Familial Aggregation of Cancer in Laredo, Texas: A Generally Low-Risk Mexican-American Population

Kenneth M. Weiss, Ranajit Chakraborty, Peter E. Smouse, Anne V. Buchanan, and Louise C. Strong

Center for Demographic and Population Genetics, University of Texas Graduate School of Biomedical Sciences, Houston (K.M.W., R.C., A.V.B.); Department of Human Genetics, University of Michigan Medical School, Ann Arbor (P.E.S.); Department of Experimental Pediatrics, University of Texas M.D. Anderson Hospital and Tumor Institute, Houston (L.C.S.)

Genealogies for the Mexican-American city of Laredo, Texas, have been assembled by computer from individual civil and church records of birth, marriage, and death. Documentation is available on vital events in the lives of over 300,000 individuals, about 80% of the city population from 1870–1981. These data were collected to determine the degree to which death from cancer is more clustered in families than would be expected by chance alone; methods specific to this data base have been developed to accomplish this task. A statistically significant excess of familial cancer was observed overall when all cancer sites were pooled, but no evidence was observed for excess familial risk at single sites except for breast cancer and perhaps for ovarian cancer. The excess of breast cancer risk is comparable to that observed in other populations. A few site-combinations manifest excess familial risk, most notably those involving and dominated by breast cancer and certain digestive system sites. We do not confirm the degree of familiality observed elsewhere for cancers of the lung, colorectum, stomach, or other sites in this generally low-risk population. Even where we find evidence of excess risk, the degree of excess is small and the number of multiply affected families too small to test etiologic models by segregation analysis. The absence of excess familial risk does not appear to be due to inadequate numbers of cases, since breast cancer is familial with no more occurrences in Laredo than other sites. These results differ to some extent from those found in a similar study of

Utah Mormons, but it is unclear whether this is because of differences in risk patterns or statistical properties of the analytic methods used in the two studies.

INTRODUCTION

It is said that at least 80% of all cases of cancer are attributable to environmental factors and hence could be prevented with proper avoidance of risk factors extrinsic to the individual [Doll and Peto, 1981]. There are some notable exceptions, cancers that appear with elevated frequency in particular families and cancers known to have a hereditary basis. Cancers known to be simply inherited include some tumors of childhood, such as retinoblastoma, and some Wilms tumors, as well as cancers of adult onset, most notably colon cancer associated with familial polyposis, multiple neoplasia syndromes, and perhaps some breast cancer. In rare instances, specific genes or chromosome regions have been identified in support of the epidemiologic evidence that the tumor has a genetic basis requiring variable environmental contributions [Schimke, 1978; Strong, 1982; Swift, 1982].

For many other cancers, there exist reports of a modest elevation of familial risks in some populations, with relative risks on the order of 2.0 [eg, Anderson, 1982; King et al, 1984; Lipkin et al, 1980; Schimke, 1978]. Some studies showing such an excess risk have been based on hospital or clinic ascertainment, allowing a possibility that selection bias is responsible for the observed excess. Population-based studies have shown familial risk [Jacobsen, 1946; Macklin, 1959; Woolf, 1955], but in one of the first population based studies very little concordance was found among Danish twins, suggesting that genetic factors were not generally important in cancer epidemiology [Harvald and Hauge, 1963].

In the only whole-population study of familial cancer to date, genealogies for the Mormon population of Utah were computerized and linked to death certificates. From these genealogies, the genetic relationships between cases of cancer have been compared to the relationships that would be expected among randomly chosen age-, sex-, and cohort-matched individuals from these same genealogies [Hill, 1980a,b; Skolnick et al, 1981; Bishop and Skolnick, 1984]. The expectations were derived from simulations, using the genealogies as a reference frame. For single cancer sites and certain sets of sites, there was an excess of cases among closely related individuals compared to chance expectation.

We report here a study similar in nature to the Utah effort; genealogies from the Mexican-American city of Laredo, Texas, have been assembled from church records and linked to all death certificates from 1900 to the present [Buchanan et al, 1984; Schwartz et al, 1984]. We have assessed the degree of familial aggregation of death from cancer based on the total death experience of the population, which does not require simulation. Mexican-Americans have characteristic cancer patterns, and here we investigate whether there are differences in the amount or nature of familial aggregation. For comparative purposes, we have tested most of the same cancer site pairs reported in the Utah studies. The epidemiology of breast cancer in Laredo, including its familiality, has been discussed in detail elsewhere [Buchanan et al, 1985].

MATERIALS AND METHODS

The city of Laredo lies on the Rio Grande River at the Texas-Mexico border and, with the exception of the Mexican sister city of Nuevo Laredo, is surrounded by

arid rangeland for 100 miles in all directions; geographic isolation makes Laredo a good sampling unit for epidemiologic study. For much of its history, Laredo was part of Mexico and has remained ethnically Mexican-American and predominantly Roman Catholic. Records of the births, marriages, and deaths of most Laredoans have been maintained by the Church since the founding of the town in 1785. Some 170,000 baptism, 99,000 marriage, and all 54,000 city death records from Laredo were photocopied and computerized, and these records imply the existence of another 130,000 individuals who have lived, but were not born, in the city and who may have died there. The death records extend from the beginning of mandatory registration in about 1900 through 1981. The data base includes records on approximately 75–80% of current residents of the city [Buchanan et al, 1984].

The effects of the twin city of Nuevo Laredo, across the Rio Grande from Laredo, and of the 20% of the population not represented are, we think, small and not of a biasing nature. It is not likely that individuals preferentially went across the river to die, undocumented, of cancer. While Mexicans do come to the medical facility in Laredo on some occasions, our study is based on those represented in the genealogies and hence is not affected by deaths of others. Similarly, while the study reported here is of the Roman Catholic population of Laredo, there is no reason to think that the data are unrepresentative of the general population.

Causes of death are given in variable number and format, depending on the civil record forms in use at any given time. For recent decades, primary and contributory causes are differentiated. Deaths were coded according to the specifications and format of the ICD-9 [WHO, 1978]. We have lists of all of the certifying physicians and for recent ones can identify a specialty (if applicable). However, in general, the number of signing physicians at any one time has been small (less than 10 for 1900–1930, less than 30 subsequently). It is not possible to judge their accuracy, except that the age patterns and cause distribution are consistent with what we know of the expected epidemiology in the area [eg, Badzioch, 1982]. This provides at least indirect confirmation of the reliability of the records. Autopsy verification has not been done to our knowledge.

Genealogies have been assembled from these 450,000 or so records by a complex, interactive-automated software system developed by us [Buchanan et al, 1984; Schwartz et al, 1984; also available as technical reports from the first author (K.M.W.)]. At all stages of the computerization and record-linkage, extensive error checking and verification have been done to ensure that the resultant genealogies are as error-free as possible. The genealogies represent the population from the late 1800s up to about 1980. The linked death records provide an essentially complete mapping onto the genealogies of deaths occurring in Laredo during this century; that is, virtually all Laredoans in our genealogies who died in the city have their deaths linked to the genealogies. The structure of the genealogies is indicated in Tables I and II, which give the distribution of sibship sizes available in the population and the distribution of the sizes of the genealogical structures making up the data base. Table I shows that there are numerous large sibships in which to observe multiple occurrences of disease should they occur. Table II shows that there are a large number of sizeable genealogies, providing ample opportunity to observe multiple affected relatives of probands.

These genealogical structures include all individuals we could identify who were related, by descent or marriage, to any starting individual. All individuals in the data

TABLE I. Sibship Size Distribution in Laredo Genealogies

Size	Frequency	Size	Frequency	Size	Frequency
1	51,863	7	1,195	13	54
2	12,297	8	742	14	16
3	6,601	9	526	15	16
4	4,027	10	329	16	8
5	2,586	11	177	17	3
6	1,808	12	109		
Total					82,357

TABLE II. Disjoint Distribution of Genealogy Structure Sizes

Genealogy structure size (unique indiv.)	No. of such structures	Total individuals involved
3	28,789	86,367
4	7,568	30,272
5	2,972	14,860
6	3,933	23,598
7	1,547	10,829
8	992	7,936
9	604	5,436
10	407	4,070
11–20	1,085	14,735
21–30	136	3,289
31–81	63	2,671
98,405	1	98,405
Total	48,097	302,468

base were used as index cases unless they had been included previously in a genealogical structure; hence in the structures of Table II, no individual is counted more than once (ie, the table gives the distribution of exhaustive, disjoint structure sizes). Many of the "families" consist only of a single individual and his/her parents; these represent incomplete families, unlinkable data, and families who either immigrated to or emigrated from Laredo. This is a sizeable number of incomplete families; on the other hand, Table II shows that one genealogical structure contains about 100,000 individuals, or about one-third of the city's residents over the past century. Although most of these individuals are related only by a multiplicity of marriage links (ie, have no genetic relationship), most members of the genealogical structure will have a substantial number of biological relatives in the structure. (One might consider how difficult it would be to identify as many connected individuals over a century's time in other places solely from birth and marriage records in a single population.)

The processes by which the records were linked or by which the assembled genealogies are ascertained do not depend on the individual with whom one starts. Hence nothing in the analysis depends on the order of identification of probands or family members. Because the analysis is restricted to those individuals lineally related to each other, the terms *family* and *genealogy* will henceforth refer only to nuclear or extended sets of lineal relatives but not to whole genealogical structures containing affinal relatives as well.

We cannot confirm paternity directly in these data, and so we have no way to assess the frequency with which incorrectly attributed paternity obscured true family connections. However, preliminary results from a study in a nearby community, which is virtually the same ethnically, indicate that the rate of exclusions based on genetic markers is only a few percent. It should also be noted that such errors will diminish rather than augment the apparent evidence for familial susceptibility and that most if not all comparative studies in the literature have unconfirmed parentage.

To assess the epidemiological characteristics of cancer in Laredo, we have used the U.S. census of Laredo for every decennial period from 1900 to the present. From these and the death records, mortality statistics have been computed to compare to those of the U.S. population at large and to find secular trends in cancer mortality.

To evaluate the familiality of cancer, we have developed our own measure based on a complete enumeration of all the person-years lived by members of the genealogies and the occurrence of cancer death among those same individuals. This represents the entire exposure history of the genealogical population. The entire exposure history of the Laredo genealogies was tabulated according to sex, age, and period of birth. The number of person-years lived by each individual was tabulated and grouped according to age in categories 0–14, 15–39, 40–59, and 60+, chosen because they relate to life-stages and can separate early-, middle-, and late-onset disease. To account for possible secular trends in occurrence or diagnosis of disease, person-years were further tabulated according to sex and to the time period when they occurred using the intervals 1870–1899, 1900–1929, 1930–1959, and 1960+, chosen largely for about equal length and for convenience and also (in regard to pre- and post-1930 periods) roughly to divide the pre- and postantibiotic eras.

The tabulation thus consists of four age groups, four time periods, and two sexes, or a breakdown of the entire Laredo experience into 32 strata. When sexes were pooled, there were a total of 16 strata. The numbers of person-years in these strata are given in Table III. The total experience is 1,064,234 male years and 1,032,734 female years (2,096,968 person-years overall).

TABLE III. Distribution of Total Person-Years (pyr) of Experience in the Laredo Data Base*

Age and sex group	Historical time period							
	Before 1900		1900–1929		1930–1959		After 1960	
	pers	pyr	pers	pyr	pers	pyr	pers	pyr
0–14								
M	7,563	75,138	16,857	142,709	17,163	155,101	4,691	23,442
F	7,328	74,417	16,409	142,537	17,092	158,808	4,960	27,876
15–39								
M	3,195	42,804	11,328	149,808	16,443	192,691	9,762	82,438
F	3,245	40,301	11,443	144,374	16,036	175,012	9,908	78,173
40–59								
M	729	8,009	3,494	34,672	6,764	74,938	3,057	18,426
F	569	5,888	3,393	34,836	5,825	67,119	2,201	12,805
60+								
M	149	1,448	943	8,437	3,165	29,127	2,712	25,046
F	96	775	858	8,034	3,313	35,282	2,675	26,497

*Note that in each cell the number of persons is given first followed by the number of person-years lived by these persons in the cell.

For each cancer site tested, the sex and age and time period when each case died were tabulated; that is, all cancer deaths were partitioned into these 32 cells. Using the person-years already tabulated and the numbers of deaths, the sex-, age-, and time-specific death rates (deaths per person-year) were computed for each cell. These are the empirical rates for Laredo, involving a complete enumeration, and they form the basis of all comparisons.

For each site, all cases were ascertained sequentially from the genealogical data base and used as probands for a family analysis. All family members of the proband were identified, and their person-year exposure history was tabulated, just as was done for the total population. Because all cases were probands, if there were n cases, the experience of n families (one for each proband) was tabulated. Thus multiplex families (with more than one affected individual) will be represented as often as they have cases. This is a form of complete ascertainment and yields the correct expected values for the probabilities of disease in the tabulated cells under the null hypothesis of no familiality, as we show in the Appendix.

The initial analysis of the patterns of familial cancer was done by tallying the first- and second-degree relatives of the probands. We also repeated this analysis, with all lineal relatives (the complete genealogy) of each proband taken into account. We report here the results of the first analysis only, because the results from the extended families were almost identical. This is consistent with theoretical expectations and with the findings in Utah [Hill, 1980a,b; Skolnick et al, 1981], where most excess risks were found to be concentrated in close relatives.

TESTS OF SIGNIFICANCE

The objective of this analysis is to determine whether there is evidence for excess cancer risk among relatives of affected individuals compared to risk in the population as a whole. Because the probands comprise all cases in the population and the genealogies represent the whole population, it was not necessary to select a control group. Rather, we determined whether the probability that an individual dies of the disease in question while passing through a particular cell is greater if that individual is related to an affected proband than if he/she is an average sex- and age-matched member of the population. To do this we have computed the exact upper tail probability (π) that at least as many as the observed number of deaths would occur by chance, with the observed number of entrants into each cell as the number of independent observations, and used the population probability of death as the true parameter under the null hypothesis of no familiality (see Appendix equation A5).

As a result, for each of the exposure cells, we compute the probability of death derived from the population and the probability of death derived from the relatives. This was done using life-table methods for the cell involved; we did not compute whole-life risks, because our study spans a century during which risks by age and sex changed greatly, and we wanted to be able to look for specific age groups, sex, or time periods in which excess risk might be manifest. In practice, only about four cells were informative for each sex, namely, those involving older individuals in recent decades. There are several reasons for this. First, cancer will occur only rarely in younger individuals, and sample sizes are not sufficient to detect excess. Second, the actual occurrence of cancer, and the ability to diagnose it, have been low until the latter two historic time periods. Finally, the sample sizes in the older age categories

are so small, in earlier time periods, that we could not expect to detect risk differences for site-specific cancers. Clearly, this data base extends well beyond the time periods for which detailed cancer epidemiological statistics can be developed for a single small population.

To determine the probability of observing the series of probabilities for a given site, we used the fact that

$$-2 \log_e \pi$$

is asymptotically distributed as a χ^2 with 2 df, where π is defined as above [Anderson and Bancroft, 1952]; thus minus twice the sum of the logarithms for the m cells for which we compute π has a χ^2 distribution with 2m degrees of freedom. This yields an approximate test of the overall significance of the results, approximate because the small sample sizes in the cells lead to an irregular rather than a continuous upper-tail probability distribution.

RESULTS
General Epidemiology

In Table IV we present, by site, the total number of malignant neoplasms recorded anywhere on the death certificates, including multiple cancers recorded in the same individual. The large number of site-unspecified diagnoses includes cases of unknown primary site as well as cases of stated primary site in addition to stated metastases or nonspecific/unspecified primary site. These latter deaths were often recorded as "carcinomatosis."

In Table V(A) we provide the age and time distributions of all cancer deaths in Laredo separately for males and females to show the nature of the available sample. Not all of these deaths were mapped onto the genealogies; such mapping depended on a person's having had some life event documented in the city, essentially having been a member of the Catholic Church, and, if not born or married in Laredo, having had a child born there. The year of death groups are chosen to span U.S. censal years

TABLE IV. Distribution of Malignant Neoplasms Recorded on Death Certificates in Laredo, 1895–1981*

Site (ICD-9 code)	No. of cases
Lip, oral cavity, and pharynx (140–149)	105
Digestive organs and peritoneum (150–159)	1,258
Respiratory and intrathoracic organs (160–165)	523
Bone, connective tissue, skin, and breast (170–175)	355
Genitourinary system (179–189)	789
Unspecified sites (190–199)	1,821
Lymphatic and hematopoietic tissue (200–208)	167
Total malignant neoplasms	5,018

*This table includes all deaths recorded in Laredo; a variable fraction of these were linked to individuals represented in the genealogies. Unspecified site category includes individuals certified only as "carcinomatosis" but does not include individuals for whom a primary site was also listed on death certificate.

TABLE V(A). Age and Year of Cancer Deaths Registered in Laredo, 1895–1981

Year of death	Age (yrs)									Total
	< 5	5–14	15–24	25–34	35–44	45–54	55–64	65–74	75+	
Males										
1895–1904	0	0	0	3	4	2	1	6	0	16
1905–1914	0	0	1	4	3	2	7	7	6	30
1915–1924	0	1	0	0	5	6	7	8	2	29
1925–1934	0	1	2	5	7	15	27	20	11	88
1935–1944	5	0	4	3	10	27	38	32	21	140
1945–1954	3	2	2	3	6	26	67	51	44	204
1955–1964	7	2	4	9	12	38	81	105	60	318
1965–1974	4	5	3	2	12	35	81	135	131	408
1975–1981	3	2	2	6	8	20	61	95	149	346
Total	22	13	18	35	67	171	370	459	424	1,579
Females										
1895–1904	0	0	1	1	6	16	10	6	2	42
1905–1914	0	1	0	2	13	20	13	9	3	61
1915–1924	1	1	0	4	9	19	20	9	6	69
1925–1934	1	1	4	13	32	42	45	22	17	177
1935–1944	0	4	4	7	27	43	51	33	20	189
1945–1954	3	3	6	11	32	56	73	73	42	299
1955–1964	0	1	3	9	26	53	88	94	70	344
1965–1974	1	2	1	10	24	55	83	135	105	416
1975–1981	1	2	2	6	14	39	60	100	142	366
Total	7	15	21	63	183	343	443	481	407	1,963

TABLE V(B). Age- and Sex-Specific Cancer Death Rates (per 100,000) in Laredo, 1895–1981

Year of death	Age (yrs)									Total
	< 5	5–14	15–24	25–34	35–44	45–54	55–64	65–74	75+	
Males										
1895–1904	0	0	0	17	29	22	20	233	0	14
1905–1914	0	0	4	23	25	22	130	287	561	27
1915–1924	0	3	0	0	30	47	93	268	179	20
1925–1934	0	2	5	15	29	89	245	398	576	43
1935–1944	20	0	9	9	35	143	322	532	778	63
1945–1954	7	3	4	8	18	103	420	584	1,114	76
1955–1964	14	3	8	24	39	141	400	803	1,071	102
1965–1974	10	5	4	6	35	129	346	818	1,637	117
1975–1981[a]	8	3	3	12	26	76	302	644	1,651	104
Females										
1895–1904	0	0	4	6	53	231	254	357	333	41
1905–1914	0	4	0	11	101	248	281	393	353	54
1915–1924	6	3	0	17	50	172	308	363	517	101
1925–1934	4	2	9	37	120	235	395	433	810	82
1935–1944	0	8	9	18	86	208	378	471	588	79
1945–1954	8	5	11	27	86	195	401	636	842	123
1955–1964	0	1	6	21	70	165	369	610	897	102
1965–1974	2	2	1	23	58	164	287	632	866	119
1975–1981[a]	3	3	3	12	39	121	233	489	1,018	101

[a]1975–1981 is a short interval because of unavailability of data at time of study; data for 1981 incomplete.

so that rates could be computed. Those rates, shown in Table (V)B, reflect the kinds of age and sex patterns observed in the contemporary U.S. population, suggesting that the diagnosis of disease is reasonable. The rates also reflect two trends. Age-specific cancer rates increase both with age and with time, as has been the national experience. Partly this has been due to improvements in the accuracy of diagnosis. Generally, however, it relates to the decrease in risk of terminal infectious disease, to decreased fertility, and to increased smoking and other exposure to cancer-causing agents in the diet and environment [eg, Doll and Peto, 1981].

The standardized mortality ratios (SMRs) are presented in Table VI for the major organ sites for both sexes in 1950 and 1970. Even with 90,000 current inhabitants, the population of Laredo is too small to estimate age-, sex-, and site-specific cancer rates with precision, so we have used indirect standardization to compare cancer death patterns in Laredo to those in the U.S. at large. Further, we have averaged the number of deaths (per site and sex) for the decade surrounding each census year to reduce the small-sample effect on the SMR. The standard rates used were taken from the U.S. white rates given in *Vital Statistics of the United States, Mortality, Volume B,* for the appropriate years.

Overall, there was a striking deficit of cancer for the major sites (lung, female breast, and colon) and a relative excess of cancer of the cervix and the middle digestive system (gallbladder, liver, stomach, and perhaps pancreas), especially in females. This is similar to the SMR patterns found among Mexican-Americans in Los Angeles and in New Mexico [see Weiss et al, 1984; JNCI Monogr 69, 1982]. The finding of an excess of cancer deaths at many occult sites suggests that the level of

TABLE VI. Standardized Mortality Ratios (SMRs) for Malignant Neoplasms in 1950 and 1970 (U.S. as Standard Population)*

Site	Males		Females	
	1950	1970	1950	1970
Buccal cavity and pharynx	1.02	0.79	1.36	0.60
Stomach	1.00	0.88	1.30	1.31
Colon and rectum	0.24	0.36	0.35	0.38
Liver	1.54	7.81	1.51	8.68
Gallbladder	NA	0.48	NA	1.62
Pancreas	0.64	1.00	1.49	0.99
Trachea, bronchus and lung	0.90	0.63	1.32	0.86
Bladder	0.30	0.86	0.74	0.78
Kidney and O/U urinary organs	0.69	2.04	1.91	0.78
Breast			0.56	0.98
Body of uterus, O/U uterus			1.69	0.76
Cervix uteri			1.12	1.49
Ovary			0.78	1.82
Brain	0.34	2.07	0.00	2.00
Bone and connective tissue	0.45	1.04	1.76	2.29
Melanoma of skin	0.00	0.28	0.21	0.52
Lymphosarcoma and reticulosarcoma	0.29	0.46	0.24	0.33
Leukemias	0.33	0.28	0.71	0.60
All sites	0.76	0.79	0.98	0.91

*O/U, other or unspecified. Source: Vital Statistics of the United States, B. Mortality, Volume II, U.S. Government Printing Office, for the years indicated; 1950 liver data include gallbladder.

diagnosis, at least in later years, is comparable to that found in an average U.S. community.

Although sample sizes are small, the shape of the age-of-onset curves for the major cancers (for which there are sufficient data) are very similar to those in other populations, indicating that similar factors are at work and, more importantly, that the pattern of diagnosis is comparable [Badzioch, 1982] (K.M. Weiss, unpublished results). One important trend that we have found is an increase in the death rates of postmenopausal breast cancer since about 1940. Death rates of premenopausal breast cancer have remained virtually unchanged over the same period and are comparable to those in the U.S. population at large [Buchanan et al, 1985].

Familial Cancer Patterns

To assess the familiality of cancer, we identified probands for whom there was at least one other family member with cancer of the same site. All types of cancer represented by a separate ICD-9 code (140–208) were tested. To identify other affected relatives, we sampled the probands' first-degree relatives and grandparents; however, inclusion of all lineal relatives did not materially change the result. The results of this procedure are presented in Table VII, which includes all instances of a multiply affected family for any single site. Only a few sites with multiplex families were found, and only in breast cancer did we find a family with more than two

TABLE VII. Frequency of Site-Specific, Multiply Affected Families in Laredo*

Cancer site tested (ICD-9 code)	Number of cases	Number of multiply affected families
Single sites		
Unspecified oropharyngeal (149)	22	1
Stomach (151)	235	2
Colon (153)	97	0
Liver (155)	205	3
Lung (162)	263	3
Female breast (174)	191	9
Prostate (185)	112	1
Site combinations		
Colon/rectum (153/154)	143	1
Gastrointestinal (150–154)	430	8
Female reproductive system (179–184)	311	5
Cervix/endometrium (179–182)	234	2
Breast/ovary (174/183)	263	13
Urinary system (188–189)	99	1
Hematopoietic system (159.1, 204–208)	82	0
CNS (191–192)	36	0
Lymphatic system (196,201)	23	0

*Numbers of cases include only those deaths that could be linked to a member of the genealogies. Except for one breast cancer family and one family for all gastrointestinal sites, among each of whom there were a total of three affected family members, all instances in the table contain only the proband and one other family member. All first-degree relatives plus grandparents were included in the family members tested. Every cancer site encoded by single ICD-9 codes, from 140 to 208, was tested, but no multiply affected families were found except as reported above. The site combinations in the second part of the table were tested and reported by Hill in the Utah study (see text). Other site combinations tested, with no multiply affected families, include: all cancer in relatives of childhood cancers, digestive system and breast (150–151, 155–157, 174), middle digestive system (151–152, 155–157), all female reproductive sites plus breast in both sexes and prostate (174–185).

affected individuals, a triad of affected individuals in a single family. Familial colon cancer, often found elsewhere, has not occurred in Laredo, at least as reflected by death certificates.

There have been many reports of multiple site familial cancer [eg, Anderson, 1982; King et al, 1984; Schimke, 1978; Strong, 1982; Swift, 1982]. We have examined familial occurrences for a selection of those grouped sites tested in the Utah study, the most comparable other data set showing familial aggregation [Hill, 1980a,b; Skolnick et al, 1981]. We did find some instances of multiple affection, but most of these were due to single sites already reported. Overall, we failed to find support for a general or common familial aggregation of cancer. Some sites, however, were occasionally familial, but because families contain different numbers, sexes, and ages of individuals a tabulation of the fraction of multiplex families does not lead to any simple conclusion.

In both of these tests (single site and multiple site combinations), we enumerated the frequency of occurrence of multiply affected kindreds. Because the amount of information available on the families varies, as does their size and structure, it is not possible to compare this directly to a simple expectation, so more rigorous methods are desirable. Thus we tested the pattern of time-, sex-, and age-specific risks as described in Materials and Methods. The results of this test for single sites of cancer, and for certain groups of organ sites, are presented in Table VIII. Except for sites noted in the table, the data are reported for sexes pooled; no additional information was contained in the sex-stratified cells.

Table VIII provides the results of the significance test described above. For most sites, there is a substantial probability that the observed excess of familial cases occurred by chance alone; that is, there is no evidence in the Laredo data base for excess familial risk for most sites. For some sites, however, some or all of the cells show only a small probability of such familial clustering by chance alone. Given the large number of tests we performed (most of which were not significant), the fact that a few sites proved to be significantly clustered in families might be unconvincing evidence of familial clustering of risk, even for the implicated sites. For those sites for which the overall probability was very small, however, an excess was observed in several cells; we therefore conclude that the excess is valid and not a chance finding.

The expected and observed probabilities of death, by sex, age, and cohort for each site or site combination shown in Table VIII, do not specify any specific measure of excess risk, since there is no single measure that would satisfy all purposes. One way to provide a measure of excess risk would be to compute the risk ratio (o_{ij}/e_{ij}), where o_{ij} and e_{ij} are the observed and expected risks of death from the designated cause for the ij^{th} cell in the table. For example, with regard to breast cancer in women who were aged 40–59 during the period since 1960, the expected risk was .023 and the observed risk .126, so the risk ratio was 4.48.

We have tested the site combinations reported by others [eg, Li and Fraumeni, 1969; Lynch, 1976; Bishop and Skolnick, 1984; Hill, 1980a,b; Skolnick et al, 1981] as showing most significant, intermediately significant, and no significant clustering in families. We find evidence for familial clustering of combined female reproductive sites (excluding cervical cancer, which has distinct and generally unrelated epidemiology), although this might be due mainly to the clustering of breast cancer. No significant clustering of nonbreast reproductive sites was observed in relatives of

TABLE VIII. Summary of Results of Tests for Excess Cancer in Family Members*

| | 1930–1959 | | | | | | | | 1960+ | | | | | | | | Global prob |
| | Age 40–59 | | | | Age 60+ | | | | Age 40–59 | | | | Age 60+ | | | | |
	rels	obs	exp	prob	rels	obs	exp	prob	rels	obs	exp	prob	rels	obs	exp	prob	
All cancer	37,769	.031	.024	<0.01	19,428	.080	.070	.02	6,941	.070	.054	<0.01	14,760	.168	.165	.03	$<10^{-5}$
Colorectum	2,062	0	.001	1.0	1,136	.010	.004	.34	263	0	.004	1.0	823	.011	.012	.66	0.94
Trachea, bronchus and lung	3,971	0	.002	1.0	1,983	.010	.006	.29	725	.015	.008	0.26	1,649	.006	.028	.99	0.74
Female breast	1,821	.019	.004	0.03	1,177	.010	.006	.45	361	.126	.023	<0.001	805	.100	.027	<.001	$<10^{-5}$
Ovary	689	.016	.002	0.14	427	.025	.002	.08	109	0	.005	1.0	288	.071	.013	.04	0.051
Female reprod.	4,159	.016	.013	0.36	2,695	.017	.016	.53	866	.108	.031	$<10^{-4}$	1,743	.093	.046	.01	$<10^{-4}$

*rels, number of person-years of exposure in relatives of probands observed; obs, prob of relatives dying in the exposure cell (Eq A4 for relatives); exp, expected risk of dying in cell in population (Eq A4); prob = π in Eq A5. Global probability is from the χ^2 (8 df) for the overall result (Eq A6). Results for first-degree relatives plus grandparents. Sexes pooled except breast and ovary. Only the most informative cells given here (includes all statistically significant results). Female reproductive sites: breast, ovary, endometrium. Other site combinations tested for multiply-affected families but where none occurred (Table VI) were not subjected to the above test.

breast cancer probands. No other combination of sites demonstrated significant familial clustering. The Li-Fraumeni familial pattern was not observed in Laredo; it is too rare to be expected in a population the size of the Laredo data base.

To address the question of the degree to which the Laredo genealogies permit us to detect excess risk, we present in Table IX the number of relatives available in the main informative cells and the 5% critical value (numbers of affected relatives required) for the sites summarized in Table VIII. The sites presented there cover the general range of values of expected risk and of sample size (number of relatives at risk in the cells) for the relatively common cancers. For these critical numbers of relatives, Table IX provides the corresponding risk ratios. The critical number of affected relatives is that minimum number that would occur by chance less than 5% of the time. Because the distributions are discrete binomial, with finite sample sizes, the level of significance chosen for these computations is only ~5%. The actual significance level is always $\leqslant 5\%$.

Table IX shows that for a wide range of sites and the sample sizes appropriate for these data risk ratios on the order of 1.5–3 can be detected at the 5% significance level. This is the general level of excess risk observed in other studies of familial cancer. When the observed risk was significantly greater than expected, the observed risk ratio, computed directly from Table VIII, was greater than the critical risk ratio given in Table IX.

Figure 1 summarizes the power of the tests used here in a data base the size of Laredo. For a range of sample sizes (numbers of at-risk relatives) of 50, 200, and 1,000, generally spanning the Laredo sample sizes, and for a range of risks from 0.001 to 0.05, Figure 1 presents the power to detect risk ratios up to 10 or more. Clearly, for risks on the order of 0.05, the power is adequate to detect even a two- or threefold increase. For rarer sites, where risk is on the order of 0.001, adequate power to detect small risk ratios requires samples on the order of 1,000, rarely attained in Laredo.

It is possible that our conclusion that excess familial risk is unusual in Laredo is wrong and that there is considerable excess risk that we do not have the statistical power to detect. This is not likely to be a serious problem, however, for several reasons: 1) Our test includes combined analysis from at least four risk cells, considerably raising the effective sample sizes. 2) Were familial cancer common, or were it to exist for many sites, more of the tests performed in our analysis should have been positive even if we could not detect all excess risk that existed. 3) When substantial excess risk existed, it was detected in more than one risk cell.

DISCUSSION

We are restricted in this study to family events that can be documented from the records available to us. Our information differs in certain ways from that of the Mormon study. One difference is that our data do not include a registry with mandatory case reporting, which was partially available in Utah. Another is that the Mormon Church requires of its members that they compile family "group sheets" that document relatives in extended nuclear families, whereas we had to document all family members directly. We cannot consider events of which we have no direct record, of course, but this was true also of the Mormon studies. Mormons have emigrated from Utah, often in systematic or organized patterns. Some of those

TABLE IX. Minimum Risk Ratios Detectable at 5% Significance Level for Cancer Sites Tested in Family Members (entries correspond to Table VIII)*

	1930–1959						1960+					
	Age 40–59			Age 60+			Age 40–59			Age 60+		
	n	cval	rr	n	cval	rr	n	cval	rr	n	cval	rr
All ca	3,288	93	1.2	1,890	151	1.1	1,193	77	1.2	1,498	270	1.1
Colorec	187	1	5.4	104	2	4.8	52	1	4.8	89	3	2.8
Lung	343	2	2.9	193	3	2.6	131	3	2.9	158	8	1.8
Breast	156	2	3.2	95	2	3.5	52	3	2.9	76	5	2.4
Ovary	63	1	7.9	40	1	12.5	17	1	11.8	27	1	2.8
Fem rep	369	9	1.7	235	7	1.9	131	8	2.0	175	13	1.6

*Sites listed are same as those in Table VIII; rr, minimum risk ratio (observed risk/expected risk) that could be detected in Laredo genealogies at the 5% significance level; n, observed number of probands' relatives in age and time category (ie, sample size); cval, number of affected relatives that would be significant at the 5% level (to nearest integer).
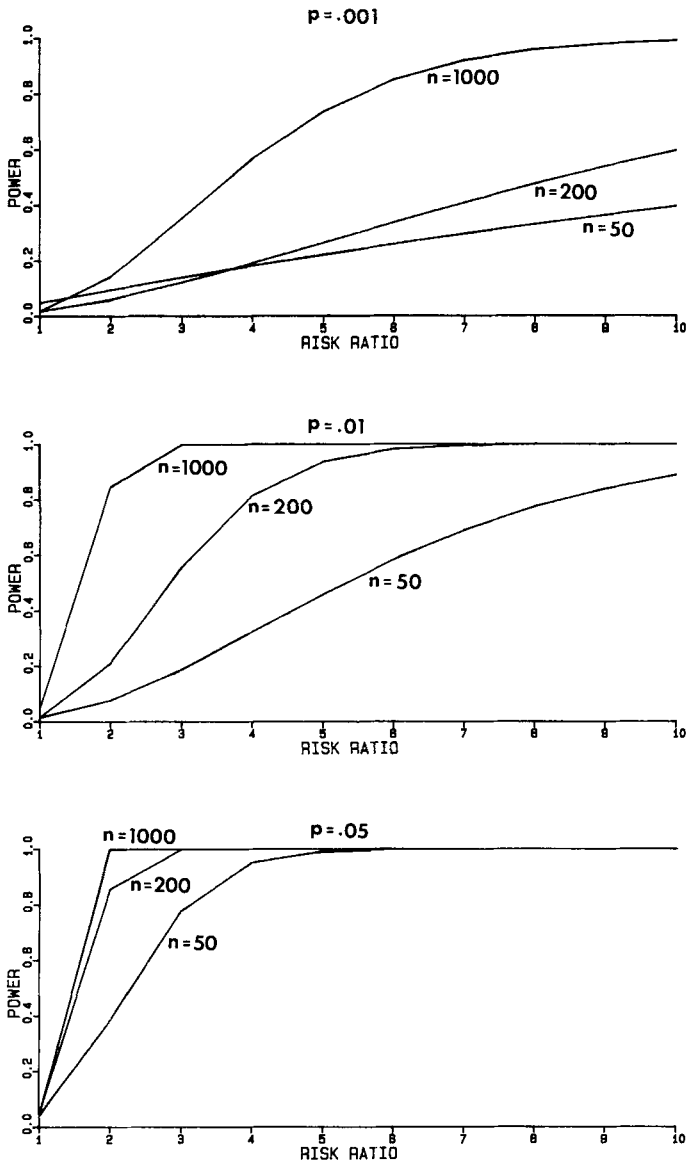
Fig. 1. Power curves for a selection of sample sizes (n), baseline population risks (p), and risk ratios. The power curves are drawn with 5% critical values corresponding to the values of baseline risk, and sample sizes. The computations were performed by evaluating binomial probabilities below the critical value, for risks corresponding to the plotted risk ratios. For P = 0.001, small risk ratios result in non-monotonic power computations (as a function of n) because of discontinuity of the corresponding binomial distributions.

individuals will be included on group sheets of Utah Mormons although their deaths will not. There have been no major organized immigration patterns into Laredo of which we are aware, though clearly there has been steady immigration (including family members) from Mexico and some emigration to other parts of the U.S. as well as occasionally back to Mexico. There are also immigrants from Europe and other parts of the U.S. There is no way to exclude the possibility that important cancer patterns are associated with these undocumented events, but there is no reason to think that our results are seriously affected by such patterns, and every family study must face limited ascertainment of information, since human pedigrees extend beyond history in many ways.

We have adopted a new method to examine familial risk in a whole population of genealogies, developing an internal standard for comparing overall risk differences, after cohort, sex, and age effects have been taken into account. Our methods are variants of stratified standardized incidence-density comparisons. There are several reasons why we feel that the use of the genealogical index (GI), as used in the Utah study [eg, Hill, 1980a,b; Skolnick et al, 1981; Bishop and Skolnick, 1984], is not ideally suited to our purposes or population sample. First, as Table I shows, the Laredo population is an assembly of disjoint genealogies of different structures and sizes. The GI between two members of a genealogy is a function of its size and structure. Therefore, if the GI between two affected pairs is to be compared to that between pairs of matched controls, those controls must be suitably matched in terms of the size and structure of their genealogies. This is not generally feasible, yet an unrestricted set of controls will yield a different baseline genealogical relationship than that applying to the observed cases. Second, when more than one pair of cases occurs within a given genealogy, the GI values for each pair are not statistically independent. Third, because of the lack of multiply affected genealogies, the GI between case pairs in our study would have produced a resultant distribution of GI values heavily concentrated at indeterminate values (ie, no known relationship between cases). This makes test contrasts inapplicable.

Finally, the GI might be biased towards close relationships, since close relatives are more likely to be present and documented in a population than are more distant relatives. Indeed, the genealogical index is a conditional test, comparing the degree of relationship among those pairs of cases/controls that were linked in the same genealogy; the relationship between other pairs is not known, and the fraction of unlinkable pairs might be different between cases and controls. The Utah study did find closer kinship among case pairs [Skolnick et al, 1981; Bishop and Skolnick, 1984; Hill, 1980a].

Some of these problems have been discussed in an analytical context by Chakraborty [1985]. Our method here is a more direct comparison of familial risk, with a baseline risk applicable to the entire Laredo population (age/sex/cohort stratified) rather than to specific genealogical relationships. The method is also free of the inherent distributional and sampling uncertainties associated with simulation as well as certain other problems [Chakraborty, 1985].

It is clear that we do not confirm the degree of familiality that has been reported in Utah and in other studies for major cancer sites (colon, lung, stomach, and other organ systems) with the single exception of breast cancer. Further, familial excess for many site combinations reported by the Utah study does not exist in Laredo, even though the large families in Laredo provide ample opportunity for multiple affection

to have occurred. If our results were due simply to low cancer prevalence, a low probability that multiple cases will occur in small groups such as families, one might expect the same problem in Utah.

It might be that a difference between Laredo and Utah is related to the low levels of exposure Mormons are known to have for many of the important risk factors for cancer, leading to their generally low rates of cancer. Although Laredo also has lower rates of many major cancers, the exposure differences might be relevant to the findings for certain sites. For example, Mormon religious beliefs and practices serve to reduce exposure to important risk factors (cigarette smoke, alcohol); they frequently have large family sizes, which might protect to some extent against female reproductive site cancers; and Mormons might also have a protective diet.

It seems unlikely that the genetic basis for susceptibility in Laredo is different from that in the largely northern European population of Utah, since the Mexican-American population of Laredo is genetically about 75% European, although largely Spanish, in ancestry [Ferrell et al, 1985; Gottlieb, 1983; Relethford et al, 1983].

The Mormon study included death and incidence data for some parts of its population and study period, and our study has been restricted to deaths alone. Other familial studies have also used incidence data. Although more cases should be discovered with the former approach, there is no reason why morbidity data should demonstrate excess familial risk when mortality data do not except in the sense that larger sample sizes have increased statistical power. Most of the tumors typically showing familial risk have been lethal, especially prior to the increased availability of endoscopic and other early detection methods. Therefore, it seems unlikely that our overall result would change with the inclusion of morbidity data.

It is unlikely that poor ascertainment or diagnosis is responsible for the lack of cancer familiality that we have observed. It is possible that immigration and emigration are nonrandom with regard to familial susceptibility to cancer, but we have no reason to suspect that this has been a systematic factor in Laredo. As is shown in Table VI, many cancers of occult or difficult-to-diagnose nature are more common in Laredo than elsewhere in the U.S. The fact that breast cancer in Laredo exhibits familiality and fertility-related epidemiology comparable to the overall U.S. experience [Buchanan et al, 1985] indicates that our findings are artifacts neither of diagnosis or classification nor of the way in which the genealogies were assembled. We conclude that there have been biases of various sorts in many previous studies and/or that there is less familial risk, or at least less *expressed* familial risk, in Laredo than typically exists in white populations.

Although we found a significant excess of cancer in relatives for all sites combined, the amount (excess or relative) of the increase was not great. This should underline the distinction between a significant excess and an important excess. The importance of an excess is to be found in the numerical increase in risk or perhaps in biological considerations such as the age at onset. Younger ages of onset, for example, may have a greater impact at both population and individual levels than a numerically comparable increase of cases at older ages.

Overall, Mexican-Americans currently have slightly less cancer than the U.S. average [Weiss et al, 1980]. However, this might be changing; rates at some of the common sites are increasing in Mexican-American populations and for some sites, such as colon, breast, and lung, might be converging with the U.S. overall rates [Buchanan et al, 1985; MacDonald and Heinze, 1978]. As this trend continues, there

may be an unmasking of familial predisposition not previously expressed. For many cancers, familiality might be due to shared environmental exposures; this is something that is difficult but important to determine.

We had originally expected to identify high-risk families in the population with the use of nonparametric tests for excess risk [Chakraborty et al, 1980, 1984; Weiss et al, 1980; Smouse et al, 1982] and then to perform a large-scale segregation analysis on the entire set of extended genealogies so identified. It is now obvious from both our results and those of the Utah study that the delayed-onset, low-probability expression of most cancers precludes the frequent occurrence of many cases in single genealogies [Weiss, 1985]. Although it has been done successfully in some samples of families for breast cancer [Williams and Anderson, 1984; Go et al, 1983; King et al, 1983], and for colon and a few other cancers, we probably will not learn much about heritable genetic risk factors for most cancers from segregation analysis.

Familial cancer is apparently not an important public health problem, but even on a population basis it might be important to intensify efforts to identify familial susceptibility to cancer precursor states. These can be much more prevalent than cancer itself, which develops only in a subset of the vulnerable individuals. When such precursors are genetic, their identification would lead to a higher detectability of the genes involved as well as provide opportunities for timely intervention [Weiss, 1985]. Also, an effective use of research resources should include the identification of those relatively rare families who do in fact manifest greatly elevated cancer susceptibility.

ACKNOWLEDGMENTS

REFERENCES

Anderson DE (1982): Familial predisposition. In Schottenfeld D, Fraumeni J (eds): "Cancer Epidemiology and Prevention." Philadelphia: Saunders, pp 483–493.

Anderson RL, Bancroft TA (1952): "Statistical Theory in Research." New York: McGraw-Hill.

Badzioch MD (1982): "A Study of Mortality in a Mexican-American Community—Laredo, Texas, 1900–1930." MPH thesis, University of Texas School of Public Health (also CDPG Report No 4, Center for Demographic and Population Genetics, University of Texas, Houston).

Bishop DT, Skolnick MH (1984): Genetic epidemiology of cancer in Utah genealogies: A prelude to the molecular genetics of common cancers. J Cell Physiol 3 [Suppl]:63–77.

Buchanan AV, Weiss KM, Chakraborty R, Anderson DE, MacNaughton NL (1985): Breast cancer epidemiology in Laredo, Texas. JNCI 74:1199–1206.

Buchanan AV, Weiss KM, Schwartz RJ, MacNaughton NL, McCartan MA, Bates SS (1984): Reconstruction of genealogies from vital records: The Laredo epidemiology project. Comp Biomed Res 17:326–351.

Chakraborty R (1985): Some analytical explorations for detecting familial aggregation of disease traits. In Chakraborty R, Szathmary EJE (eds): "Diseases of Complex Etiology in Small Populations: Ethnic Differences and Research Approaches." New York: Alan R. Liss, Inc., pp. 21–38.

Chakraborty R, Weiss K, Majumder P, Strong LC, Herson J (1984): A method to detect excess risk of disease in structured data: Cancer in relatives of retinoblastoma patients. Genet Epidemiol 1:229–244.

Chakraborty R, Weiss KM, Schull WJ (1980): A test for the randomness of the occurrence of a disease trait in familial or other similar ordered sequences of epidemiological data. Proc Natl Acad Sci USA 77:2974–2978.

Doll R, Peto R (1981): "The Causes of Cancer." Oxford: Oxford University Press.

Ferrell RE, Chakraborty R, Stern MP (1985): Prevalence of non-insulin dependent diabetes mellitus and Amerindian ancestry in a Mexican-American (MA) population. Am J Human Genet 37:A195.

Go RP, King MC, Bailey-Wilson J, Elston RC, Lynch HT (1983): Genetic epidemiology of breast cancer in families. I. Segregation analysis. JNCI 71:455–461.

Gottlieb K (1983): Genetic demography of Denver, Colorado: Spanish surname as a marker of Mexican ancestry. Hum Biol 55:227–234.

Harvald B, Hauge M (1963): Heredity of cancer elucidated by a study of unselected twins. J Am Med Assoc 186:749–753.

Hill JR (1980a): "A Kinship Survey of Cancer in the Utah Mormon Population." PhD thesis, University of Utah.

Hill JR (1980b): A survey of cancer sites by kinship in the Utah Mormon population. In Cairns J, Lyons J, Skolnick M (eds): "Banbury Report 4: Cancer Incidence in Defined Populations." Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, pp 299–318.

Jacobsen O (1946): "Heredity in Breast Cancer: A Genetic and Clinical Study of Two Hundred Probands." London: H.K. Lewis.

King MC, Go RCP, Lynch HT, Elston RC, Terasaki PI, et al (1983): Genetic epidemiology of breast cancer and associated cancers in high-risk families. II. Linkage analysis. JNCI 71:463–467.

King MC, Lee GM, Spinner NB, Thomson G, Wrensch MR (1984): Genetic epidemiology. Annu Rev Public Health 5:1–52.

Li FP, Fraumeni JF (1969): Soft-tissue sarcomas, breast cancer and other neoplasms. A familial syndrome? Ann Intern Med 71:747–752.

Lipkin ML, Sherlock P, DeCosse JJ (1980): Risk factors and preventive measures in the control of cancer of the large intestine. Curr Prob Cancer Vol. IV, No. 10.

Lynch HT (ed) (1976): "Cancer Genetics." Springfield, IL: Charles C. Thomas.

MacDonald E, Heinze E (1978): "Epidemiology of Cancer in Texas." New York: Raven Press.

Macklin MT (1959): Comparison of the number of breast cancer deaths observed in relatives of breast cancer patients and the number expected on the basis of mortality rates. J Natl Cancer Inst 22:927–954.

Relethford J, Stern M, Gaskill S, Hazuda H (1983): Social class, admixture, and skin color variation in Mexican-Americans and Anglo-Americans living in San Antonio. Am J Phys Anthropol 61:97–102.

Schimke N (1978): "Cancer Genetics in Man." London: Churchill Livingstone.

Schwartz R, Weiss K, Buchanan A (1984): Looking into the black box: Rational approaches to record linkage. Ann Demog Historique 84:119–128.

Skolnick M, Bishop DT, Carmelli D, Gardner E, Hadley R, Hasstedt S, Hill JR, Hunt S, Lyon JL, Smart CR, Williams RR (1981): A population-based assessment of familial cancer risk in Utah Mormon genealogies. In Arrighi F, Rao PN, Stubblefield E (eds): "Genes, Chromosomes, and Neoplasia." New York: Raven Press, pp 477–500.

Smouse P, Weiss K, Chakraborty R (1981): A simple test for aggregation of disease occurrence in genealogical data. Hum Hered 31:334–338.

Strong LC (1982): Genetic-environmental interactions. In Schottenfeld D, Fraumeni J (eds): "Cancer Epidemiology and Prevention." Philadelphia: Saunders, pp 506–516.

Swift M (1982): Single gene syndromes. In Schottenfeld D, Fraumeni J (eds): "Cancer Epidemiology and Prevention." Philadelphia: Saunders, pp 475–482.

Weiss K (1985): Phenotype amplification, as illustrated by cancer of the gallbladder in New World peoples. In Chakraborty R, Szathmary E (eds): "Etiology of Complex Diseases in Small Populations: Ethnic Differences and Research Approaches." New York: Alan R. Liss, Inc., pp 179–198.

Weiss K, Chakraborty R, Schull WJ, Rossmann D, Norton S (1980): The Laredo Epidemiology Project. In Cairns J, Lyons J, Skolnick M (eds): "Banbury Report 4: Cancer in Defined Populations." Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, pp 267–284.

Weiss K, Ferrell R, Hanis C (1984): A New World syndrome of metabolic diseases with a genetic and evolutionary basis. Yearbook Phys Anthropol 27:153–178.

Williams W, Anderson D (1984): Genetic epidemiology of breast cancer: Segregation analysis of 200 Danish pedigrees. Genet Epidemiol 1:7–20.

Woolf CM (1955) Investigations on genetic aspects of carcinoma of the stomach and breast. Univ Calif Publ Public Health 2:265–350.

World Health Organization (1978): Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death." Geneva: World Health Organization.

APPENDIX

The purpose of this appendix is to describe the methods by which cancer mortality was tabulated and contrasted for the population as a whole and for family members of affected probands.

Obtaining the Observed and Expected Death Risks

All person-years of experience for all individuals in the genealogical data base were tabulated. The tabulation was divided into classes according to sex (s: 1 = male, 2 = female, 3 = both), to age (a: 1 = 0–14, 2 = 15–39, 3 = 40–59, 4 = 60+), and to chronological time (c: $1 \leqslant 1899$, 2 = 1900–1929, 3 = 1930–1959, 4 = 1960+). In the analysis, we have typically used only sexes-combined or one sex, but it is possible to use both; the use of all three sex categories would not be correct, of course, since the strata are not disjoint. For all individuals (i: 1–302,468) in the data base, we tabulated $Y_i(a,c,s)$ as the number of years the individual i spent while age a, in the appropriate sex and year category. For example, a male born in 1910 who lives to be age 48 contributes person-years as follows:

$$Y_i(1,2,1) = 15$$
$$Y_i(2,2,1) = 4$$
$$Y_i(2,3,1) = 21$$
$$Y_i(3,3,1) = 9.$$

In the first category, he reaches age 15 in the year 1925. He spends 4 years aged 15–39 still in the historical period 1925–1929, and 21 years still aged 15–39 in the historical period 1930–1959. He spends his final 9 years of life, from age 40 to 48 (inclusive), in the historical period 1930–1959.

For the whole data base, we tabulate values of $Y_i(a,c,s)$ as

$$Y.(a,c,s) = \sum_{i=1}^{N} Y_i(a,c,s), \qquad N = 302,468, \qquad (A1)$$

for each category, summed over all individuals. Each individual who enters a cell contributes one "person" to the number of individuals who entered the cell, that is, increments a person-counter $P(a,c,s)$. Each death is recorded for a cell in a death counter $D(a,c,s,j)$, where j is the cause code for the disease causing death. For example, our hypothetical individual added one person to the counters $P(1,2,1)$,

P(2,2,1), P(2,3,1), and P(3,3,1). If he died of cause j, he contributed 1 to the counter D(3,3,1,j). The mortality density in each cell is given by

$$R(a,c,s,j) = D(a,c,s,j)/Y.(a,c,s). \tag{A2}$$

The mean years lived in each cell, by members of the entire population, is determined by

$$\overline{Y}(a,c,s) = Y.(a,c,s)/P(a,c,s), \tag{A3}$$

and now we can compute the probability that a person who entered the cell survived the cell (with respect to the disease in question). We do this by assuming that each individual who enters a cell is exposed to a constant mortality density $R(a,c,s,j)$ for that cell, as given by eq A2. The probability of surviving $\overline{Y}$ (a,c,s) years at that risk is given by the standard formula

$$S(a,c,s,j) = \exp[-\int_0^{\overline{Y}(a,c,s)} R(a,c,s,j)dt]. \tag{A4}$$

These values are all tabulated initially for the whole population and for each specific disease category (eg, cancer site). Then, using similar tabulation methods, we compute the analogous values (which we can symbolize by adding a prime to the variable name) for those who are relatives of each proband. Hence the totals are now summed over all relatives of cases rather than over all N members of the data base.

Resulting from this are a set of probabilities, $S(a,c,s,j)$ and $S'(a,c,s,j)$, and a pair of mortality rates, $R(a,c,s,j)$ and $R'(a,c,s,j)$. In most cells, however, the numbers are too small and the distributions too sparse to make useful comparisons of the rates themselves. To determine the probability, $\pi(a,c,s,j)$, that we would observe at least as many deaths as we have observed, we computed the size of the upper-tail probability of a binomial distribution with P', D', and S as parameters. Suppressing the series of variables (a,c,s,j), this is

$$\pi(a,c,s,j) = \sum_{r=D'}^{P'} \binom{P'}{r} S^r [1 - S]^{(P'-r)}. \tag{A5}$$

This provides a test of the significance of the observations in each cell, and to obtain an approximate overall significance test we use the fact that

$$T(j) = -2 \sum_a \sum_c \sum_s \ln[\pi(a,c,s,j)] \tag{A6}$$

is distributed as χ^2 with 2m degrees of freedom, where m is the number of cells evaluated [Anderson and Bancroft, 1952]. This is the method by which the analysis of familiality was computed. Separate runs were made for first- and second-degree relatives of the probands, and for all lineal relatives.

Tabulation of Multiply Affected Families When There Is Complete Ascertainment

In assessing familial risk in this study, all affected cases were used as probands, and the rates of disease in their relatives, divided into 32 risk classes, were compared to those in the entire population. This practice differs from most studies in that: 1) the probands represent 100% of the cases in the population and 2) every affected family member will also be a proband. Thus, many family members' exposure and disease experience will be counted more than once (each time a member of the family is used as a proband). Here we show that this is a valid way to assess family risk in our sampling frame and that it leads to an unbiased estimate of the population risk under the null hypothesis of no familial aggregation of disease.

Families of size n with uniform risk p to each member. Let the data consist of a series of families, all of the same size, n, in which the risk of disease is the same, p, to each member. If there is no excess familial risk, then the number of affected individuals in any family, r ($r = 0, 1 \ldots n$) will be a binomial function of the risk, p, and family size, n. If there are N cases in a total population of size T, then $p = N/T$, and the probability that a family of size n has r affected is simply the corresponding term from a binomial distribution:

$$\text{Prob. } (r \text{ affected} \mid n,p) = \binom{n}{r} p^r (1 - p)^{n - r}. \tag{A7}$$

Because each of the affected individuals will be a proband, this family will be ascertained r times. The observed fraction of nonproband members of this family who are affected by disease is $(r - 1)/(n - 1)$. Hence the weighted total count of affected relatives in families consisting of $r - 1$ affecteds (not including proband) is

$$r \binom{n}{r} p^r q^{n - r} (r - 1). \tag{A8}$$

In the whole population of families, then, the estimate of p will be given by the weighted average of all possible families of size n (families with $n - 1$ relatives). The expected value of this estimate, $E(\hat{p})$, is given by

$$E(p) = \frac{\dfrac{1}{n - 1} \displaystyle\sum_{r=1}^{n} r \binom{n}{r} p^r q^{(n - r)}(r - 1)}{\displaystyle\sum_{r=1}^{n} r \binom{n}{r} p^r q^{(n - r)}}, \tag{A9}$$

where the denominator is the sum of all probabilities for families having at least one affected individual. The denominator is recognizably equal to the expected value of a binomial distribution, np. Equation A9 can be simplified in the following way:

$$E(p) = \sum_{r=1}^{n} (r - 1) \frac{r \binom{n}{r} p^r q^{(n-r)}}{np \cdot (n - 1)}$$

$$= \frac{n(n - 1)p^2}{n(n - 1)p} \sum_{r=2}^{n} \binom{n - 2}{r - 2} p^{r-2} q^{(n-r)} \qquad (A10)$$

$$= p(p + q)^{(n-2)} = p.$$

That is, the expected probability that a relative of a proband is affected, if there is no excess familial association of risk, is the population risk, p. This proves that if p and n are constant then it is correct to include each family as often as there are cases within it.

A mixture of family sizes. In reality, we have a mixture of family sizes. Let $n_1, n_2 \ldots n_N$ be the family sizes (including proband) for the N cases. The derivation above implies that

$$E(r_i - 1) = (n_i - 1)p \quad \text{for } i = 1, 2 \ldots N, \qquad (A11)$$

so that a pooled estimate of p is

$$p = \sum_{i=1}^{N} (r_i - 1) / \sum_{i=1}^{N} (n_i - 1), \qquad (A12)$$

which has expectation

$$E(p) = \frac{\sum_{i=1}^{N} E(r_i - 1)}{\sum_{i=1}^{N} (n_i - 1)} = \frac{p \sum_{i=1}^{N} (n_i - 1)}{\sum_{i=1}^{N} (n_i - 1)} = p, \qquad (A13)$$

since the family sizes, n_i's, are all fixed.

Mixture of risk levels. In our data, we subdivided the population into 32 categories, within each of which risk was assumed to be approximately constant. These categories involved age, sex, and birth cohort subdivisions of the data. Since all relatives within a category have the same risk level (p_k, k = 1, 2 . . . 32), and the only varying element over which pooling is done is the number of relatives in the kth risk category, the above argument on the mixture of family sizes shows that p_k's are estimated without bias.

The unbiased nature of the estimated risks and the appropriateness of the test criterion used in the text were also checked by means of 100 random trials, for each of which 250 random probands and their available relatives were selected; patterns of death from cancer in the relatives were shown to be comparable to those expected in the population as a whole.

Edited by D.C. Rao