

HEREDITY, STRESS AND BLOOD PRESSURE, A FAMILY SET METHOD—I

STUDY AIMS AND SAMPLE FLOW*

ERNEST HARBURG, JOHN C. ERFURT,† WILLIAM J. SCHULL,‡
M. ANTHONY SCHORK§ and ROBERT COLMAN¶

Department of Epidemiology, School of Public Health and Department of Psychology,
The University of Michigan, Ann Arbor, MI 48103, MI, U.S.A.

Abstract—This first article, in a series of five, describes the method of sampling family sets. Family sets are composed of three persons having a genetic relation: an index, his/her sibling and first cousin, and two persons having an environmental nexus, a spouse of index and an unrelated person matched to the index. The target populations were four census areas in Detroit, a black high stress area, a black low stress area, a white high stress, and a white low stress area. These areas were selected by a factor analysis of census rates which indicated extremes of stressor conditions. Within each area a complete census was taken, potential sample members were selected and verified by another interviewer, then assigned as an index; sibs and first cousins, selected as closest in age to index, were verified independently, then an unrelated person was chosen, and all five persons were independently interviewed and blood pressures taken. This article details the full sampling process in each of the four census areas, and tests the final sample of 461 family sets in several ways which confirmed expectations.

INTRODUCTION

THE PROFOUND consequences of high blood pressure levels have now been documented by a number of major longitudinal studies [1-6]. These studies have also reconfirmed the familial presence of high blood pressure, but have not provided tests of hereditary contributions to its occurrence. This series of five reports should provide a start to further research in this critical area by delineating methods of inquiry useful for the broad aim of partitioning hereditary and environmental components in human populations. Specifically, our objective in these articles are

*Funded by National Center for Chronic Disease Control, No. CD-00209-01, 02, 03; National Center for Health Services, Research and Development, No. HS 00164-04; National Association of Mental Health; Michigan Heart Association; National Heart and Lung Institute, No. HE 13329-01, 02, 03; American Heart Association; National Institute of Mental Health, MH 20621-01; Fannie E. Rippel Foundation.

†Co-Director, Worker Health Program, Institute of Labor and Industrial Relations, The University of Michigan.

‡Director, Center for Demographic and Population Genetics, The University of Texas, Houston.

§Professor, Department of Biostatistics, School of Public Health, The University of Michigan.

¶Research Assistant, Program for Urban Health Research, Department of Psychology, The University of Michigan.

(1) to assess the genetic contribution to blood pressure readings within black and white family sets in Detroit, and (2) to describe a method of sampling and analyzing family set data to assess hereditary and environmental components in phenotypic variation, e.g. blood pressure.

The idea of sampling family sets arose from an interdisciplinary project in the early 1960's [7] focused on rheumatoid arthritis. Subsequent modification of the early family set design resulted in a pilot project, 1966–1967, which sought to develop further the general method and describe hereditary and psychosocial stressor contributions to blood pressure [8, 9]. In 1968, a third project was initiated using an enlarged sample of over 400 family sets, again focused on blood pressure. This third project provided the data for the present series of five articles. In this first article, the overall design, the sampling methodology and the description of the sampling stages and results will be described in sufficient detail to aid in evaluating the method and the genetic results.

THE STRATEGY OF FAMILY SETS AND STRESSOR AREAS

Genetic theory specifies the proportion of independently segregating autosomal genes which various biological relatives will share on the average. Thus, parent and offspring or siblings will share half of their genes in common; first cousins, one-eighth; and so forth. These proportions may also be interpreted as the probability that for any given (autosomal) gene, specified in one person, the stated relative will possess the same gene, in the sense of a gene of common origin. These proportions are termed coefficients of relationship, and are clearly measures of genetic similarity. Commonness of origin implies that both genes have been derived from a single ancestral gene.

The complements of the aforementioned coefficients may be regarded as measures of genetic dissimilarity or distance from commonness of origin. In this latter view, the average percentage of genes of dissimilar origin between (1) an index case and himself (or herself) is zero, (2) index and his (her) sibling is 50%, (3) index and his (her) first cousin is 87%, (4) index and his (her) spouse is 100%, assuming no biological relationship, and (5) index and his (her) control (or unrelated) is 100%. This genetic scale, it will be noted, ranges from 0 to 100% and, to reiterate, represents the expected proportion of independent autosomal genes of dissimilar origin under random mating. The study analysis attempts to relate these known differences in genetic distance between an index case and other members of his or her family set to differences in blood pressure with due allowance for other sources of variability, primarily socio-environmental.

The socio-environmental variation of interest in this study is that which obtains among major racial groups in an urban context. We conceive of socioecological stressors as social processes in local populations which indicate the relative availability of socioeconomic resources, and the stability of major institutional patterns. Thus, for example, both level of income and education, as well as marital stability and crime, were considered in estimating relative degree of stress-inducing conditions in local populations. This conceptual framework follows the logic that socially disorganized residential areas generate problem situations requiring adaptation more often and with less resources for solution than more organized areas; in turn, these socioecological high stressor areas should be related to higher morbidity

and mortality, e.g. [10, 11]. It is important that these external conditions be measured independently of the emotional states of individuals in the local populace [12].

Other indicators of stressors measured in this project that were also of interest in accounting for blood pressure variation were subjective reports by individuals about their residential security, financial concerns, marital discord [13] and coping with anger [14]. These measures were collected on all five members of a unit we term a family set.

A 'family set' in this study consists of three persons having a genetic relation, an Index, his or her Sibling and First Cousin, and two persons having an environmental nexus, the Spouse of the Index and an Unrelated person matched to the Index. The spouse serves as an 'immediate environment' control for the index, e.g. sharing the same dwelling unit, diet, water and air, and the unrelated person serves as a control for other critical environmental factors outside the dwelling unit but also within the same urban, socio-environmental area as the spouse and index. Both spouse and unrelated then are persons not genetically related to the index, but who do share certain broad aspects of the urban socioecologic niche. If morbid processes are also shared between index and spouse and index and unrelated more than between index and sib or among index, sib, and first cousin, then certainly environmental factors become prominent in any search for etiology 'or intervention.'

DESIGN OF THE MAJOR STUDY

Overview

To execute the strategy of family sets and stressor areas, we conceived the following design: First, four census areas in Detroit were selected, in ways to be described shortly, to represent extremes of socioecological stressor areas for blacks and whites. We call these areas black high stress, black low stress, white high stress, and white low stress. Within each area a census was taken and persons who reported having the characteristics for being a member of a family set were listed and randomly selected for verification. These traits were then verified by another visit, and selection of a nearest-in-age sib and first cousin was followed by their further verification. Once all five members of a set were verified, nurse-interviewers obtained medical and personal history and blood pressure readings from each set member. The sampling and verification and nurse interviews were stopped when a sufficient number of family sets were obtained from each area, i.e. over 100. For a detailed specification of this entire process, see [10, 15].

Selection of high and low stress areas

The first objective was to select residential areas in Detroit which varied in extremes of high stressor and low stressor conditions relative to the City [16]. Rates were computed for each of 382 census tract areas in Detroit on 1965 data for variables reflecting economic deprivation, residential instability, family instability, crime, and population density. It was assumed that such combined rates, at their end-points, indicate social environments which vary objectively in chronic exposure to stressor events. The rates per census tract were factor analyzed and

TABLE 1. RATES FOR PRIMARY STRESS AREAS, DETROIT, 1965*

Characteristics	Black		White	
	High Stress	Low Stress	High Stress	Low Stress
A. Socioeconomic variables				
1. Median income (\$)	4627	8670	5417	8030
2. Median education (yr)	9.6	13.2	9.0	11.7
3. % unemployed	4	0	0	0
4. % home ownership	19	92	40	90
5. % professional/managerial	9	49	7	19
B. Instability variables				
1. Adult crime rate (per 10,000)	89.0	55.9	60.0	9.9
2. Juvenile crime rate (per 10,000)	17.2	6.4	13.5	1.3
3. Marital instability†	1:2.9	0.00	1:12	0.00
4. % in residence 5 yr or more	27%	51%	48%	86%

*These data are from a 4% sample of the City of Detroit by the Transportation and Land Use Study (1965), except for crime data supplied by the City of Detroit Police Department.

†This is a ratio of the number of separations and divorces to the number of marriages. No separations or divorces were reported in the low stress areas in the 4% sample drawn to represent the City in 1965.

the 382 tracts were each assigned factor scores for the two emergent factors: Socio-economic Status (SES) and Instability. The factor scores were separately rank ordered within all predominantly black tracts (50% or more Negro) and within all the residual white tracts. Census tracts within the upper and lower quartiles for each factor score list were then selected for having *both* the upper range for the Instability score and the lower range for the SES scores. These tracts were labeled 'high stress.' The converse was done to delimit the 'low stress' tracts. Data, not shown, indicate that the final selection of four primary study areas (black high and low stress, white high and low stress) were in the extreme quartiles of the two factor scores; such data and further descriptions of the sociodemographic background of these areas are presented in [10].

Table 1 gives actual rates of the study tracts after selection on the basis of factor scores. The rates show predicted sharp differences between high and low stress across median income, median education, crime rates, and marital breakup. This same pattern is revealed in other rates (not shown), e.g. school truancy, 'drop-outs,' welfare registration, Aid to Dependent Children, and so forth. Not shown is the fact that tracts were also required to have a minimal per cent of married pairs (50%) and the age distributions of the 30-54 population were normal with a median around 40 ± 4 yr. It should be noted that both low stress areas are each single census tracts; both high stress areas, however, are each four contiguous census tracts equivalent in social characteristics. In the latter areas, multiple tracts were required to increase the yield of families required by the genetic design in the larger Project.

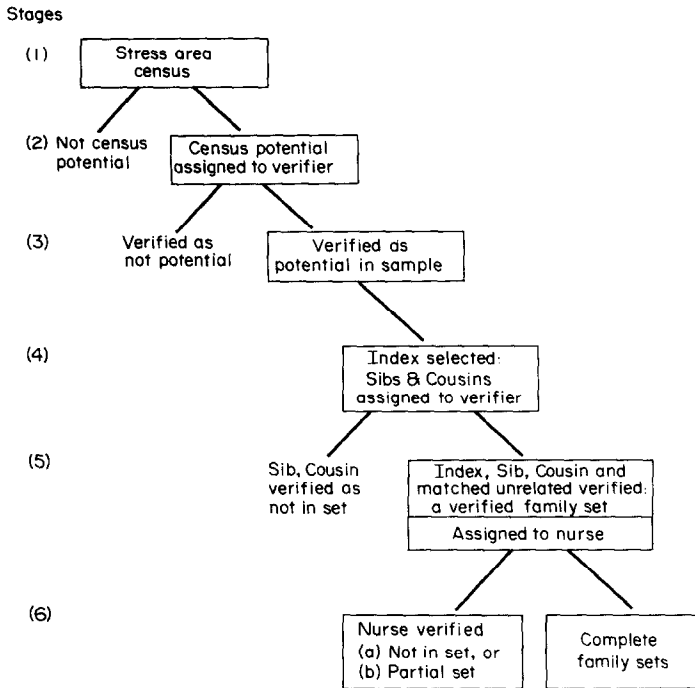


FIG. 1. Overview of sampling flow.

Overview of sampling flow

Figure 1 permits a visual overview of the important steps in locating persons who reported having family sets, verifying and sampling such index persons, verifying all persons in the family set as the population of such sets, and assigning to the nurse-interviews the final sample of set members. This flow was followed in each of the four stress areas independently from the others; thus each area yielded a separate population list of persons with family sets within which samples could be drawn.

The six major steps in the sample flow (Fig. 1) are as follows: (1) First, a complete census of dwelling units in the stress area was carried out. (2) Persons with reported family sets in the census we refer to as Census Potentials, i.e. persons found in the initial screening census reporting traits making them potential sample members. A strict definition of Census Potential is a person who is: (a) residing in the given stress area and of the majority race of that area; (b) married and living with spouse; (c) 25–59 yr of age; (d) reporting a full sibling of either sex living in the greater Detroit area; and (e) reporting a full first cousin also residing in the greater Detroit area. Persons who were Not Census Potentials were excluded from the target population list. The strategy here and throughout the sample flow was to accept the loss of false negatives, but to check or verify so that false positives could be eliminated. Tactically, even certain types of reported negative cases were checked in order to enlarge the base of true positives. We call these types 'Don't Know Cousins' and 'Non-respondents' and will discuss these potential

false negatives later. (3) The third major process was to verify the initial census reports by assigning an interviewer to visit and check the required traits of *all* Census Potentials. At this visit a complete listing of names, addresses, age, married name, phone numbers and so forth was elicited for all siblings and first cousins (within 10 yr of the respondent's age) residing in the Detroit Area. Family historian names were also collected. At this stage, a list of Verified Potential index persons was compiled, and all persons verified as Not Potential sample members were excluded from the population list.

The next Step (4) in Fig. 1 shows that an index person was selected. This was done by random sampling from the list of index persons whose sibs and first cousins closest in age (± 6 yr) were verified as to their relationship, age, and availability. When such information was negative, that potential index person was excluded from the family set listing. Step (5) indicates that after the index, sib, and first cousin were verified, an unrelated person from the verified index list was matched to the index on residence in area, race, sex, marital status, age and having sibs and first cousins in the Detroit area. In this process, there was an incremental listing of Verified Family Sets whose members were then randomly assigned in Step (6) to nurse-interviewers. Nurses then took an hour and a half interview, during which blood pressure was measured three times in the first half hour and twice at the close. Medical history, personal stress, and demographic data were also obtained. For details of the blood pressure measures, see the next paper in this series [17].

The design of nurse assignments to interview family set members was guided by these principles: (a) the nurse was of the same race as the respondent; (b) no one nurse could interview more than one family member of each set with the exception of index and spouse, where the latter always had a half-hour interview. This exception was required by limited funds and not by design; (c) nurses were randomly assigned to family set members and also by random time assignments during the day of the interview; (d) nurses were alternated weekly between the high and low stress areas; (e) assignments of sibs and cousins were clustered by area of the city or suburb to facilitate completion of interviews per day, but such clusters in different parts of the Detroit area were also alternated among nurses. These procedures aided in assuring that the blood pressures within each family set were obtained by random observers and offset this source of bias within sets. Analysis of variance results, (not shown) indicate that such precautions achieved their objective, e.g. index, sib, first cousin and unrelated persons, as *individuals*, did not differ significantly on height, weight, age, pulse rate, smoking, and blood pressure. Finally, Step (6) in Fig. 1 indicates that when all five members of a set were interviewed, the set was designated as a Complete Family Set. Again, persons nurse-verified as not having the requisite traits were dropped along with other members of their set, and those sets in which less than five persons were interviewed are labeled 'Partial Sets,' useful for psychosocial information.

It should be noted that 11 independent checks from census to verifier to nurse were made to ascertain that the family relationships required by the genetic model did in fact exist [8]. Furthermore, the names and addresses of set members were cross-checked with other family set members to reduce the error of non-independence among family sets, i.e. the design requires each set be independent in family

TABLE 2. SAMPLING FLOW: CENSUS, VERIFICATION, SAMPLE SELECTION, AND NURSE INTERVIEW, DETROIT, 1968

Phases of sample	Total (%) (N)	Stress area			
		Black		White	
		High stress (%) (N)	Low stress (%) (N)	High stress (%) (N)	Low stress (%) (N)
A. Census					
1. No. of dwelling units in census	12,294	4379	1952	4183	1780
2. No. of persons interviewed	17,961	5422	3272	6419	2848
3. Yield of census potentials, index or unrelated (of all persons interviewed)	4205 (23%)	986 (18%)	915 (28%)	1205 (19%)	1099 (39%)
B. Verification of census potentials					
1. Verified as sampling potential (of all census potentials)	1859 (44%)	408 (41%)	444 (49%)	442 (37%)	565 (51%)
C. Sample selection					
1. Persons selected as index or unrelated	1195	294	303	300	298
2. Unused sample	664	114	141	142	267
D. Persons interviewed by nurse					
1. Index and unrelated in 5-member family set (of all persons selected)	992 (77%)	224 (76%)	234 (77%)	234 (78%)	230 (77%)
2. Number of family sets with index, sib, cousin, spouse and unrelated	461	112	117	117	115

ties from all the others (indeed, in the white high stress area, a cluster of first cousin marriages was exposed by such checking). Table 2 gives a statistical summary of the sampling flow in Fig. 1.

Census results

To understand better the process of sampling family sets, we must take a closer view of each major phase. Table 3 presents data describing the census, using dwelling units as units of analysis. Line A2 shows that 93% of all DU's were inhabited and contact was made with these units by the team of census interviewers. It is further validity of the black high stressor area that 13% of all DU's in the area were Not Inhabited, i.e. vacant or abandoned, or double the rate for the white high stress and 13 times that of the low stress areas. Line B3 shows that of all the inhabited DU's, nearly 90% (89%) were interviewed (with only a 2% refusal rate). Next, C2 indicates that of all the DU's interviewed, about 24% had Census Potentials; however, the high stress areas had about half the yield compared to the low stress areas. We had censused contiguous high stress tracts having similar socioecological rates; while this allowed us to attain the number of potentials to fill the design of 100 family sets in each area, the final yields were still

TABLE 3. RESULTS OF CENSUS OF DWELLING UNITS (DU'S) BY STRESS AREA

	Stress area									
	Total		Black				White			
	(%)	(N)	High stress (%)	High stress (N)	Low stress (%)	Low stress (N)	High stress (%)	High stress (N)	Low stress (%)	Low stress (N)
A. Total dwelling units (DU's)	100%	(12,294)	100%	(4379)	100%	(1952)	100%	(4183)	100%	(1780)
1. Not inhabited	7	(806)	13	(579)	1	(26)	5	(190)	1	(11)
2. Inhabited	93	(11,488)	87	(3800)	99	(1926)	95	(3993)	99	(1769)
B. Total DU's inhabited	100%	(11,488)	100%	(3800)	100%	(1926)	100%	(3993)	100%	(1769)
1. Not interviewed (not at home, etc.)	9	(1033)	11	(419)	9	(166)	7	(286)	9	(162)
2. Refusals	2	(288)	3	(107)	4	(75)	2	(74)	2	(32)
3. Total DU's interviewed	89	(10,167)	86	(3274)	87	(1685)	91	(3633)	89	(1575)
C. Total DU's interviewed	100%	(10,167)	100%	(3274)	100%	(1685)	100%	(3633)	100%	(1575)
1. Without potential cases*	76	(7779)	82	(2676)	68	(1146)	83	(3021)	59	(926)
2. With potential cases	24	(2398)	18	(598)	32	(539)	17	(612)	41	(649)
D. Total DU's with Potential cases	100%	(2398)	100%	(598)	100%	(539)	100%	(612)	100%	(649)
1. Husband only is potential	33	(796)	39	(231)	31	(169)	35	(211)	29	(185)
2. Wife only is potential	36	(864)	35	(207)	35	(191)	39	(241)	35	(225)
3. Both husband and wife are potential	31	(738)	26	(160)	33	(179)	26	(160)	37	(239)

*Potential refers to a person who is (1) of the same given race of the stress area, (2) married and living with spouse, (3) 25-59 yr old, (4) with a full sibling living in the Detroit area, and (5) a full first cousin in the Detroit area. These are 'strict' Census Potentials. However, we have also included here persons with criteria (1) through (4) who 'didn't know' if they had first cousins in the Detroit Area or not. These are 'DK' (didn't know cousin) Census Potentials.

relatively lower than in the middle-class areas. Finally, in Table 3, line D3, the data show that about one-third (31%) of DU's with Census Potentials had both husband and wife potentials. The family set design however requires independence of dwelling units and family living, thus only one member of such a pair was eventually selected for inclusion in the sample as an index case by a random procedure *after* verification of each spouse.

Table 4 reveals the same census with individuals, not DU's, as the unit of analysis. In Section A, the reasons for being classified as Not Potential can be seen: about 70% of all persons were excluded on the basis of race, marital status, and age (A1, 2, 2). Line A2 shows that the black high stress area had over two-fifths (45%) reported as not married or living with spouse or about eight times the rate of the black low stress and double that of the white middle class. Among the whites, and more so in the low stress area, the primary reason for exclusion is shown in line A3, i.e. being outside the age range. The percentages excluded for reasons of lacking relatives in lines A4, 5, and 6 are highly similar; thus only about 30% of the Not Potentials (total, lines A4, 5, and 6) were due to not having siblings or first cousins, who lived in the Detroit area.

TABLE 4. RESULT OF CENSUS OF INDIVIDUALS BY STRESS AREA

	Total (%) (N)	Stress Area					
		Black			White		
		High stress (%) (N)	Low stress (%) (N)	High stress (%) (N)	Low stress (%) (N)	High stress (%) (N)	Low stress (%) (N)
I. Total persons interviewed	100% (17,961)	100% (5422)	100% (3272)	100% (6419)	100% (2848)		
A. Census not potentials:*	100% (13,756)	100% (4436)	100% (2357)	100% (5214)	100% (1749)		
1. Not majority race of area	18 (2539)	11 (483)	52 (1219)	16 (810)	1 (27)		
2. Not married or not living with spouse	30 (4060)	45 (2005)	6 (141)	29 (1513)	23 (401)		
3. Not age 25-59	21 (2906)	18 (818)	5 (124)	26 (1341)	36 (623)		
4. No sib; has cousin	6 (891)	6 (280)	10 (240)	3 (204)	10 (167)		
5. No cousin; has sib	7 (920)	5 (198)	7 (171)	7 (365)	10 (186)		
6. No sib and cousin	18 (2440)	15 (652)	20 (462)	19 (981)	20 (345)		
B. Census potentials:†	100% (4205)	100% (986)	100% (915)	100% (1205)	100% (1099)		
1. Not assigned to verifier‡	25 (1055)	22 (212)	21 (195)	32 (389)	24 (259)		
2. Assigned to verifier	75 (3150)	78 (774)	79 (720)	68 (816)	76 (840)		
C. Assigned to verifier	100% (3150)	100% (774)	100% (720)	100% (816)	100% (840)		
1. Strict CP	87 (2750)	83 (640)	94 (677)	79 (641)	94 (792)		
2. 'Maybe' CP§	13 (400)	17 (134)	6 (43)	21 (175)	6 (48)		

*The six reasons for classification as a 'Census Not Potential' are rank-ordered in priority for data analysis. Presence in a higher rank deducts that person from the count in lower ranks. e.g. a person 'not married or not living with spouse' is not in the count of age exclusions.

†This refers to both Strict and DK Census Potentials as defined in the Table 3 footnote plus 'Non-respondents'. A Non-Respondent refers to a person having all criteria for potential except not having a first cousin *and* the census data about this person was obtained from another person in the house. Of all Census Potentials about 71% were Strict, 26% were Non-respondents, and 3% were DK's only (see text).

‡These persons were 'left over' Census Potentials after the estimated yield had been randomly selected from the incoming listing. About 75% of these were 'Non-Respondents.'

§'Maybe' CP's refers to both Non-Respondents and DK Census Potentials.

Several results in Table 4, Section B require further explanation. Persons in B1, labeled Not Assigned to Verifier are residuals of possible false negatives due to the use of a dynamic sampling frame model. It was not feasible to allow time first to collect a complete, up-to-date listing of Census Potentials before verification started. The residential mobility in the high stress areas required short turn-around-time between the census and verification and interview, and the full utilization of field staff required continuous assignment because, subsequently, as Census Forms arrived at the office, they were screened for Census Potentials and this list was used for a random assignment of persons to be verified. The reason that 25% of all Census Potential were not assigned (line B1) was largely due to the subset of potentials we refer to as 'Non-respondents.' Non-respondents were defined as persons having most of the criteria to be a Census Potential, but the census data were obtained from *another* person in the house who reported absence of a first cousin. Our experience shows that about 26% of all Census Potentials were classed as Non-respondents, and about 20% of such Non-respondents did in fact have a first cousin known to the persons themselves or by others. These 'false negatives' were assigned for verification only *after* all the positive or Strict

Census Potentials were assigned and, therefore, most of these persons were not assigned for verification.

The next type of possible false negatives who were assigned to be verified we called the 'Don't Know Cousins' or DK type. These persons have all the criteria of being a Strict Census Potential except they 'don't know' if they have a first cousin in the Detroit area. Results in the Pilot Study and this Major Study reveal that over one-fourth (about 27%) of these persons can be shown to have first cousins within 6 yr of their age. This requires questioning of spouse, or siblings, or at times, the family historians. Line C2 in Table 4 shows that of all 3150 persons assigned to the verifier, about 400 or 13% were 'Maybe Census Potentials' composed of primarily Non-respondents and DK's. (Not shown is the result that 85 or 21% of these Maybe's were in fact verified as members of the target population of persons with family sets.) The final results in Table 4, line C2, show that most of these Maybe types were obtained from the high stress areas where the initial low yield of Census Potentials was aided by this kind of subroutine engineering (developed in the Pilot Study). Such a mechanism allowed an increase in the final yield of numbers to attain 100 or more family sets from each stress area. From a sampling viewpoint, these measures to verify both the true positive and possibly false negative are desirable in order to enlarge the base population of potential family sets and ensure a more representative sample.

Results of verification

Table 5, line A, shows that of the eventual 3150 individuals assigned to be verified, about 13% were non-interviews, with the lowest cooperation among white low stress persons and (not shown) among black low stress females. (This rate of 13% compares favorably with the per cent survey rate of about 15–20% on national samples in high stress urban areas [18]). Next, Sections B and C indicate the reasons why these persons were not included in the population listing of family sets. While the time between census and verification ranged from 2 weeks to about 3 months, it averaged about 8 weeks. Results in lines B1 and B3 show clearly that residential mobility and marital separations were highest in the black high stress area as expected. For section C, dealing with loss of potential sample members due to failure to verify positive sibs and cousins, the percents across the four race-social class groups are similar, but lower in the white low stress area. At this stage, family relationships contribute more to exclusion than do the sociodemographic factors (status changes). Finally, Section D shows that about three-fifths (59%) of all individuals assigned to be verified were indeed verified as having family sets; again this per cent varied from about half in the high stress to about two-thirds in the low stress areas. At this stage in sampling, therefore, 1859 persons were verified (by an interviewer independent from the census) to be Potential Index/or Unrelated persons.

Sampling classifications

The original logical design of the sampling process called for a random assignment of verified potentials to be either an index or an unrelated person, and for sibs and cousins to be within 6 yr of age of the index. The daily statistics compiled in the field office compelled a revision in this plan. Table 6 exposes

TABLE 5. RESULTS OF VERIFICATION: REASONS FOR CLASSIFICATIONS AS VERIFIED NOT-POTENTIALS AND VERIFIED POTENTIAL BY STRESS AREA

Individuals	Stress Area									
	Total		Black				White			
	(%)	(N)	(%)	(N)	(%)	(N)	(%)	(N)	(%)	(N)
I. Total assigned to verify:	100%	(3150)	100%	(774)	100%	(720)	100%	(816)	100%	(840)
A. Non-Interviews: (refusals, not-at-home, etc.)	13	(416)	9	(71)	13	(91)	14	(113)	17	(141)
B. Change in census status:	6%	(189)	10%	(76)	4%	(32)	7%	(56)	3%	(25)
1. Moved out of area	35	(66)	53	(40)	25	(8)	23	(13)	20	(5)
2. Not of given race in area	27	(51)	4	(3)	37	(12)	63	(35)*	4	(1)
3. Separated from spouse since census	17	(32)	28	(21)	13	(4)	7	(4)	12	(3)
4. Not age 25-59	21	(40)	15	(12)	25	(8)	7	(4)	64	(16)
C. No sib and/or cousin in Detroit area:	22%	(686)	28%	(219)	21%	(153)	25%	(205)	13%	(109)
1. No sib; DK or yes cousin	16	(111)	21	(46)	16	(25)	13	(27)	12	(13)
2. No cousin; DK or yes sib	65	(445)	60	(132)	63	(96)	68	(139)	72	(78)
3. No sib and cousin	19	(130)	19	(41)	21	(32)	19	(39)	16	(18)
D. Total verified potential:	59%	(1859)	53%	(408)	62%	(444)	54%	(442)	67%	(565)
1. Strict CP†	95	(1774)	93	(379)	98	(434)	92	(406)	98	(555)
2. Maybe CP's†	5	(85)	7	(29)	2	(10)	8	(36)	2	(10)

*These were Mexican-Americans whom we excluded from the study for reason of cultural differences.

†Defined in footnotes in Tables 3 and 4.

the manner in which these sampling rules were relaxed. First, not shown, was the decision that all addresses of potential sibs and cousins should show that they resided in the Detroit area, or at least, the address in the area was not known. Second, a rank order of relaxing the age range for sibs and cousins was agreed upon in the order of age groups listed in A and B. Third, line B1 shows persons of age less than 30 and greater than 54 who could not be index persons and must be unrelated by definition. Line B2 ($N = 265$) indicates we also assigned to the class of Potential Unrelateds those persons 30-54 where either the siblings or first cousins were over 10 yr of age distant from the index. We assumed that the age range of the relatives of the index or unrelated would not be associated with differences in the index/unrelateds' own physical traits of interest to this study. And in fact, after data collection, separate analysis of variance (not shown) of the groups of index or unrelated persons in Table 6, Sections A and B2, revealed no significant differences of index or unrelateds in age, weight, systolic or diastolic blood pressures. This means that, e.g. blood pressure among index or unrelateds did not differ within stress areas regardless of the age groups of their sibs and first cousins. We have already indicated that there were no significant differences in the final sample between index and unrelateds on these traits (t -tests, data not shown).

TABLE 6. SAMPLING CLASSIFICATIONS OF VERIFIED POTENTIALS BY AREA

Verification	Stress area									
	Total		Black				White			
	(%)	(N)	(%)	(N)	(%)	(N)	(%)	(N)	(%)	(N)
I. Total verified potentials	100%	(1859)	100%	(408)	100%	(449)	100%	(442)	100%	(565)
A. Classified potential index or unrelated with age difference from:	64%	(1182)	58%	(238)	68%	(302)	63%	(280)	64%	(362)
Sib										
Cousin										
0-6 yr	78	(922)	71	(170)	80	(242)	76	(214)	82	(296)
0-6 yr	12	(145)	17	(41)	13	(39)	13	(37)	8	(28)
7-10 yr	8	(98)	9	(21)	5	(16)	10	(27)	9	(34)
7-10 yr	2	(17)	3	(6)	2	(5)	1	(2)	1	(4)
B. Classed as potential unrelateds:	36%	(677)	42%	(170)	32%	(142)	37%	(162)	36%	(203)
1. Age of index*	61	(412)	59	(100)	46	(66)	60	(97)	73	(149)
2. Age of sibs or cousins > 10:†	39	(265)	41	(70)	54	(76)	40	(65)	37	(54)
Sib										
Cousin										
0-6 yr	78	(206)	81	(57)	78	(59)	85	(55)	65	(35)
7-10 yr	6	(15)	6	(4)	5	(4)	6	(4)	6	(3)
10+ yr	9	(25)	7	(5)	8	(6)	5	(3)	20	(11)
10+ yr	4	(10)	3	(2)	4	(3)	1	(1)	7	(4)
10+ yr	3	(9)	3	(2)	5	(4)	3	(2)	2	(1)

*Persons aged 25-29 or 55-59 could only be assigned as unrelated persons ± 5 yr of age to be matched to index persons 30-54.

†Persons 30-54 with sibs or cousins > 10 yr were not assigned as index persons.

Table 7, line A, shows that about 36% of the 1859 Verified Potentials were not assigned to the final nurse-interview, either because of exclusion from the population list because of being a resident spouse (A1) or because sampling of incoming Verified Potentials was terminated due to attaining the required number

TABLE 7. SAMPLE POTENTIALS ASSIGNED TO NURSE INTERVIEWER

Verification	Stress area									
	Total		Black				White			
	(%)	(N)	(%)	(N)	(%)	(N)	(%)	(N)	(%)	(N)
I. Total verified potentials:	100%	(1859)	100%	(408)	100%	(444)	100%	(442)	100%	(565)
A. Verified potentials not assigned to nurse:	36%	(664)	28%	(114)	32%	(141)	32%	(142)	47%	(267)
1. R's spouse was assigned*	30	(202)	27	(31)	38	(54)	31	(44)	27	(73)
2. Potential index or unrelated†	22	(146)	8	(9)	26	(36)	23	(32)	26	(69)
3. Potential unrelated†	48	(316)	65	(74)	36	(51)	46	(66)	47	(125)
B. Total verified index/unrelated potentials assigned to nurse:	64%	(1195)	75%	(294)	68%	(303)	68%	(300)	53%	(298)

*The design required that index and unrelated reside in separate dwelling units.

†These were residuals after estimated yield had been randomly selected from incoming listing of verified potentials.

of family sets in the area. It will be noted that almost half of the Not Assigned persons were from the class of Potential Unrelateds (see Table 6, Section B) and that only 8% of 9 persons (see Table 7, I, A, 2) were not assigned in the black high stress area, i.e. the sample in this stress area was practically the population. These non-assignments left 1195 randomly selected Verified Index/Unrelated persons, or 64% of the total Verified Potentials, to be assigned to the nurses.

Results of interviewing the family set sample

The difficulty in describing a dynamic sampling frame is that any attempt to point to a static population list and final sample is simply an arbitrary matter. One might suggest that in Table 8, line I, the 1195 persons assigned to the Family Set list can be considered the population list. However, Section A indicates that even while this listing was compiled for the nurses, the verifiers were adding further exclusions to the list (line A, 7%) and it was expected and found that the nurses also would find persons to be excluded as shown in Section B, or 3%. On line B3, the reason for exclusion is of interest because, while not shown in prior tables, constant surveillance of independence of family sets required excluding family sets all during the verification, the nurse-interview, and even after data collection was over for the final listings! Therefore, about 10% (lines A and B) of the so-called final listing of family sets assigned to nurses were subsequently excluded from the list.

Table 8, lines C and D both indicate the nurses' success rate was high. Only 6% assigned were in fact not interviewed. Much of this achievement can be credited not only to the nurses' skills but also to the cooperation and quasi-appointment elicited by the verifiers and handled with efficiency by the field office. If one does in fact exclude the persons verified as a true negative in this stage (Sections A and B in Table 8), then about 94% of all 'true positives' assigned to the nurses were interviewed.

Final samples for analysis of data

The original design and logic of this Major Study, based on experience in the Pilot Study, directed that there be three final samples: a Socioecological Sample, Area Genetic Sample, and Family Set Sample. The sample of index and matched unrelateds we call the Socioecological Sample includes (1) those with complete family sets (sets in which all five members were interviewed by nurses), and (2) those with partial family sets (sets in which the sib and/or cousin was *not* interviewed by nurses). The total *N* with these criteria across all four areas was 1000 (see Table 8, IIA). These data are used for purposes of socio-psychological analyses of personal stress factors obtained in the nurse-interview [10].

The second and target sample is referred to as the Area Genetic Sample (*N* = 922) which is simply the Socioecological Sample *minus* those index and unrelated pairs with partial family sets, i.e. only those index and matched unrelated pairs with *complete* family sets. While this is the sample base for the Family Set Sample, it is not useful for genetic analysis but must be understood as a technical sub-sample of this larger final genetic sample.

TABLE 8. RESULTS OF FAMILY SET INTERVIEWING BY STRESS AREA

Final sample flow	Stress area									
	Total		Black				White			
	(%)	N	High stress (%)	Low stress (%)	High stress (%)	Low stress (%)	High stress (%)	Low stress (%)	High stress (%)	Low stress (%)
I. Total verified index or unrelated assigned to family set list:	100%	(1195)	100%	(294)	100%	(303)	100%	(300)	100%	(298)
A. Sib or cousin verified not available:	7%	(88)	6%	(17)	5%	(14)	8%	(23)	11%	(34)
1. No sib in Detroit area	17	(15)	24	(4)	0	—	22	(5)	18	(6)
2. No cousin in Detroit area	11	(10)	12	(2)	0	—	9	(2)	18	(6)
3. Sib or cousin refused, or not available	72	(63)	64	(11)	100	(14)	69	(16)	64	(22)
B. Verified by nurse as not in family set list:	3%	(32)	4%	(12)	2%	(5)	3%	(11)	1%	(4)
1. Moved from area	56	(18)	75	(9)	60	(3)	55	(6)	0	—
2. Separated; divorced	19	(6)	17	(2)	40	(2)	9	(1)	25	(1)
3. Related to other family set member	25	(8)	8	(1)	0	—	36	(4)	75	(3)
C. Index refusal or not available to nurse:	6%	(68)	6%	(17)	6%	(20)	5%	(14)	6%	(17)
D. Total nurse interview of assigned index or unrelated:	84%	(1007)	84%	(248)	85%	(264)	84%	(252)	82%	(243)
II. Final samples										
A. Socioecological sample*		(1000)		(246)		(262)		(250)		(242)
B. Area genetic sample†		(922)		(224)		(234)		(234)		(230)
a. Per cent of total assigned to family set list (I above)	77%		76%		77%		78%		77%	

*'Socioecological Sample' refers to *all* verified index and matched unrelated pairs who were interviewed by nurses, including (1) those with complete family sets, *and* (2) those with partial family sets (sets in which the sib and/or cousin was *not* interviewed by nurses). The *N* of 1007 was reduced to 1000 when 7 Index persons were dropped for lacking a matched unrelated person. This sample is primarily useful for social and psychological analyses.

†The 'Area Genetic Sample' is the Socioecological Sample *minus* those index and matched unrelated pairs with partial family sets, i.e. only those index and unrelated pairs with complete family sets (sets in which all five members were interviewed by nurses).

Family set sample

Table 9 exposes the final family set sample used in the genetic analyses reported in the next four articles. The total family set sample includes the index and unrelateds selected in each stress area with their verified, nurse-interviewed sib and first cousins and spouse of index. Table 9 reveals in the lower right corner that 461 complete family sets were finally obtained, or 2305 persons. Another requirement of the design can be seen clearly in the column headings and totals, namely that the index persons were balanced for sex within each stress area. Comparison

TABLE 9. TOTAL FAMILY SET SAMPLE. NUMBER OF FAMILY SETS BY STRESS AREA, SEX OF INDEX PERSON, NUMBER, AND MEAN AGE OF SET MEMBERS—RESULTS OF NURSE-INTERVIEWING, OCTOBER, 1968-JUNE, 1969

Family set members	Black						White						No. of persons interviewed						
	High stress			Low stress			High stress			Low stress									
	Family set when Index person is:			Family set when Index person is:			Family set when Index person is:			Family set when Index person is:									
	Male	Female	Mean age (yr)	Male	Female	Mean age (yr)	Male	Female	Mean age (yr)	Male	Female	Mean age (yr)		Male	Female	Mean age (yr)			
(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	
(57)	(55)	41	(61)	39	(56)	40	(55)	40	(62)	41	(58)	43	(57)	44	(57)	44	(57)	461	
Spouses of index person	(57)	(55)	45	(61)	36	(56)	44	(55)	37	(62)	45	(58)	40	(57)	47	(57)	47	(57)	461
Siblings of index person	(57)	(55)	41	(61)	40	(56)	41	(55)	39	(62)	40	(58)	43	(57)	44	(57)	44	(57)	461
First cousin of index person	(57)	(55)	41	(61)	40	(56)	42	(55)	39	(62)	41	(58)	43	(57)	45	(57)	45	(57)	461
Unrelated persons*	(57)	(55)	41	(61)	37	(56)	39	(55)	39	(62)	41	(58)	43	(57)	44	(57)	44	(57)	461
Total persons	(285)	(275)	(305)	(280)	(275)	(310)	(290)	(285)	(2305)										
Number of Family Sets	57	55	61	56	55	62	58	57	461										

*The unrelated person is matched to the index case by sex, age, race, marital status, census area, and possession of a Family Set.

of the mean ages of each set member for each index group shows that for all paired comparisons in each column, the range of mean age differences is not larger than ± 5 yr and the average is about four. Another analysis, not shown, indicates that 95% of all index-sib pairs were within ± 6 yr; as were 98% of all index-unrelateds, 89% of all index-first cousins, and 79% of index-spouse. This last pairing varied by high and low stress areas—approx 70% of high stress or working class marital pairs met the criterion of 6-yr age differences, but about 88% was found in the middle-class areas. The overall results suggest that the experimental control of age differences was fairly successful; however, data not shown suggest that within certain subsets of 5-yr age groups, significant pair differences do appear and require statistical adjustments in analysis.

The design also allowed for the sex of the sib or first cousin to vary. It was held open for testing whether biases in ascertaining sibs and cousins would result in unequal sex patterns. In our procedure, this would not be expected for sibs because they were selected as being *closest* in age to the Index, and results not shown reveal that the sex of sibs for each sex-race-area index group was therefore almost 50% male and female in each group. Procedures for ascertaining first cousins, however, allowed the respondent to report two of each of four parent-sib lines, (e.g. father's brother, mother's brother, etc.) who were closest in age. Results (not shown) indicate that for the total, and for each race-area index and unrelated person, their first cousins were also approximately 50% male and female; however, within each parent-sib line, there was a distinct bias, in general, for reporting same-sex first cousins. This bias was least for high stress females, and most for white males and black low stress females. Nevertheless, Table 10 views this matter in terms of family sets and suggests that, in general, expected chance patterns seem to prevail. The results would be even stronger except for the low number of Pattern No. 5 among blacks, which could be accounted for, though not explained, by the refusal bias of the black low stress females. Finally, Patterns Nos. 6 and 8 can be used as family sets which are experimentally controlled for sex, excluding the spouse.

Ideally, not only the sex of the siblings and cousins would be controlled, but also their marital and socioeconomic status. Again, differences in the latter variables can be adjusted statistically in the analyses. However, certain patterns of differences on such traits can be described and should be expected [19]. First, for marital status, sibs and cousins of blacks were three times more often 'non-married' (separated, divorced, single, widowed) than whites. This is partly due to the high per cent of Catholic families in the two white stress areas, but it also reflects ethnic differences in such status in the United States. Thus, for blacks, while all the index persons are married and living with a spouse, about one third of their sibs and cousins in the family set sample were non-married, compared to about 10% on the average for whites. Next, differences in education and family income varied by race-stress area groups in the following ways (data not shown): First, for all groups there were no differences in average educational and family income levels between index and unrelateds, indicating that matching the unrelated to the index on five other criteria already mentioned carried over to socioeconomic status (hereafter called SES level). Second, for black high stress family sets, there were no significant differences among the five set members on education and family

TABLE 10. CHI-SQUARE TEST OF PATTERNS OF SEX DISTRIBUTIONS WITHIN FAMILY SETS VS EXPECTED FREQUENCIES FOR THE GENETIC SAMPLE, BY RACE*

Pattern No.	Sex patterns†			Distribution of Family Sets					
	Index	Sib	Cousin	Total (%)	(N)	Blacks (%)	(N)	Whites (%)	(N)
1	M	M	M	18	(84)	19	(43)	18	(41)
2	M	M	F	9	(43)	8	(17)	11	(26)
3	M	F	M	11	(49)	10	(23)	11	(26)
4	M	F	F	12	(55)	15	(35)	9	(20)
5	F	M	M	9	(40)	5	(12)	12	(28)
6	F	M	F	15	(71)	17	(39)	14	(32)
7	F	F	M	12	(54)	14	(32)	9	(22)
8	F	F	F	14	(65)	12	(28)	16	(37)
				100%	(461)	100%	(229)	100%	(232)
				$\chi^2 = 12.9$		$\chi^2 = 15.8$		$\chi^2 = 6.1$	
				$P = 0.07$		$P = 0.03$		$P = 0.52$	

*The null hypothesis is that the number of sets in each pattern should be equal to those generated by probability, i.e. one-eighth or 12%. Spouses are always the opposite sex from the Index, and Unrelateds are always the same sex of Index; Sib and First Cousin varied. It can be seen that Pattern No. 5 among blacks was far less than the expected N of 29 or 12%.

†M = Male; F = Female.

income. For black low stress index persons, however, whose SES levels were among the top 5% of all 382 census tracts in the City of Detroit, and in the top 1% of all black census tracts, the education and income of their sibs and cousins were predictably significantly lower. Third, for whites, the converse pattern held for index persons in the high stress working-class groups whose sibs and cousins were, generally, higher in SES status. There were no patterns for the white low stress areas, except the trend appeared to be that sibs and cousins who lived in Detroit suburbs tended to be higher in family income. Finally, regarding geographic distribution, white relatives tended to live outside the City in suburban locales while black relatives tended to reside within the segregated areas of Detroit. This again is an expected finding [20].

Table 11 shows the average number of sibs and types of first cousins reported by family set index and unrelated persons in each stress area. The high average number of sibs per respondent partially results because the sample design requires the existence of one or more siblings in order to be an index or unrelated person. We will return to this constraint on family set analyses later.

Because of the prevalence of sibs and cousins for family set sampling, crude estimates of yield may be made from United States Census data using Dwelling Units as the unit of analysis. Table 12 shows the funneling process for estimating yield of family sets and the comparison of percents between the estimates and actual results from two studies. Line B1 shows that an estimate of 25% was used for the Major Study data (i.e. that 25% of the DU's with married heads of the majority race in an area, ages 25-59, would yield potential family sets). Line C1b shows that the actual average yield across the four stress areas was 28%, indicating that the original estimator was fairly accurate. In the Pilot Study, an estimator of 33% was used. Line C1a shows that the actual average yield across the two Pilot stress areas was only 19%. Thus, we adjusted the estimator downward for

TABLE 11. MEANS FOR NUMBER OF SIBS AND OF FIRST COUSINS BY AREA (GENETIC AREA SAMPLE)

Family relations	Stress area									
	Total		Black		White					
	(\bar{X})	(<i>N</i>)	High stress (\bar{X})	Low stress (<i>N</i>)	High stress (\bar{X})	Low stress (<i>N</i>)	High stress (\bar{X})	Low stress (<i>N</i>)		
A. Siblings										
1. Total Sibs	4.15	(922)	4.31	(224)	4.50	(234)	4.27	(234)	3.50	(230)
2. Sibs in Detroit	2.98	(922)	2.93	(224)	3.25	(234)	2.96	(234)	2.77	(230)
3. Total sibs ± 10 yr in Detroit	2.31	(922)	2.16	(224)	2.48	(234)	2.30	(234)	2.28	(230)
B. First cousins as children of:										
1. Father Brother	3.89	(922)	4.08	(224)	4.71	(234)	3.56	(234)	3.20	(230)
2. Father Sister	3.24	(922)	2.97	(224)	3.39	(234)	3.76	(234)	2.83	(230)
3. Mother Brother	4.04	(922)	4.61	(224)	4.24	(234)	3.97	(234)	3.36	(230)
4. Mother Sister	5.19	(922)	4.29	(224)	4.79	(234)	4.36	(234)	3.32	(230)
Total first cousins	15.15	(922)	15.48	(224)	16.95	(234)	15.44	(234)	12.71	(230)
C. First cousins in Detroit as children of:										
1. Father Brother	1.45	(922)	1.17	(224)	2.02	(234)	1.09	(234)	1.50	(230)
2. Father Sister	1.01	(922)	0.69	(224)	1.09	(234)	0.73	(234)	1.52	(230)
3. Mother Brother	1.48	(922)	1.70	(224)	1.45	(234)	1.07	(234)	1.74	(230)
4. Mother Sister	1.53	(922)	1.27	(224)	1.58	(234)	1.54	(234)	1.74	(230)
Total first cousins in Detroit	5.50	(922)	4.87	(224)	6.15	(234)	4.45	(234)	6.51	(230)
D. Total first cousins ± 10 yr in Detroit										
	2.66	(922)	2.01	(224)	2.69	(234)	2.49	(234)	3.43	(230)

the Major Study. The accuracy of estimation will vary with cultural and socio-demographic facts of life of any given target population; nevertheless the data support the view that the yield can be roughly estimated in the design stage. The dynamic sampling frame allows for monitoring of desired numbers.

DISCUSSION

The results of the pilot and this major study sampling effort are encouraging in confirming the feasibility of ascertaining and obtaining data about family sets in an urban area. Internal reliability tests are also encouraging, e.g. no differences on key variables among set members when treated as individuals, a random sex pattern across family sets, random sex patterns of sibs and first cousins, and no differences on key variables among randomly selected index persons regardless of their closest age sibs and cousins. External validity tests on such traits as age, weight, and blood pressure to be presented in the next article also provide supporting evidence that family set members treated as individuals conform to empirical norms in other populations of similar demographic traits. This resemblance holds not only for physical traits but for expected rates of social, economic and attitudinal behavior, e.g. higher separation and divorce rates for blacks than whites, working and middle class, and higher percent of working wives for blacks than for whites.

Basic to the feasibility of sampling family sets is the evidence in urban literature suggesting that active family networks have a high prevalence in urban society.

TABLE 12. ESTIMATE AND ACTUAL PER CENT OF DWELLING UNITS (DU's) HAVING A POTENTIAL FAMILY SET FROM CENSUS AND VERIFICATION: DETROIT, 1968-69

	Total		Stress area							
	(%)	(N)	Black		White					
			High stress (%)	Low stress (%)	High stress (%)	Low stress (%)				
			(N)	(N)	(N)	(N)				
A. Total number of DU's	100%	(12,294)	100%	(4379)	100%	(1952)	100%	(4183)	100%	(1780)
1. DU's with head married, and of given race	47	(5726)	34	(1482)	50	(977)	49	(2052)	68	(1215)
2. DU's with head married, of given race, ages 25-59	34	(4200)	25	(1075)	46	(895)	32	(1366)	49	(864)
B. Estimate of DU's having potential family sets										
1. As % of DU's with head married, of given race, ages 25-59	25	(1050)	25	(269)	25	(224)	25	(341)	25	(216)
2. As % of total DU's	9	(1050)	6	(269)	12	(224)	8	(341)	12	(216)
C. Actual DU's with verified potentials*										
1. As % of DU's with head married, of given race, ages 25-59:										
(a) (Pilot study)†	(19)		(20)		(18)					
(b) Major study‡	28	(1188)	22	(235)	32	(282)	20	(275)	46	(396)
2. As % of total DU's:										
(a) (Pilot study)	(8)		(7)		(9)					
(b) Major study	10	(1188)	5	(235)	14	(282)	7	(275)	22	(396)

*We chose the verified potential (with sib and first cousin, ± 6 yr of age) as the sampling stage to estimate, because it seemed closest to the arbitrary point of where the population list of family sets first appears.

†The pilot study population (1966) was black only and in different census areas than those of the major study.

‡Pilot study data were collected from single census tracts per area, as were major study data for the low stress areas. High stress areas in the major study were composed of four contiguous tracts (two partially censused).

Axelrod's early work [21] and subsequent studies support the conclusion that kinship visiting, for example, is a primary social activity among urban dwellers and outranks such activities with friends, neighbors and co-workers. More recent inquiry in more detail confirms this finding in a smaller city; in a white, married, young and middle-aged population, yearly contact at least several times was reported to be over 90% with parents and age-near siblings and over 60% with 'best-known' cousins [22]. Sussman and Burchinal conclude in a review of the 'outer extensions' of American urban families that 'There exists an American kin family system with complicated matrices of aid and service activities which link together the component units into a functioning network. The network... is composed of nuclear [family] units related by blood and affinal ties. Relations extend

along generational lines and bilaterally where structures take the form of sibling bonds and ambilineages, i.e. the family circle or cousin club.' [23]. There is evidence that the destination of migrants is heavily influenced by existence of kin in the place of destination [24]. From a socio-biological view, normal reproduction cycles of age-graded cohorts provides the population base for a high incidence of siblings and first cousins in urban areas [25].

Limitations in the family set sample

We can now explore certain limitations inherent in the family set approach. It has already been mentioned that the population of index persons must have at least one or more siblings and one or more first cousins; the sampled families will be larger than 'average.' This will restrict interpretation and constrain the kinds of problems fitting to the model. One can, however, statistically control on parity. Furthermore, new techniques, to be mentioned later in this series, will allow for sampling family sets with multiple and uneven numbers of first and second degree relatives.

A second constraint arises from comparison of mean ages of persons at each stage of the sampling process with the final sample. Table 13 presents means and *t*-test comparisons between the final Genetic Area Sample (index and unrelateds) and the other persons present at each stage of the sampling flow. The first comparison between the age and education means of the final sample and others at the point of census shows a steady bias, i.e. the family set members

TABLE 13. *T*-TEST RESULTS COMPARING MEANS OF AGE AND EDUCATION LEVELS BETWEEN THE SAMPLING STAGES AND THE FINAL GENETIC AREA SAMPLE BY AREA (AGE RANGE 30-54 yr)

	Black		White	
	High stress	Low stress	High stress	Low stress
I. Genetic area sample	Final genetic area sample†			
Age	40.6	39.8	40.7	43.4
Education	10.8	12.8	10.3	11.8
A. vs others in census				
Age	42.1*	40.9*	42.7*	44.6*
Education	10.4	12.8	9.6*	11.9
B. vs others assigned to be verified				
Age	40.7	39.7	42.4*	44.1
Education	10.7	12.6	9.7*	11.7
C. vs other verified potentials				
Age	40.2	40.1	41.7	43.0
Education	10.4	12.5	10.0	11.8
D. vs others assigned to nurse-interviewer				
Age	40.2	41.8*	42.8*	43.0
Education	10.4	12.5	10.0	11.6
E. vs other index/unrelated with partial sets				
Age	40.9	43.0	46.3*	42.6
Education	10.7	13.0	9.2	11.7

†The \bar{X} 's of age and education level in the genetic area sample are compared to \bar{X} 's in A, B, C, D, and E.

An asterisk () means $P < 0.05$. All other comparisons are not significant.

are younger and more educated. From results (not shown) of frequency tables and chi-square tests in tables using the full range of both age (in 5-yr intervals) and education levels, there is a uniform pattern for all the comparisons shown in Table 13 that persons 50–54 yr of age are significantly more often excluded from each stage of sampling and persons 20–34 yr of age are more often included. This can reasonably be interpreted as due to differential mortality of aged sibs and first cousins compared to younger families. There is also some evidence in kinship studies that the highest interaction between family members in terms of visiting, get-togethers, financial aid, gifts, and services occurs between younger, marrieds with first child and younger children than at other stages of the life-cycle [26]. These two major facts converge to increase the prevalence of family sets among 25–50 yr old age groups. The upper age limit of adequate yields for sampling family sets in our population appears sharply at 50. Such results again are relative to health-related cultural and socio-demographic parameters of target populations.

Finally it must be noted that there is a cost in obtaining lists of potential sample members. Obviously, in most populations, such lists are not available, nor current, much less complete, except in special circumstances where family registers are part of official records. Our experience in screening for index persons actually began in The University of Michigan Arthritis Clinic where with a minimal effort, we obtained 21 persons having diagnosed rheumatoid arthritis who also had a full family set [7]. Unfortunately these data were never analyzed for genetic validity by using the variable of height. In the pilot and major studies, a census of dwelling units began the sampling process. In small residential areas, this is a relatively rapid and low-cost operation. In larger populations, such as a sample of a city, state, or nation, the screening can be done by inclusion of necessary screening items in the context of another study sample. There are a small number of continuous local surveys and large-sample national studies now on-going in the United States within which screening can be accomplished, e.g. local health monitoring surveys and national health and economic surveys. It is important in this multi-stage sampling process that adequate machinery for data management be available for monitoring the sample flow during data collection; ideally this would include computer programs designed for storing, up-dating, partitioning, analyzing, and describing a dynamic process (e.g. Ref. [27]). A complete description of the procedures used to execute the present study has been documented [15].)

The remaining articles in this series will present further results of the family set method as we explore the problem of estimating an hereditary component of blood pressure among blacks and whites in Detroit.

Acknowledgements—A project as large as this must necessarily draw on the knowledge and skills of many agencies and persons. Acknowledgment of support to all is not possible: therefore the following list is a partial one.

Initial support in the form of seed money from the Office of Research Administration of The University of Michigan was critical in allowing this project to start. Our special thanks to Dr. Rudolph Schmerl, Dr. Samuel Fox's encouragement, funds from the Center for Chronic Disease Control, U.S. Public Health Service, and Mr. Robert Thorner's administrative skills within this Center, made the pilot survey viable.

Local resources of The University of Michigan in Ann Arbor have also been utilized in various ways with an equally high degree of cooperation. Informal consultations with a core group of authorities in ecology, biostatistics, hypertension, survey methods, sampling, and human genetics were utilized

for the duration of the Project. The computer facilities of The University of Michigan, the Institute for Social Research, and Wayne State University, including program consultation, have been used extensively.

Active consultants to this Project include: Dr. Jeremiah Stamler, Dr. Sibley W. Hoobler, and Dr. Theodore M. Newcomb. Other persons informally consulted were: Dr. Fredrick Epstein, Dr. Otis D. Duncan, Dr. J. E. Keith Smith, and Dr. John Scott. They each have contributed much to this Project. Our special thanks to Dr. Stevo Julius, Dr. Robert Smith, Dr. Adrian Ostfeld, and Dr. John Cassel.

Cooperation from local agencies in Detroit was superb. The Visiting Nurses Association and the City Department of Health were most helpful in recruitment, and in solving minor problems. The City of Detroit, Department of Police, furnished through its statistical units, all the crime and delinquency records requested, in an efficient and most pleasant manner. The Mayor's Committee for Community Renewal Program was most helpful in every request for data, maps, etc. made by Project personnel. Wayne State University campus facilities and graduate students' aid were arranged by the Department of Sociology and Anthropology, which also allowed use of a field office in the department, and appointed the Project Director an Adjunct Professor. Wayne State University provided facilities for training purposes, and facilitated administrative matters. Other agencies aiding the Project were the City of Detroit Planning Commission and the Metropolitan Services agency. Many persons in Detroit were consulted on a variety of problems, and all gave freely of their time and ideas.

The public health nurses who actually collected the data through snow, rain and High Stress areas did an excellent job, guided by two superb nurse-supervisors: Mrs. Mildred Harvin (Pilot Study) and Mrs. Revera Munce (Major Study).

Our Detroit Advisory Committee must be thanked for their time, helpful criticisms and support: Dr. Mel Ravitz, Dr. Francis Kornegay, Dr. John Caldwell, Miss Sylvia Peabody, Dr. Leonard Moss, Dr. Ross Stagner, Dr. Robert Smock, Mr. Homer Hall, and Dr. Milton Palmer.

Appreciation must also be expressed to Mr. William Ash, the mathematician-programmer who served this project well at its inception, and to Mr. Eugene Beauregard for initial direction of field work in the Pilot Study.

We must also express appreciation for the competent administrative duties during data collection performed by our office staff in Detroit: Lorraine Ward, Field Director; Ann Stark, Field Coordinator; Fredericka Harvey, Nurse Assignment Coordinator; Bettjean Harris and Dorrie Shaefer, Field Office Assistants. And in Ann Arbor: Karen Donahue, Administrative Assistant; and Heidi Bergemann, Patricia Strauch, Eileen Kavenaugh and Robert Colman, Research Assistants. During data analyses Thomas Smedes, Study Director, served the Project in managing the complex data files.

REFERENCES

1. Stamler J, Epstein FH: Coronary heart disease: Risk factors as guides to preventive action. *Prev Med* 1: 27-48, 1972
2. The Framingham Study. An epidemiological investigation of cardiovascular disease. Kannel WB, Gordon T (Eds.), Washington, DC: US Government Printing Office. Section 27, December 1970
3. Epstein FH, Francis T Jr, Haynes NS, Johnson BC *et al.*: Prevalence of chronic diseases and distribution of selected physiologic variables in a total community, Tecumseh, Michigan. *Am J Epid* 81: 307-322, 1965
4. Cassel JC (Ed.): Evans County Cardiovascular and cerebrovascular epidemiologic study. *Arch Int Med* 128: 883-986, 1971
5. Clark VA, Chapman JM, Coulson AH: Effects of various factors on systolic and diastolic blood pressure in the Los Angeles Heart Study. *J Chron Dis* 20: 571-581, 1967
6. Keys A, Taylor HL, Blackburn H *et al.*: Mortality and coronary heart disease among men studied for 23 years. *Arch Int Med* 128: 201-214, 1971
7. Cobb H, Schull WJ, Harburg E, Kasl S: The intrafamilial transmission of rheumatoid arthritis—VII. Summary of findings. *J Chron Dis* 22: 295-296, 1969
8. Harburg E, Schull WJ, Erfurt JC *et al.*: 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, 1966-67. *J Chron Dis* 23: 69-81, 1970
9. Schull WJ, Harburg E, Erfurt JC, Schork MA, Rice R: 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, 1970
10. Harburg E, Erfurt JC, Chape C, Hauenstein LS, Schull WJ, Schork MA: Socioecological stressor areas and black-white blood pressure. Detroit. *J Chron Dis* 26: 595-511, 1973

11. James SA, Kleinbaum DG: Socioecologic stress and hypertension related mortality rates in North Carolina. *Am J Pub Hlth* 66: 354-358, 1976
12. Cassel J: Appraisal and implications for theoretical development. In social stress and cardiovascular disease. *Milbank Memorial Fund Quart* Syme SL, Reeder LG (Eds.) 45 (2), 1967
13. Kasl SV, Harburg E: Mental health and the urban environment: Some doubts and second thoughts. *J Hlth Soc Behav* 16: 268-282, 1975
14. Harburg E, Erfurt JC, Hauenstein LS *et al.*: Socio-ecological stress, suppressed hostility, skin color and black-white male blood pressure: Detroit. *Psychosom Med* 35: 276-296, 1973
15. Donahue KM, Erfurt JC, Harburg E, Strauch PA, Bergemann H: Procedures for sampling, verifying, and interviewing members of family sets in a study of the roles of stress and heredity in black-white blood pressure differences. Ann Arbor, Michigan, Program for Urban Health Research, University of Michigan, 1974
16. Erfurt J, Harburg E, Rice R: A method for selection of census tract areas differing in ecological stress. Program for Urban Health Research, 1970
17. Harburg E, Schork MA, Erfurt JC, Schull WJ, Chape CA: Heredity, stress and blood pressure, a family set method—II. Results of blood pressure measurement. *J Chron Dis* 30: 649-658, 1977
18. John Scott, Director, Field Section, Survey Research Center, University of Michigan, Personal communication, 1975
19. U.S. Department of Labor: Social and Economic Status of Negroes in the United States—1970. Washington, DC: US Government Printing Office, Series P-23, No. 38, 1970
20. Taeuber KE, Taeuber AF: *Negroes in Cities*. Chicago: Aldine, 1975
21. Axelrod M: Urban structure and social participation. *Am Soc Rev* 21: 13-18, 1956
22. Adams BN: *Kinship in an Urban Setting*. Chicago; Markham, 1968
23. Sussman MB, Burchinal L: Kin family network: Unheralded structure in current conceptualizations of family functioning. In: *Sourcebook in Marriage and the Family*, Sussman B (Ed). Boston: Houghton Mifflin 1968
24. Tilly C, Brown CH: On uprooting, kinship and the auspices of migration. *Int J Comp Soc* 8: 139-164, 1967
25. Hajnal J: Concepts of random mating and the frequency of consanguineous marriages. *Roy Soc Proc Ser B* 159: 1963-64
26. Aiken MT: Kinship in an Urban Community, PhD Dissertation, University of Michigan, Department of Sociology, 1964
27. Brill RC: The Taxir Primer, MTS Version, University of Michigan Taxir Users Group, Center for Human Growth and Development, May 1974