel of the Department of Human Genetics for their critical comments on earlier drafts of this paper. Portions of this paper were presented at the annual meeting of the American Society of Human Genetics, October, 1974.

LITERATURE CITED

- Armitage, P., W. Fox, G. A. Rose and C. M. Tinker 1966 The variability of measurements of casual blood pressure. II. Survey experience. *Clin. Sci.*, 30: 337-344.
- Armitage, P., and G. A. Rose 1966 The variability of measurements of casual blood pressure. I. Laboratory study. *Clin. Sci.*, 30: 325-335.
- Bancroft, T. A. 1968 *Topics in Intermediate Statistical Methods I*. The Iowa State University Press, Ames.
- Brod, J. 1970 Hemodynamics and emotional stress. In: *Psychosomatics in Essential Hypertension*. *Bibl. Psychiat.*, 144. M. Koster, H. Musaph and P. Visser, eds. Karger, Basel.
- Cruz-Coke, R. 1960 Environmental influences and arterial blood pressure. *Lancet*, 2: 885-886.
- Dawber, T. R., W. B. Kannel, A. Kagan, R. K. Donabedian, P. M. McNamara and G. Pearson 1967 Environmental factors in hypertension. In: *The Epidemiology of Hypertension*. J. Stamler, R. Stamler and T. N. Pullman, eds. Grune and Stratton, New York.
- Epstein, F. H. 1964 Hereditary aspects of coronary heart disease. *Am. Heart J.*, 67: 445-456.
- Feldman, J. G., M. A. Ibrahim and H. A. Sultz 1973 Differential filial aggregation of coronary risk factors. *Hum. Biol.*, 45: 541-552.
- Harburg, E., W. J. Schull, J. C. Erfurt and M. A. Schork 1970 A family set method for estimating heredity and stress. I. A pilot survey of blood pressure among Negroes in high and low stress areas, Detroit, 1966-1967. *J. Chron. Dis.*, 23: 69-81.
- Hoobler, W. W. 1961 Current concepts of the mechanism of essential hypertension. In: *Hypertension: Recent Advances*. A. N. Brest and J. H. Moyer, eds. Lea and Febiger, Philadelphia.
- Johnson, B. C., F. H. Epstein and M. O. Kjelsberg 1965 Distributions and familial studies of blood pressure and serum cholesterol levels in a total community: Tecumseh, Michigan. *J. Chron. Dis.*, 18: 147-160.
- Kannel, W. B., T. R. Dawber, A. Kagan, N. Revotzkie and J. Stokes 1961 Factors of risk in the development of coronary heart disease—six year follow-up experience: The Framingham Study. *Ann Intern. Med.*, 55: 33-50.
- Keyfitz, N., and W. Flieger 1971 *Population: Facts and Methods of Demography*. W. H. Freeman and Company, San Francisco.
- Koster, M. 1970 Patterns of hypertension. In: *Psychosomatics in Essential Hypertension*. *Bibl. Psychiat.*, 144. M. Koster, H. Musaph and P. Visser, eds. Karger, Basel.
- Levi, L. 1970 Emotional stress and sympathoadrenomedullary and related physiological reactions with particular reference to cardiovascular pathology. In: *Psychosomatics in Essential Hypertension*. *Bibl. Psychiat.*, 144. M. Koster, H. Musaph and P. Visser, eds. Karger, Basel.
- Lowe, C. R., and T. McKeown 1962 Arterial pressure in an industrial population. *Lancet*, 1: 1086-1092.
- Lowenstein, F. W. 1961 Blood pressure in relation to age and sex in the tropics and subtropics. *Lancet*, 1: 389-392.
- McKeown, T., R. G. Record and A. G. W. Whitfield 1963 Variation in casual measurements of arterial pressure in two populations. *Clin. Sci.*, 24: 437-450.
- Miall, W. E., P. Heneage, T. Khosla, H. G. Lovell and F. Moore 1967 Factors influencing the degree of resemblance in arterial pressure of close relatives. *Clin. Sci.*, 33: 271-283.
- Pickering, G. W. 1961 *The Nature of Essential Hypertension*. J. and A. Churchill LTD, London.
- Record, R. G., and A. G. W. Whitfield 1964 The effect of mortality on the relation of arterial pressure to age. *Clin. Sci.*, 27: 385-391.
- Schull, W. J., E. Harburg, J. C. Erfurt, M. A. Schork and R. Rice 1970 A family set method for estimating heredity and stress. II. Preliminary results of the genetic methodology in a pilot survey of Negro blood pressure, Detroit, 1966-67. *J. Chron. Dis.*, 23: 83-92.
- Scotch, N. A. 1963 Sociocultural factors in the epidemiology of Zulu hypertension. *Amer. J. Publ. Hlth.*, 53: 1205-1213.
- Scotch, N. A., and H. J. Geiger 1963 The epidemiology of essential hypertension: A review with special attention to psychologic and sociocultural factors. II. Psychologic and sociocultural factors in etiology. *J. Chron. Dis.*, 16: 1183-1213.
- Shaper, A. G. 1967 Blood pressure studies in East Africa. In: *The Epidemiology of Essential Hypertension*. J. Stamler, R. Stamler and T. N. Pullman, eds. Grune and Stratton, New York.

- Shreffler, D. C., C. F. Sing, J. V. Neel, H. Gershowitz and J. A. Napier 1971 Studies on genetic selection in a completely ascertained Caucasian population. I. Frequencies, age and sex effects, and phenotype associations for 12 blood group systems. *Amer. J. Hum. Genet.*, 23: 150-163.
- Sing, C. F., D. C. Shreffler, J. V. Neel and J. A. Napier 1971 Studies on genetic selection in a completely ascertained Caucasian population. II. Family analyses of 11 blood group systems. *Amer. J. Hum. Genet.*, 23: 164-198.
- Sinnock, P., and C. F. Sing 1972a Analysis of multilocus genetic systems in Tecumseh, Michigan. I. Definition of the data set and tests for goodness-of-fit to expectations based on gene, gamete, and single locus phenotype frequencies. *Amer. J. Hum. Genet.*, 24: 381-392.
- 1972b Analysis of multilocus genetic systems in Tecumseh, Michigan. II. Consideration of the correlation between nonalleles in gametes. *Amer. J. Hum. Genet.*, 24: 393-415.
- Snedecor, G. W., and W. G. Cochran 1971 *Statistical Methods*. The Iowa State University Press, Ames.
- Sokal, R. R., and F. J. Rohlf 1969 *Biometry*. W. H. Freeman and Company, San Francisco.
- Stamler, J., D. M. Berkson, H. A. Linberg, W. A. Miller, R. Stamler and P. Collette 1967 Socioeconomic factors in the epidemiology of hypertensive disease. In: *The Epidemiology of Hypertension*. J. Stamler, R. Stamler and T. N. Pullman, eds. Grune and Stratton, New York.
- Strehler, B. L. 1962 *Time, Cells and Aging*. Academic Press, New York.
- Weiner, H. 1970 Psychosomatic research in essential hypertension: Retrospect and Prospect. In: *Psychosomatics in Essential Hypertension*. *Bibl. Psychiat.*, 144. M. Koster, H. Musaph and P. Visser, eds. Karger, Basel.