INHERITED RISK FOR COMMON DISEASE

by

Helen Banava

S.B. Chemistry, Massachusetts Institute of Technology, 1991

M.S. Chemistry, California Institute of Technology, 1993

S.M. Toxicology, Massachusetts Institute of Technology, 2002

Submitted to the Division of Biological Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Toxicology

at the
Massachusetts Institute of Technology
February 2007

Author.....
Division of Biological Engineering
January 5, 2007

Certified by.....
William G. Thilly
Professor of Toxicology
Thesis Supervisor

Accepted by.....
Peter C. Dedon
Professor of Toxicology
Chair, Ph.D. Thesis Committee

Accepted by.....
Douglas A. Lauffenburger
Director, Division of Biological Engineering

Accepted by.....
Kari Hemminki
Professor of Mol. Genetic Epidemiology, DKFZ

Accepted by.....
Stephan Morgenthaler
Chair of Applied Statistics, EPFL

INHERITED RISK FOR COMMON DISEASE

by

Helen Banava

Submitted to the Division of Biological Engineering
on January 5, 2007 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in
Toxicology

ABSTRACT

Linkage disequilibrium studies have discovered few gene-disease associations for common diseases. The explanation has been offered that complex modes of inheritance govern risk for cancers, cardiovascular and cerebrovascular diseases, and diabetes. Such studies, however, depended on the untested assumption of monoallelic risk. My research advisor and I set out to investigate whether simple forms of inherited risk, monoallelic or multiallelic, could be excluded by analysis of familial risk for a common disease, such as colorectal cancer (CRC). First, we derived formulæ that describe the risk for monogenic, multigenic, and polygenic possibilities of Mendelian inheritance. Next, we obtained an estimate of minimum lifetime risk for CRC of >0.26. Then, we examined the case of late-onset CRC, using the Swedish Family Cancer Database (1958-2002) to estimate the familial relative risk for CRC diagnosis at age 50 or older, and obtained an estimated range of 1.5 to 3.0. We compared this range of actual values to the ranges of expected values for monogenic, multigenic, and polygenic modes of inheritance. We delimited bounds that can be placed on the conditions for various modes of inheritance. The key observation is that monogenic risk for CRC is included among various possibilities, and cannot be eliminated by existing observations. The arguments herein indicate that further efforts can and should be made to obtain more precise estimates of familial risk for CRC and other common forms of cancer.

Thesis Supervisor: William G. Thilly
Title: Professor of Toxicology

'Development of Western science is based on two great achievements: the invention of the formal logical system (in Euclidean geometry) by the Greek philosophers, and the discovery of the possibility to find out causal relationships by systematic experiment (during the Renaissance).'

Albert Einstein (1953)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Rates of most common forms of cancer rose in urbanizing societies in the early- to mid-twentieth century, while gastric cancer rates declined (http://epidemiology.mit.edu). With the exception of the role of cigarette use in lung cancer, the environmental changes responsible for these changes in cancer rates are unknown. For each form of cancer, it has been postulated that the changing environmental conditions interacted with one or more mutant forms of unknown genes to change the rate of oncogenic mutations, the rate of growth of preneoplastic lesions, or both (Herrero-Jimenez, Thilly et al. 1998; Herrero-Jimenez, Tomita-Mitchell et al. 2000). In addition to environmental exposure and genetic susceptibility, it is also possible that individual risk is dependent on stochastic factors, which may themselves be functions of genetics and/or environment. For example, if tumor initiation mutations were limited to the juvenile period, otherwise susceptible persons reaching adulthood without an initiated stem cell would be without subsequent risk by chance (Gostjeva and Thilly 2005). Family studies have clearly established a familial risk for most common forms of cancer (Lichtenstein, Holm et al. 2000; Czene, Lichtenstein et al. 2002). These observations serve as the basis for the hypothesis that there are genes that carry mutant alleles conferring lifetime risk, subject to environmental exposure and stochastic events. Most of the ~2000 recognized rare inherited syndromes have been accounted as a multi-allelic set of mutations in a single gene (monogenic risk) while a few rare syndromes have been shown to arise from multi-allelic sets of mutations in any of two or more genes (multigenic and/or polygenic risk) (http://www.hgmd.org). To date, only two genes have been convincingly associated with common diseases, the melanocortin 1 receptor gene, *MC1R*, with skin cancers (Rees 2000; Rees 2004), and the complement factor H gene, *HF1/CFH* with macular degeneration (Hageman, Anderson et al. 2005). (The term "common diseases" as used in this thesis refers to diseases that may afflict ~1% or more of the population during a lifetime, and includes vascular diseases, cancers, diabetes, and late-onset conditions broadly associated with aging.)

In the current research, we are exploring the intersection of estimates of lifetime risk and familial risk for common forms of cancer in an attempt to delimit the kinds of genetic risk

that may govern familial risk. We seek to discover if there is any reason to assume that such risks involve one or many genes, and whether or not these risks exclude ordinary modes of Mendelian inheritance in which heterozygosity or nulizygosity for one or more genes define risk. In this collaborative effort, Prof. K. Hemminki is kindly providing data from the Swedish Family Cancer Database (1958-2002), that permit independent calculation of the minimum fraction of persons at risk both in the general population and in the subpopulation with a family history of common cancers. As an example of a common cancer, we use colorectal cancer throughout this thesis, but derive simple algebraic models applicable to all forms of cancers or other common diseases. We have derived formulae that describe the expectation of inherited risk from monogenic, multigenic, and polygenic possibilities for the several Mendelian possibilities. These formulae have made it possible to compare theoretical expectations with actual data drawn from a large human population (Sweden, 1958-2002) to discover if any class or classes of simple forms of inheritance are arithmetically excluded. To carry out this objective, it was necessary to revise existing practices in calculations of age-specific familial relative risk to make them applicable to the quantitative estimation of the relative risk conveyed by inheritance. The revision involved broad reconsideration of the processes of carcinogenesis, to account for established facts, and to maintain a parallel relationship between mathematical and physiological models. Our results applied to colorectal cancer demonstrate that not all monogenic forms of risk are eliminated by the intersection of data on lifetime age-specific risk, familial risk and the quantitative distribution of gene-inactivating mutations in the human population.

The results of these analyses are already being applied to the design of experiments to discover the genes conferring risk for common forms of disease including all common forms of cancer. Researchers planning pan-genomic searches, such as Prof. G.M. Church of Harvard University, Prof. R.W. Davis of Stanford University, and my research advisor, can proceed with the expectation that while, as in the case of *MC1R* and skin cancers, genetic risk for other common cancers may in fact be conferred by mutations in a single gene, search strategies must comprehend the possibility of multi- and polygenic risks.

BACKGROUND

Goal and Motivation

This thesis intends to contribute to our understanding of inherited risk for late-onset forms of common cancers, which constitute more than 95% of all cancer deaths. In particular, we seek to discover if the simplest monogenic forms of Mendelian inheritance are excluded by any combination of data and/or logic. Motivation has been given by the failure of the extensive effort to use linkage disequilibrium mapping methods applied in the general population to discover any gene that carries risk for any common cancer.

Search for genes carrying mutations conferring inherited risk for common diseases

Public apologia for the method of linkage disequilibrium has hypothesized that inherited risk must be conferred independently by multiple genes, each gene carrying but a single risk-conferring mutation in the general population. This is a scenario of mono-allelic, multigenic risk. This idea seeks to explain the failure to discover gene-common diseases association by presenting a scenario in which the "true" bases for inherited risk for common diseases are essentially too complex for discovery using linkage disequilibrium studies in genetically heterogeneous populations. In practical terms, multigenicity has the effect of reducing the fraction of an affected cohort carrying risk from each particular contributing gene studied. The idea seems to be that if there were simple conditions of monogenic risk with Mendelian inheritance, linkage disequilibrium studies would have found them.

This failure has, however, been more formally addressed in terms of a direct contradiction of the central hypothesis underlying linkage disequilibrium methods: they depended on the arguments of Kimura and Crow (1964) for mono-allelic, as well as

monogenic risk for common diseases (Lander and Botstein 1989; Lander and Schork 1994; Tomita-Mitchell, Muniappan et al. 1998; Morgenthaler and Thilly 2006).

In Morgenthaler and Thilly (2006), the many known examples of generative mutational spectra encoding risk for rare inherited diseases in humans, the age and population size of the species *H. sapiens* were presented as a basis for the expectation that inherited risk for common diseases would be expected to be markedly multiallelic, absent selection of any particular mutant allele. But Morgenthaler and Thilly accepted the possibility that inherited risks were conferred by a single gene (monogenic risk) or multiple genes (multigenic and/or polygenic risk). They offered a combination of statistical logic and technical advances to discover gene-disease association via pangenomic pair wise trials that they argued would detect most, if not all, associations independently of the mono- or multi-allelic, mono-, multi-, or poly-genic nature of the inherited risk(s).

While Morgenthaler and Thilly (2006) accepted the need to design an analytical strategy that would capture both monogenic and multigenic risks in genetically heterogeneous populations, we questioned the assumption that genetic risks were perforce multigenic and/or polygenic, *i.e.* complex. What data supported such a contention? Were there any specific data that excluded simple monogenicity? If we could provide answers to these questions our effort could be used in planning the search for genes associated with any and all common diseases. If risks were monogenic then enumeration of the point mutations in the exonic sequences of a test gene in a case-control study with cohorts as small as 1,000 persons could discover the gene in a pan-genomic scan enumerating all point mutations in the exonic sequences of all genes; if monogenic risks were excluded then case cohorts of 10,000 or even larger would be required (Morgenthaler and Thilly 2006).

This is not a minor point. Public skepticism expressed in reduced support for public health research by the U.S. Congress is in part due to the failure of the promised genomic revolution to discover gene-common disease associations, an effort that has consumed an estimated fifteen billion dollars of public and private investment in the past decade

without tangible results. The estimated cost of pan-genomic studies of ~100 common diseases has been estimated at ~$250 billion using the most recent high throughput strategies if 10,000 persons are required in case cohorts. But various elements of the National Genome Institute are proceeding on the assumption that only a modest subset of known genes need be screened in case cohorts of ~1,000 persons, a strategy dependent on a dependable means to identify genes conferring risk (which does not exist) and a condition of monogenic or near monogenic risk for each disease studied (Morgenthaler and Thilly 2006). The predictable failure of these efforts will further erode public confidence in population genetics and science in general. It is necessary to analyze the data regarding risk of common cancers and other common diseases in order to create and execute a plan that will discover such genes that carry inherited risks. Science and society cannot really afford a continuing obfuscation of the truth about genetic risks.

Insofar as there are rare inherited syndromes that are monogenic and multi-allelic, such as phenylketonuria and cystic fibrosis, or multigenic and multi-allelic, such as the recessively inherited xeroderma pigmentosum (XP), where any of 7 nucleotide excision repair genes are involved (*XPA-XPG*), the dominantly inherited hereditary non-polyposis colorectal cancer (HNPCC, Lynch syndrome), where any of 5 mismatch repair genes are implicated (*MLH1, MSH2, MSH6, PMS2, PMS1*), and others, there is no reason to assume that risk for any particular common disease must be either monogenic or multigenic. Indeed, the medical classification of diseases must often be expected to group conditions of differing genetic etiology with common, if not explicitly identical, sets of symptoms. This being recognized, we set out to gather, organize, and analyze the existing forms of epidemiological and clinical data to discover if there were any bases for excluding the hypothesis of monogenic risk. We began with a new model of the quantitative distribution of gene-inactivating mutations among genes within the general human population (Morgenthaler and Thilly 2006).

The distribution of gene inactivating mutations in the human population
(Abstracted and edited from Morgenthaler and Thilly (2006), with permission)

As we are concerned with detecting perhaps one to many genes carrying risk-conferring mutations per common disease, it is important at the outset to have some concept of the expected number of both gene-inactivating and neutral gene-inactivating mutations/gene exonic segments, and their distributions over all genes. These expected numbers and their distributions have not yet been satisfactorily determined by observation. Here are employed: (i) observations of some 135 gene-inactivating mutation rates, (ii) an estimate of the age of the human species, and (iii) the relative numbers of gene-inactivating and neutral mutations distributed in the human genome's exons. The confluence of these data permits an initial estimate of the mean frequency and distribution of non-deleterious but gene-inactivating mutations in humans.

The mean rate of inherited gene-inactivating mutations in humans has been estimated on the basis of 135 gene "loci" for deleterious mutations to be about $(3 \pm 0.7) \times 10^{-6}$ ($\pm$ *S.E.* of the mean) per generation (Sankaranarayanan and Chakraborty 2000).

One might thus expect that accumulation of non-deleterious gene inactivating mutations summing per gene to an average approaching 0.03 has occurred during the ~250,000 years or ~10,000 generations estimated for the age of *H. sapiens* (McDougall, Brown et al. 2005).

To these must be added such high frequency non-deleterious polymorphisms in a small fraction of genes that were carried forward from the earlier hominid population or arising in the aboriginal human population. Among the single locus estimate, a range of mutation rates of 1×10^{-6} to 22×10^{-6} was also noted. Insofar as the upper limit of 22×10^{-6} inactivating mutations/generation was observed in 135 "trials", we may use a "Fermi approximation" to posit that ~ 1% of genes carrying non-deleterious gene inactivating mutations would carry an inactivated gene copy fraction at a level of about 0.22. Summed gene-inactivating allele frequencies for some genes, such as 0.3 for cytochrome *P450 2D6* or 0.8 for glutathione methyl-transferases, demonstrate that frequencies exist above

this roughly approximated 99% upper confidence limit (Strange and Fryer 1999; Cascorbi 2003).

Genes carrying recessive deleterious mutations also require consideration insofar as heterozygosity may confer risk for a common disease. In Hardy-Weinberg equilibrium, the steady state fraction of gene copies carrying deleterious mutations is approximately equal to the square root of the forward mutation rate, or about 1.7×10^{-3}, which would result in an average heterozygote fraction of $2pq \sim 3.4 \times 10^{-3}$ in the general population, where q is the fraction of gene copies carrying gene-inactivating, risk-conferring mutations and $p = 1-q$. Were three such genes' conditions of heterozygosity to confer risk for the same disease, a common, multigenic, ~ 1% risk would be achieved. Given an approximation of 22×10^{-6} as of the 99% upper confidence limit on forward mutation rates, one would expect a heterozygote fraction of $\sim 9.3 \times 10^{-3}$ in the general population, so that some recessive deleterious conditions may be expected to create risk for less common "common " diseases. Absent more precise knowledge of the architecture of human population genetics, a prudent technical strategy would comprehend potentially important risk-conferring mutant fractions, *i.e.* gene-inactivating, to values of q between 10^{-3} and 1.0. Based on experimental measurements by Sankaranarayanan and Chakraborty (2000) of inactivating mutations per gene copy, the mean and upper confidence limits have been estimated as follows:

Gene-inactivating mutations/gene's exonic segments:

Mean ~ 0.03 ± 0.081 (*SD*)

90% u.c.l. ~ $0.03 + 1.28 \times 0.081 = 0.134$

99% u.c.l. ~ $0.03 + 2.33 \times 0.081 = 0.22$

Estimates of neutral mutation frequencies have also been made based on the expectations of gene-inactivating mutations and recognition that the ratio of gene-inactivating to total exonic point mutations is approximately 12 in humans (derived from analyses of the SNP Database by Dr. B.J. Glassner, as reported in Morgenthaler and Thilly, 2006)

Neutral mutations/gene's exonic segments:

Mean ~ 0.33 ± 0.894 (*SD*)

90% u.c.l. ~ 0.33 + 1.28 x 0.894 = 1.48

99% u.c.l. ~ 0.33 + 2.33 x 0.894 = 2.41

Scanning of five genes, *HBB*, *POLB*, *CTLA4*, *PPARγ* and *TPMT*, by Prof. W.G.Thilly and his collaborators provide some assurance, based on comparisons to the estimates of neutral mutation frequencies, that the estimates of dispersion for gene-inactivating mutant fractions, q, are reasonably precise. Though five genes is too small a sample to use to describe the mean and distribution for all genes, scans of the exons of these genes found a range of total mutations/gene's exonic segments to be 0.1 to 0.8 with a mean of about 0.4 in reasonable accordance with the calculated estimates of 0.33 + 0.03 = 0.36 with a 90% u.c.l. of ~1.5 cited above. These uncertainties must be accepted *pro tempore*. Here, we employ these rough approximations to consider the conditions of monogenicity, multigenicity and polygenicity that are consistent with inherited risk for common diseases, such as colorectal cancer.

Large population databases with age-specific cancer rates

There were essentially two major sets of epidemiological data to explore. The first comprises the national historical records of age-specific mortality rates from common diseases, including many common forms of cancer. For instance, in the United States such records have been kept in a way that they can be matched to age-specific population numbers since 1900. Similar data have been recorded in Japan since 1952. These are available via http://epidemiology.mit.edu. Many other countries have maintained cancer registries; Sweden began its national cancer registry in 1958. Analyses of these data permit inferences about the history of changes in environmental risk for particular diseases. They also provide a means to calculate a minimum estimate of the fraction at genetic risk of a particular disease.

These age-specific national data sets do not, however, permit an explicit calculation of the relative risk that affects the first-degree relatives of an individual with a particular disease. Each first-degree relative would, in general, carry half of the mutant autosomal genes of the afflicted individual. Currently, the largest available data set for analyses of inherited risk for common diseases is the Swedish Family Cancer Database created and maintained by the research group of Prof. K. Hemminki, formerly at the Karolinska Institute, Huddinge, Sweden, and now at the Deutches Krebsforschungszentrum (DKFZ), Heidelberg, Germany (Hemminki, Li et al. 2001; Hemminki, Granstrom et al. 2005). Another database, which has yielded valuable results, and includes fewer individuals, but more generations, is the Utah Population Database (Goldgar, Easton et al. 1994).

Concept of an age-specific familial relative risk of a common disease, *FRR(t)*

In developing my arguments we used examples from the Swedish Family Cancer Database to consider the age-specific cancer rates in the general population and in the subpopulations of children with at least one parent with a late-onset cancer and of parents of at least one child with a late onset cancer. We have chosen the incidence of colorectal cancers (CRC) in Sweden and the U.S. to illustrate the value and limitations of our results. Using a general "two-stage" model of carcinogenesis, we argue that the ratios of age-specific risk early in the earlier late-onset cancer age intervals (ages 50-54,...,70-74), *FRR (t; late-onset)*, between the subpopulations with a parent or child diagnosed with a late onset cancer to the age-specific risk in the general population of parents or children approximates the Genetic Relative Risk, *GRR*. We chose these age intervals because early-onset forms of CRC, such as FAP and HNPCC syndromes, have already had their effects by this age, and the fraction of persons at risk of late-onset CRC has not yet been significantly diminished by CRC mortality.

Heretofore, the familial relative risk was calculated either on the basis of two or more cancers within a parent-child relationship, without regard to age or gender, or by

grouping all familial concordant cases, wherein both afflicted persons were of age 50 or greater.

Grouping by age and gender are, however, logical requirements in the definition of familial genetic risk. For late-onset colorectal cancer in persons between 50 and 74, the estimated familial relative risk is about 1.89 (Hemminki and Chen 2004). Nonetheless, about 40% of the cases of CRC among "children" born since 1932 were recorded at age 0-49, and a significant number of parent/child concordant cases actually involved a late-onset diagnosis in a parent and an early-onset case in a child, leading to the expectation that the value 1.89 is an underestimate of the value of familial relative risk for late-onset CRC. In contrast, the use of data clustered over all ages of late-onset disease as in the use of an age-interval 50-104, automatically leads to an overestimate of the value of genetic risk, because the values of *FRR(t; late-onset)* must increase significantly with age.

With regard to gender, for most cancer sites, with the exception of gallbladder cancer and thyroid cancer (in particular, papillary thyroid cancer), age-specific incidence or mortality rates are greater in males than females. Thus, genetic risk for cancer in a father or mother would have a higher probability of being observed in a son than in a daughter. The probability of observing genetic transmission from afflicted parents would be expected to be greater for sons than for daughters. Calculations based on pooled gender data would be expected to create a bias leading to underestimates of *GRR*.

The formal algebra underlying these general points is derived below in the section **Familial Relative Risk.**

As this thesis is being written, recognition of limitations of these two forms of expressing familial relative risk in estimating genetic risk is having a positive effect. Data from the Swedish Family Cancer Database are being re-organized to permit calculation of *FRR(t)* for common late-onset cancers using our specific re-definition of *FRR(t)*, and to account for gender.

For our purposes herein, we use a broad preliminary estimate of *FRR* between 1.5 and 3.0 for *CRC* in Sweden.

Based on observations of unchanging CRC rates in birth cohorts of the 20th-century in Sweden and the U.S.A., and demonstration of the absence of any spousal risk for CRC in Sweden (Hemminki, Dong et al. 2001; Hemminki and Jiang 2002), we further argue that *FRR (t = 50-54,...,70-74)* for male to male transmission for the birth decades of 1900-1909 and onward, represents the genetic relative risk (*GRR*), for late-onset cancers. Then, we formally consider a wide set of modes of Mendelian inheritance and calculate their expected *GRR* values over a wide range of fractions at genetic risk in the general population.

Formal expressions for genetic relative risk, for autosomal and sex-linked monogenic disorders, were previously derived for various kinds of relatives, using stochastic matrices (Li and Sacks 1954). Another method to obtain such expressions is using path analysis. Yet another approach is via mating tables, where parental and offspring combinations are numerated, probabilities are calculated, and values are summated. The latter approach allows a more tractable extension to multigenic and polygenic modes of inheritance, and was the method followed. We have confirmed the calculations of Li and Sacks for monogenic risks, and have extended them as the first algebraic derivation for expected values of *GRR* for multigenic and polygenic conditions of inherited risk.

One point should be emphasized: the data of the Swedish Family Cancer Database, organized by the Hemminki group, clearly demonstrate that there is a familial risk for common late-onset cancers. The derivation of algebra to use these data to obtain a more accurate estimate of *GRR* builds on this accomplishment, and seeks to use it to delimit the possible modes of risk inheritance.

Using these derived algebraic formulae, we have created and provided tables herein by which researchers may use the estimate of the minimum lifetime risk, *mLR*, for any common disease, the estimates of *GRR* as derived from values of *FRR(t)* in the earlier

years of late-onset diseases, and the observed mean and 99% upper confidence levels (u.c.l.) of q for gene inactivating allelic fractions in humans (Morgenthaler and Thilly 2006), to discover if any hypothesized form of monogenic, simple multigenic or simple polygenic risk are excluded by this set of disparate observations.

For illustrative purposes, we use the present rough estimates of *GRR* for CRC in Sweden, the minimum lifetime risk, *mLR*, for CRC in Sweden and the U.S., and an estimate of the range of values of q for inactivated gene copies in the human population. We apply this to the case of CRC risk. The value of our derivations and logic will increase as better estimates of these three parameters, particularly *GRR*, accrue.

Common cancers: a historical overview

Before launching into the specifics of our effort, we owe our readers an overview of the cancer problem that underlies our effort. As this thesis is written, the field of cancer research has received many powerful analytical tools and opportunities from the disciplines of statistics, epidemiology, population genetics, somatic genetics, histology, cytology, toxicology, molecular biology and biochemistry.

Unfortunately, these advances have not yet reduced age-specific mortality rates from the set of all-too-common late-onset cancers. Many improvements are observable for specific cancer types, however: the effect of fluorouracil therapy on skin cancers, Gleevec on chronic myelogenous leukemia, decreased cigarette use on lung cancer, and the unexplained historical decrease in gastric cancer. Insofar as a statistically significant decrease in overall reported cancer mortality rates in the U.S.A. since 1960 parallels the nationwide decrease in autopsies, some skepticism as to the validity of the conclusion that cancer rates are dropping is justified. The historical record of deaths from all malignancies in U.S. European American Males (EAM) and Females (EAF) (Figure 1a, b.) appears to indicate that by the birth cohorts of the late 19^{th}- and early 20^{th}-century, a new but relatively constant age-specific rate of cancer mortality had been established in

these ethnic- and gender-specific cohorts. Whereas women born in the early 19th-century had somewhat higher age-specific cancer mortality rates than men, increases in rates among males rose above the female rates in a historical process that can be seen in successive birth decades beginning with that of the 1820s.

However, plotting the age-specific mortality data, 1900-1997, for EAM for the birth cohorts from 1890-1899 to 1970-1979 (Figure 2) betrays no discernible differences among birth decade cohorts. While there are some increases and decreases in age-specific mortality rates in different sites for European Americans born during the 20th-century, such as cervical cancer (Pap Test), lung (smoking cessation), and non-melanoma skin cancers (fluorouracil), these can be accounted by the effects of medical and public health progress. However important this progress has been for particular cancer types, it cannot be gainsaid that there is no evidence to support the contentions of the American Cancer Society and the National Cancer Institute that age-specific cancer mortality rates have significantly decreased over the past few decades. Indeed, the 1972 "War on Cancer" was announced at a time when the U.S. autopsy rates were already in a decade-long decline that continues to this day, a medical-economical phenomenon that would lead one to expect a growing underestimation of death rates from cancers without external diagnostic features.

Estimation of the minimum fraction at lifetime risk, *mLR*

These macro-epidemiological points having been made, it may also be seen that changes principally increasing overall male cancer mortality rates in the latter half of the 19th- and early 20th-century (*e.g.* cigarette use) were well established as risk factors by the birth cohorts of 1890-1899, 1900-1909, as shown in Figure 2. Inspection of the data of Figure 2 and for nearly all common forms of cancer, save a few, such as leukemias and lymphomas, shows a maximum value at age intervals ranging from 80-84 to 95-99 depending on cancer type. Given the demonstration by the Hemminki group of a significant degree of familial risk for common forms of cancer, Herrero-Jimenez et al.

(1998, 2000) devised cancer models to account for these maxima in terms of a specific population sub-fraction at lifetime risk with a mortality rate perforce greater than in the sub-fraction not at risk. Applying this reasoning, the area under the lifetime age-specific incidence curves or mortality curves adjusted for survival through medical intervention, represented the minimum average number of conditions of cancers carried/experienced by members of the population. For instance, using the birth decade cohort of 1890-1899 depicted in Figure 2, one may note that the area under the curve is somewhat greater than 1.3. Correcting for survival rates across all cancers of about ~15% this number would increase to about 1.5. If such risks were Poisson distributed over the entire population some $e^{-1.5}$ or ~ 0.22 of the population would not be at lifetime risk of any late-onset cancer while ~0.78 of the population would be at risk of one or more forms of cancer.

One major possibility that overshadows this area of research is that atherosclerosis, responsible for perhaps half of all late-onset deaths, has common risk factors with some common forms of cancer. Atherogenesis and carcinogenesis share some important physiological features, among which are that an atherosclerotic plaque originates as "fatty streaks" in the juvenile period, is apparently derived from a single precursor cell (clonal), and creates a log-linear increase in mortality from cardiovascular disease, having the same approximate slope as most late-onset cancers, ~0.11 (W.G. Thilly, unpublished observation).

Thus, we carry the idea that overall risk of cancers are distributed over at least 80% of the population and individuals within the group at risk of any cancer may carry independent risk(s) for one more cancers. Now, we will use more specific information about CRC to apply these general logical points to a particular category of cancer diagnoses.

Figure 3 presents the historical record of lower digestive tract age-specific cancer mortality, which is constituted to some 95% by colorectal cancer. This approximation is necessary as the data in this form have been recorded since 1900 in the United States, whereas the specific designations colon cancer and rectal cancers were introduced more recently in the public record. As in the set of all cancers, CRC mortality for women was

greater than that for men in the birth cohorts of the 19th-century, but male age-specific rates were greater than female rates by the 20th-century.

Using the data of Figure 4 for the birth decade of 1890-99 for European American males, it is possible to make an initial calculation of the area under the curve, as a preliminary estimate of the fraction of this cohort at lifetime risk of death by CRC. This works out to ~0.18. However, improvements in early detection and treatment of CRC during the 20th-century acted on this cohort as it grew older, especially after 1950 when the cohort entered its 60s. Herrero-Jimenez organized the medical literature reports of CRC survival rates as a function of historical year and age so that an age-specific correction could be applied to the data of Figure 4.

Accounting for improved medical efficacy in early detection and treatment of CRC as illustrated in Figure 5 yields a better approximation to CRC incidence as a function of age in the 1890-99 EAM birth cohort. As there were significant improvements in CRC detection and treatment during the lifespan of this cohort, one is not surprised to note that the area under the curve has increased to about 0.28. Note that these calculations have been made for males based only on the assumption that males and females have the same inherited risks, but that physiological differences, such as number of cells at risk of initiation, lead to lower lifetime risks in females than males. Thus, our estimates of mLR are based on male data alone.

Figure 6 depicts the data forwarded by Prof. K.Hemminki for incidence of CRC in male parents by first-time diagnosis in Swedish hospitals. In this case, the estimate of the area under the curve needs no correction for survival, and leads to an estimated minimum fraction of males at lifetime CRC risk of ~0.26. The rough estimates of minimum lifetime risk of CRC in U.S. males (0.28) and Swedish male parents (0.26) are remarkably similar, and as we will use estimates of familial risk from the Swedish population, we will adopt *pro tempore* the value of 0.26 as the minimum estimate of the Swedish population at lifetime risk of CRC.

Another set of data that may be used to try to estimate the lifetime fraction at risk of colorectal cancer in European males is the study of the fraction of males in the United Kingdom with one or more adenomatous polyps in a study organized by Dr. W. Atkin of St. Mark's Hospital, London (Atkin, Rogers et al. 2004). Among males, this study found an average of 15% of males over 50 with one or more adenomatous polyps or a more advanced stage of CRC. The average number of adenomatous polyps among males with polyps was 1.3. Were said polyps Poisson-distributed among persons at risk of polyp development and presumably at risk of CRC, then by the formula $1.3 = \lambda/(1- e^{-\lambda})$, one may estimate the average number of polyps among persons at risk as $\lambda \sim 0.55$. This permits conversion of the fraction with polyps, 0.15, into an estimate of the fraction at risk of polyp development of $0.15/(1- e^{-\lambda}) \sim 0.35$. The author opined that the fraction with polyps may be actually larger than 0.15, with a bias due to lack of experience or expertise among proctologists.

From these three examples we see that a minimum value of CRC lifetime risk of 0.26 (Sweden) and 0.28 (U.S.) is smaller than the estimate of 0.35 based on serial proctoscopic examinations in the United Kingdom. This latter fraction might be considered an estimate of actual lifetime risk, insofar as it comprised observations in men under 65, when competing forms of death would not interfere with observations. Because of the smaller population scanned in the Atkin study, there is a greater uncertainty as to the derived estimate of *mLR*. We must be satisfied to consider the *mLR* in European males for CRC as greater than 0.26 and therefore a minimum estimate of *GRR* is also about 0.26. As the *mLR* estimate is an intersection of genetic, environmental and stochastic risks, these data do not define an upper limit for *GRR*.

Confounding variables affecting the estimate of lifetime risk; pleiotropic risk

Interference with other forms of death with risk factors unrelated to those of CRC would not interfere with the estimate of minimum lifetime risk in the Swedish and U.S. datasets as they are comprised of the conditional probabilities of detection or death by CRC, given

that an individual is still alive. However, another possibility to be accounted is that the same genetic risks may underlie several different forms of cancers, *i.e.* mutations in one gene might confer risk for cancers in different tissues, a condition of pleiotropic risk. This is known for breast and ovarian cancers, and suspected for several other combinations, as revealed by the Hemminki group's efforts to identify familial cross-sensitivity. Cure or avoidance of one form of such a set of cancers would in such cases be expected to be accompanied by increases in the related forms with common genetic etiologies. The presence of such conditions would cause the area under the incidence curve to be an underestimate of the lifetime. This confounding possibility is algebraically addressed by Herrero-Jimenez et al. (1998, 2000), and is formally addressed in the section **Cancer Models.**

Environmental risk

We argue that the lifetime risks for late-onset cancers must be functions of environmental and genetic risk factors acting independently or in concert. We further argue that inherited risk factors could not have changed to any significant degree in the U.S. population in the latter part of the 19^{th}-century, because the wave of immigration consisted mainly of European origins genetically similar to the earlier European immigrants. Actual gonadal rates of genetic change of ~ 3×10^{-6} gene inactivations per generation would be glacial in comparison to the observed rate of changes in lifetime risk (Sankaranarayanan and Chakraborty 2000).

If genetic risk factors did not change, then all changes in lifetime cancer risk in the birth decade cohorts of the latter 19^{th}-century must be ascribed to known or unknown environmental factors. It is prudent, however, to maintain some skepticism that under-diagnosis and under-reporting in the early 20^{th}-century may have exaggerated the apparent magnitude of the historical increases in some cases.

Key to the approach that we take in this thesis is the premise based on the data of Figures 1 and 2 that whatever the set of environmental risk factors may have been for any or all mortal cancers, the condition of environmental risk was established at a historical period that comprehended the birth decade cohorts of and subsequent to 1900-1909 and that the fraction(s) of the population at lifetime environmental risk(s) has not changed significantly for persons in the U.S. born since 1900.

Relevant to these arguments are the data of the Swedish Family Cancer Database, organized and analyzed by the Hemminki group over the past decade. As in the U.S.A., their data recorded for persons born since 1932 and diagnosed with a malignant neoplasia since 1958, demonstrates no change in age-specific incidence in the birth decade cohorts beginning with 1930-1939 through the present day. These observations apply to the twenty most common forms of adult cancers, as well as the sum of all cancers. Some cancer data recorded in Sweden after 1958 for the oldest Swedes indicates that a rise in cancer rates occurred in birth decade cohorts of the late 19^{th}-century similar to that observed in the U.S.A. (Figure 1), but the data are not sufficient to permit more than this general conclusion. What is important about the findings, such as the data of Figure 2 for the U.S.A. and Sweden, is that the age-specific cancer rates are essentially unchanged for the birth decades after 1900-1909; this permits us to consider risks of cancers in the 20^{th}-century birth cohorts as being constant with regard to environmental risk. Thus, one may regard the Swedish cohorts of "parents" and "children" as having experienced essentially identical conditions of environmental risk.

Epidemiologic studies have sought but not found evidence of spousal risk for colorectal cancer (Hemminki, Dong et al. 2001; Hemminki and Jiang 2002). Shared environment in adulthood does not appear to play a role in differential risk for most cancers, including colorectal cancer. A comparison of CRC mortality rates among American communities (1958-1997), whether urban or rural, by J.A.Vatland (Analysis of community cancer mortality rates, MIT Ph.D. Thesis, 2001) did not find differences significantly greater than those expected by chance alone. See Figure 7 for the case of colon cancer mortality among 520 communities in the U.S. Commonwealth of Pennsylvania.

The importance of this logical point requires emphasis. It means that we may reasonably consider familial risks to arise solely from genetic risk for the birth year cohorts of the 20^{th}-century. It is tempting to imagine that the homogeneity of cancer risks among communities discovered by J.A. Vatland in the latter decades of the 20^{th}-century in America indicates a saturation of the population with the unknown environmental risk factors with all or nearly all of the population at uniform environmental risk (E uniform, though of unknown value), and that all persons now experience unknown uniform kinds of environmental risk. But this assumption is not necessary for unambiguous determination of the degree of familial risk for common cancers, using data from the Swedish Family Cancer Registry as formally derived below in the section ***FRR(t)***.

Development of Mathematical Models of the Physiological Processes of Carcinogenesis

Our next line of logical construction arises from the fifty year-old process of creating mathematical models for the deterministic and stochastic processes hypothesized to underlie carcinogenesis in humans. In general, such models are based on the "two-stage" model first put forth by Armitage and Doll (1957), in which the first stage, "initiation", consisted of "n" mutations required to change a phenotypically normal cell "cell at risk" into a cell that could give rise to a slowly growing preneoplastic colony, in which a preneoplastic "cell at risk" could over time accumulate "m" mutations that would transform a preneoplastic cell at risk into a neoplastic cell, that would give rise to a rapidly growing tumor that untreated, kills within a few years of neoplastic transformation. In this model "progression", the events that occur during growth of the tumor, may accelerate tumor growth and death but as they inexorably occur within a short period of years and are essentially ignored.

While a number of mathematical models have attempted to build on Armitage and Doll (1957), two in particular influenced the present mode of analysis.

The first is that of Stein and Stein (1990) and Stein (1991), in which it was noted that the apparent constant slope of the log-linear plot of most late-onset cancer mortality data had a constant value of ~ 0.11. The second is that of Herrero-Jimenez et al. (1998, 2000), in which this constant exponential increase was ascribed to a constant growth rate of preneoplastic colonies, which was recognized as being approximately equal to the growth rate of human juveniles from 18 months to maturity. In Gostjeva and Thilly (2005), the argument was extended to note that the expected growth rate of epithelial sheets in juvenile organs was ~ 0.11, the same rate noted by the Steins as the slope of the age-specific increase in cancer mortality rates.

Herrero-Jimenez et al. (1998, 2000) attempted to apply linear approximations to derive estimates of initiation and promotion rates that were arguably functions of mutation rates, as well as to estimate the growth rate of preneoplastic colonies, such as adenomatous colonic polyps. In this effort they misled themselves into believing a true computational minimum had been discovered, and their specific estimates of initiation and promotion mutation rates and population risk parameters in CRC are now seen as but one set of estimates of many possible sets (Gostjeva and Thilly 2005). However, this effort introduced the only extant treatment of populations as containing persons of differing genetic risk, and also the first model to account for the possibility that different forms of cancer shared common environmental and or genetic risk factors.

Other cancer model variants have been offered, such as those of Luebeck and Moolgavkar (2002), Michor et al. (2004) in the Nowak group, and Frank (2004). These efforts have, however, explicitly or implicitly assumed that all members of the population are at risk (Luebeck and Moolgavkar 2002), or attempted to model the hypothesis that oncogenic mutation rates accelerate during growth of preneoplastic colonies. Insofar as the studies of the Hemminki group and others with smaller population sizes have established a clear demonstration that familial cancer risks exist for common forms of cancers and the obvious division of lung cancer risk among persons who do or do not smoke cigarettes, we reject the modeling approaches, which, inexplicably to me, assume

that all individuals are at equal cancer risk. With regard to modeling to explain the level of loss of heterozygosity and point mutations in tumors, these matters seem to me best remanded to the time in which direct measurements in human tissues will permit distinguishing between hypotheses that assume serial mutator mutations (Beckman and Loeb 2006) or a continuation of a hypothesized juvenile mutator phenotype throughout preneoplasia (Gostjeva and Thilly 2005).

With regard to the many phenomena that might pertain to the effects of unknown or known environmental risk factors they are beyond the scope of my effort insofar as the overall cancer rates have remained unchanged in the birth cohorts of the 20th-century. Of course, it must be recognized that inherited conditions may well define sensitivity to these known or unknown environmental risk factors. But whether genetic variation in pharmacokinetic parameters, metabolism of specific environmental agents, reactions with specific macromolecules or cellular and physiological responses to damage modulate oncogenic mutation or preneoplastic growth rates cannot be distinguished among the set of possible reasons for genetic risk for common diseases. Our effort is independent of the arguments and data provided by Prof. W.G. Thilly, about causes of human somatic mutations or their time of occurrence in a human lifetime (Muniappan and Thilly 2002; Thilly 2003; Gostjeva and Thilly 2005; Zheng, Khrapko et al. 2006). Our goal is simply to discover if, among the myriad of genetic possibilities, the hypothesis that familial risk for a common cancer, *e.g.* CRC, is monogenic is excluded by existing observations.

Carcinogenesis Model

The carcinogenesis model we employ is a variant derived from Herrero-Jimenez' collaboration with Prof. S. Morgenthaler and Prof. W.G. Thilly modified since 2000 by their interactions with students and other researchers, especially Dr. E.V. Gostjeva. A differential equation has been developed that, for each age t, accounts for the number of expected cancer deaths, in which n oncomutations occurring in presumptive stem cells early in life create preneoplastic colonies that grow at juvenile growth rates and mutation

(mutator) rates until a single preneoplastic stem cell has experienced *m* required onco-events and gives rise to a clonal, rapidly growing lethal tumor. Their function $P_{OBS}(h,t)$ represents the expected chance of cancer incidence in a cohort such as Swedish males born in year h and observed at age *t*. The elements of this function need not be defined or defended for my purposes. It need only be noted that the same function was expected to apply to all persons at risk born in Sweden or the U.S. since 1900 and thus apply equally to cohorts of "parents" and "children" identified in the Swedish Family Cancer Database.

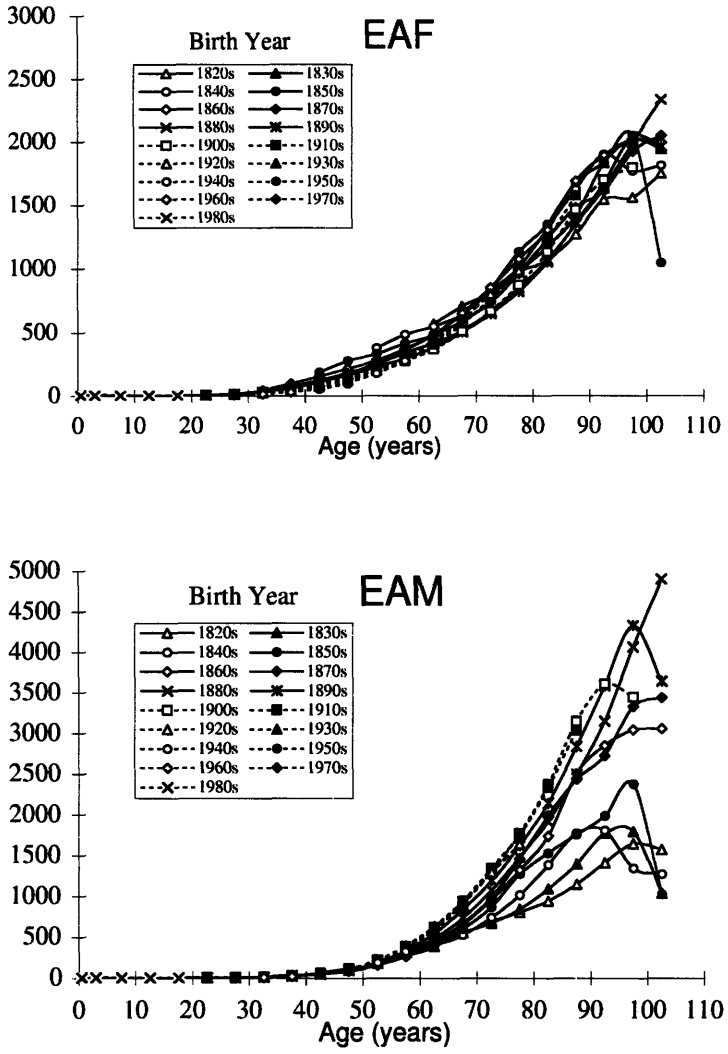


FIGURE 1: (a) The age-specific mortality rate, *OBS(t)* from all malignant neoplasms recorded between 1900 and 1997 for European American males. *OBS(t)* is calculated as the observed number of deaths recorded in each calendar year in successive quinquennial age-intervals, 0-4, 5-9,...,100-104, divided by the product of the number of persons alive of that age interval and the fraction of persons surviving death by any cause in that calendar year. *OBS(t)* is expressed as deaths from the observed disease, here all malignant neoplasms, per 100,000 population. (b) *OBS(t)* for European American females. These data are available at http://epidemiology.mit.edu.

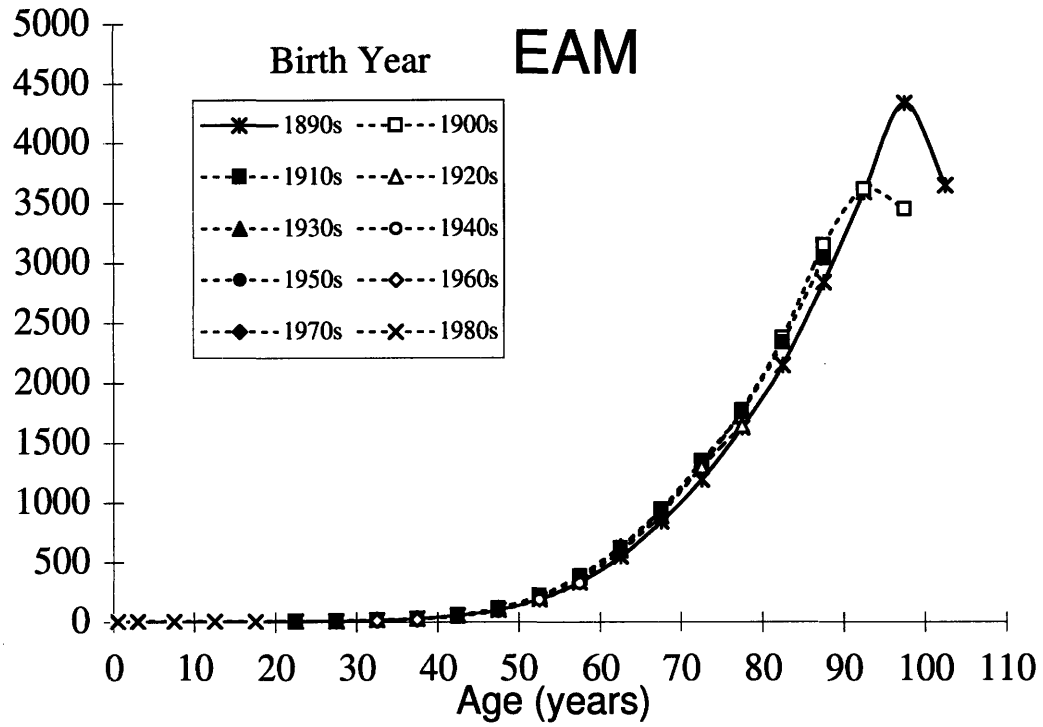
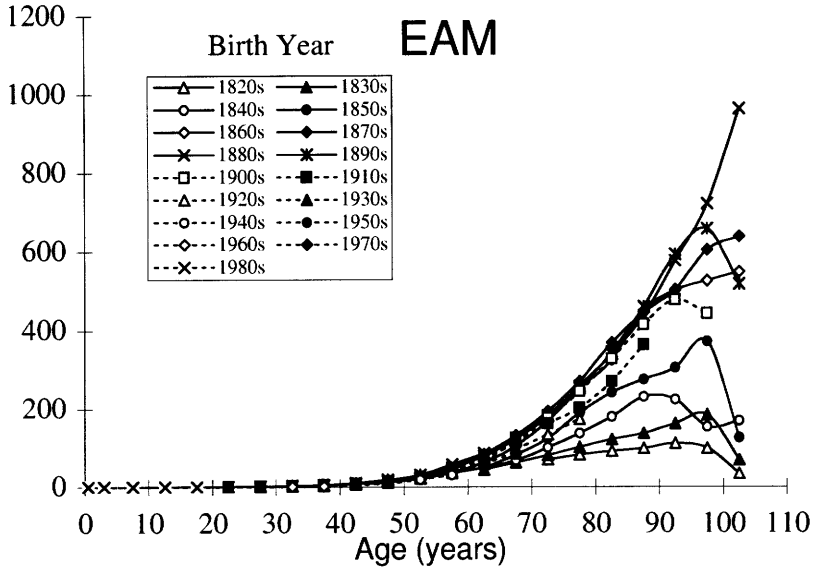


FIGURE 2: The age-specific mortality rate, *OBS(t)* from all malignant neoplasms recorded between 1900 and 1997 for European American males born in the decades 1890-1899, 1900-1909, ..., 1980-1989. These data support an interpretation that the many environmental changes of the 20th-century have had but minor effects on the overall mortality rate from all malignant neoplasms, and that cohorts of parents and children born in the 20th-century may be considered to have been at essentially identical environmental risk.



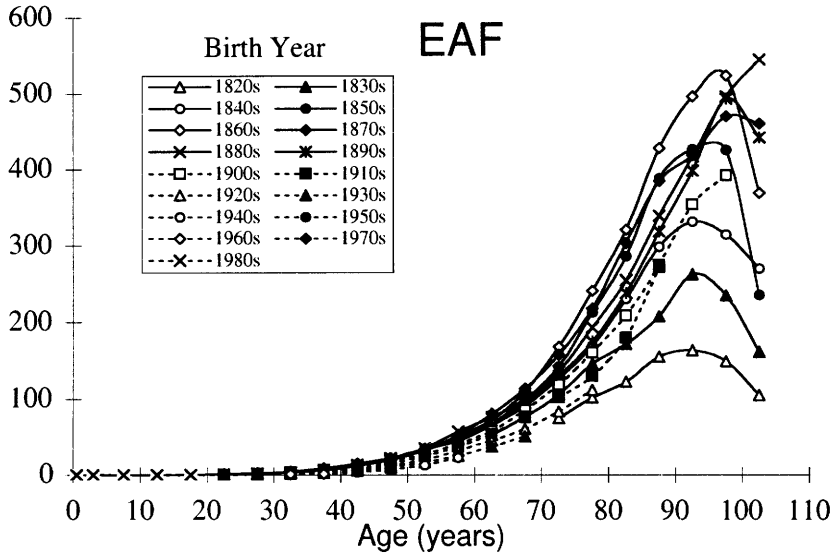


FIGURE 3: (a) The age-specific mortality rate, *OBS(t)* from all cancers of the lower digestive tract that are comprised to an extent of ~95% by colorectal cancers recorded between 1900 and 1997 for European American males. *OBS(t)* is calculated as the observed number of deaths recorded in each calendar year in age-intervals, 0-4, 5-9,...,100-104, divided by the product of the number of persons alive of that age interval and the fraction of persons surviving death by any cause in that calendar year. *OBS(t)* is expressed as deaths from the observed disease per 100,000 population. (b) *OBS(t)* for European American females. These data are available at http://epidemiology.mit.edu.

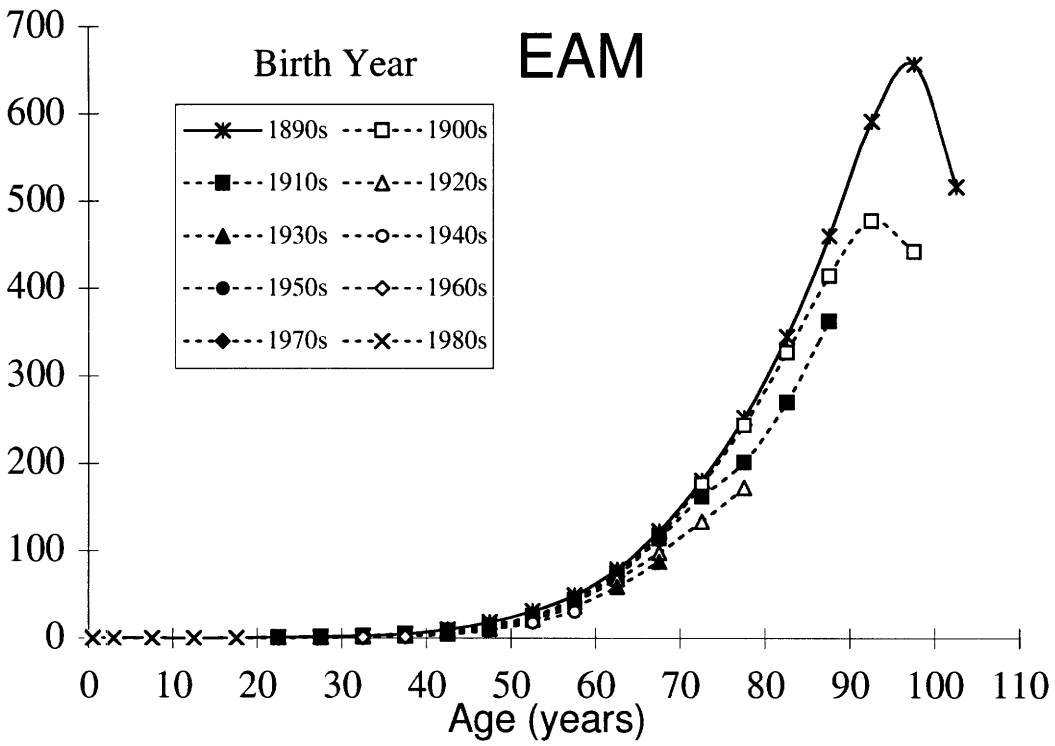


FIGURE 4: The age-specific mortality rate, *OBS(t)*, from neoplasms of the lower digestive tract recorded between 1900 and 1997 for European American males born in the birth decades 1890-1899, 1900-1909,...,1980-1989. Maximum rates of the birth decade cohorts of 1880-1889 and 1890-1899 declined in each successive decade, a statistically significant decrease ascribable in part to a decreasing autopsy rate and advances in medical practice.

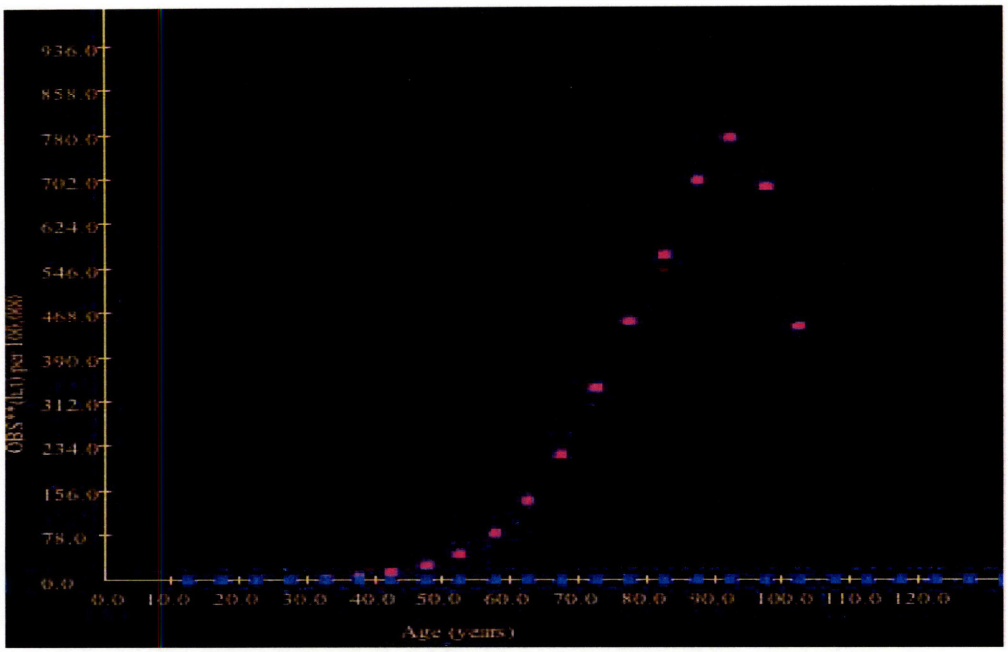


FIGURE 5: The age-specific mortality rate, *OBS(t)*, from neoplasms of the lower digestive tract recorded between 1900 and 1997 for European American males born in the decade 1890-1899 and corrected for historical increases in survival after diagnosis, *S(h,t),* as in Herrero-Jimenez et al. (1998, 2000). Note that age-specific values are greater than those for the crude mortality data of Figure 4.

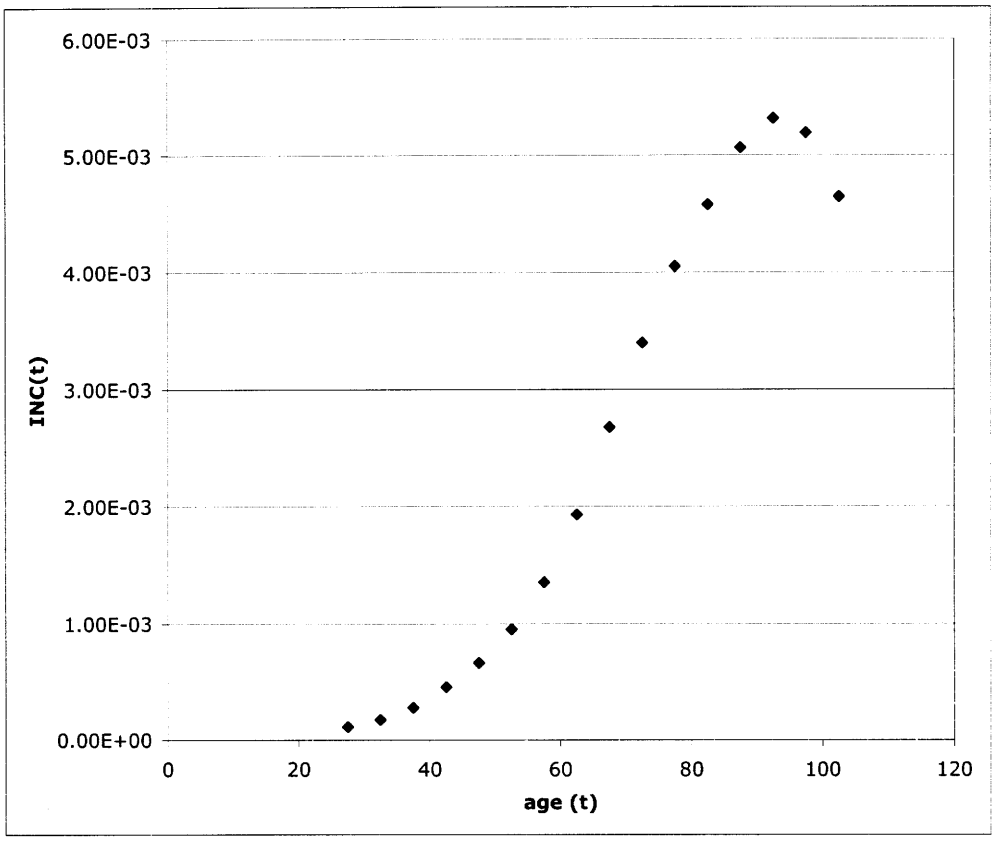


FIGURE 6: CRC incidence recorded as first time diagnoses of primary CRC among Swedish fathers whose children were born 1932-2002 (Swedish Family Cancer Database. Communicated by Prof. K. Hemminki, and prepared by Prof. W.G. Thilly).

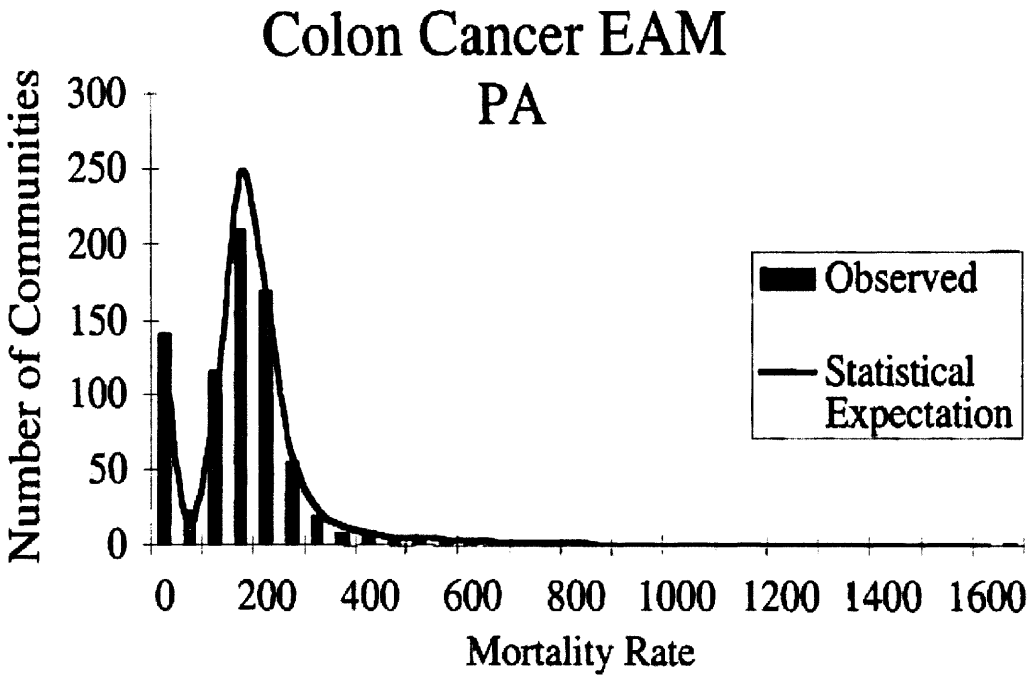


FIGURE 7: Distribution of colon cancer deaths among 520 Pennsylvania communities recorded 1958-1997 among European American males aged 65-84. Histogram indicates the observed distribution of mortality rates among communities whereas the continuous black line indicates the sum of binomial distributions expected by chance for each community. The expected and observed distributions were not significantly different (Kolmogorov-Smirnov test) indicating that neither genetic nor environmental risk factors for colon cancer differed significantly among these communities during the period of observation (J.A. Vatland, MIT Ph.D. Thesis, 2001).

METHODOLOGY

Definition of terms

h: calendar year of birth

t: biological age at diagnosis

$y = h + t$: calendar year of diagnosis

0: subscript referring to the population of all parents or all children in Sweden (1958-2002)

1: subscript referring to the subpopulation of all children with parents diagnosed with CRC in Sweden (1958-2002)

N_h: number of persons born in calendar year h

$TOT(h,t)$: total mortality rate for birth year h and age t

$INC(h,t)$: crude age-specific incidence rate for birth year h and age at diagnosis t, *i.e.* number of persons diagnosed with CRC for h and t, divided by the number of persons alive for h and t

$SUR(h,t) = S(h,t)$: fraction surviving observed disease after 5-yr interval

$REP(h,t) = R(h,t)$: fraction of reported deaths from specified causes

$P_{OBS}(h,t)$: expected chance of CRC incidence within group at risk

$F(h,t)$: population fraction at primary risk for late-onset CRC

$E(h,t)$: population fraction at environmental risk for late-onset CRC

$G(h,t)$: population fraction at genetic risk for late-onset CRC

f_h: fraction of persons at risk who would die of CRC in a fictitious population which is at risk only from CRC and diseases sharing risk factors with CRC; f_h increases, as connected mortality decreases; f_h accounts for competing forms of death, if any, with the same risk factors as CRC

'*late-onset*': diagnosis at $t \geq 50$

The data and assumptions about the values of the parameters that underlie age-specific cancer incidence

For individuals in the Swedish Family Cancer Database (1958-2002), the values of $INC(h,t)$ have been unchanged as of $h \geq 1900$ for CRC. Thus, it is assumed that the carcinogenic parameters are not historically variant, but constant, since 1900. For simplicity, we drop the designation for birth year h, and combine CRC data within age intervals t, regardless of h. The terms simplify to $INC(t)$, $TOT(t)$, $S(t)$, $R(t)$, $P_{OBS}(t)$, $F(t)$, $E(t)$, $G(t)$, and f.

In this treatment, we assume that all persons at risk from CRC in the general Swedish population are members of a single population at risk having identical carcinogenic parameters $S(t)$, $R(t)$, $P_{OBS}(t)$ and f. In particular, a child with late-onset CRC is assumed to be a member of this population whether CRC is observed in either of its parents or not. This assumption is supported by the observation that for the period $50 \leq t < 74$, the age-specific CRC incidence rates are identical for the parents and children in the general population.

$TOT(t)$ is the total number of deaths in 5-year age intervals, 0-4, 5-9,...,90-94, divided by the number of persons alive in the same year in calendar year $y = h + t$. It has been introduced in the modeling equation to account for the expectation that other independent forms of death compete with the observation of CRC, in particular given the obvious significant increases in death rates in senectitude. Death rates from CRC would be but a small fraction of total deaths in any interval, so that the differences between CRC rates in those at risk, and those not at risk, are small, and can be ignored.

The observed diagnostic rate, corrected for competing forms of death, for birth year h and age at CRC diagnosis t, is (Herrero-Jimenez, Thilly et al. 1998; Herrero-Jimenez, Tomita-Mitchell et al. 2000):

$$\frac{INC(h,t)}{1-TOT(h,t)} = \frac{\#DIAGNOSED(h,t)}{\#ALIVE(h,t)\cdot(1-TOT(h,t))}$$

$$= \frac{\left[F_h \cdot (1-S(h,t)) \cdot R(h,t) \cdot P_{OBS}(h,t) \cdot e^{-\frac{1}{f_h}\int P_{OBS}(h,t)(1-S(h,t))dt}\right] N_h}{\left[F_h \cdot e^{-\frac{1}{f_h}\int P_{OBS}(h,t)(1-S(h,t))dt} + (1-F_h)\right] N_h}$$

Simplifying and rearranging, this expression reduces to

$$\frac{INC(t)}{1-TOT(t)} = \frac{F \cdot (1-S(t)) \cdot R(t) \cdot P_{OBS}(t)}{F + (1-F) \cdot e^{\frac{1}{f}\int P_{OBS}(t)(1-S(t))dt}}$$

This function provides fits with summed statistical variances equal to the sum of variance for the observed values of *INC(t)* for most but not all common forms of cancer using data from Sweden, the U.S. or Japan. Breast cancer is one exception, where the data suggest superposition of incidences of two separate forms, arising in the mid 20s and mid 50s (P. Herrero-Jimenez 2000, MIT Ph.D.Thesis).

Application of the function INC(t) to calculation of FRR(t); The importance of independent definition of INC(t) according to gender

As shown in Figure 3 the age-specific mortality for cancers of the lower gastrointestinal tract differ significantly between men and women. However, most familial data analyses combine data for fathers and mothers as "parents" and sons and daughters as "children". This creates an algebraic bias that needs to be addressed in order to reach an unbiased estimate of genetic relative risk.

Consider letting the incidence for fathers be the number of paternal cases at age *t* be "A" and the number of living fathers of that age be "B"; let the appropriate terms for mothers be "C" and "D". When expressed as *INC(t)* as parents the term is calculated as (A+B)/(C+D) when in fact the average rate among "parents" is (A/B + C/D)/2, which terms are not equal to each other. As paternity is not established in an unknown percentage of cases, which could be about 10% for children in the Swedish Family Cancer Database, the fact that B<D adds another bias in the use of (A+B)/(C+D) as an approximation of *INC(t)*.

Another source of error in combining male and female data arises from the expectation that the larger size of males provides them in the teenage years with a larger number of tissue stem cells at risk of initiation, and thus larger lifetime values of $P_{OBS}(t)$. Assuming that males and females carry identical genetic risks, and experience nearly identical environmental risks, the only term in the driving function for *INC(t)*, $FP_{OBS}(t)$, is $P_{OBS}(t)$.

This point may be illustrated by considering the expected concordance of observed disease in the sons or daughters of fathers or mothers with late-onset CRC. Sons would appear to "inherit" CRC risk at a higher level than daughters because their values of *INC(t)* are higher than their sisters' values.

It is clear that calculations of parent-to-child transmission on the basis of values of *INC(t)* must be refined on the basis of father-to-son or mother-to-daughter transmission. The ratios of age-specific incidences corrected for age-specific total mortality (which also differs between men and women) should yield identical estimates of *FRR(t)* for both genders.

In the following calculations and discussions, it is assumed that parent-to-child relationships are defined as mother-to-daughter or father-to-son, anticipating the future use of the Swedish Family Cancer Registry in this form. Father-to-son transmission will be used in the examples.

The corrected CRC diagnostic rate for fathers whose sons were diagnosed with CRC at $t \geq 50$ is

$$\frac{INC_1(t)}{1-TOT_1(t)} = \frac{F_1 \cdot (1-S_1(t)) \cdot R_1(t) \cdot P_{OBS\,1}(t)}{F_1 + (1-F_1) \cdot e^{\frac{1}{f_1}\int P_{OBS_1}(t)(1-S_1(t))dt}},$$

while the corrected CRC diagnostic rate for all fathers (as well as all sons) in the population for $t \geq 50$, is

$$\frac{INC_0(t)}{1-TOT_0(t)} = \frac{F_0 \cdot (1-S_0(t)) \cdot R_0(t) \cdot P_{OBS\,0}(t)}{F_0 + (1-F_0) \cdot e^{\frac{1}{f_0}\int P_{OBS_0}(t)(1-S_0(t))dt}}$$

The ratio of the observed incidence rates is

$$\frac{INC_1(t)}{INC_0(t)} = \frac{1-TOT_1(t)}{1-TOT_0(t)} \cdot \frac{F_1 \cdot (1-S_1(t)) \cdot R_1(t) \cdot P_{OBS\,1}(t)}{F_0 \cdot (1-S_0(t)) \cdot R_0(t) \cdot P_{OBS\,0}(t)} \cdot \frac{F_0 + (1-F_0) \cdot e^{\frac{1}{f_0}\int P_{OBS_0}(t)(1-S_0(t))dt}}{F_1 + (1-F_1) \cdot e^{\frac{1}{f_1}\int P_{OBS_1}(t)(1-S_1(t))dt}}$$

Since 1930, the values of *R(t)* have been approximately 1. Survival *S(h, t)*, like *f*, is assumed to be approximately equal for the two subpopulations. The ratio of the observed incidence rates thus simplifies to

$$\frac{INC_1(t)}{INC_0(t)} \approx \frac{F_1}{F_0} \cdot \frac{F_0 + (1-F_0) \cdot e^{\frac{1}{f}\int P_{OBS_0}(t)dt}}{F_1 + (1-F_1) \cdot e^{\frac{1}{f}\int P_{OBS_1}(t)dt}}$$

This ratio has interesting properties for the lowest and highest values of $\frac{1}{f}\int P_{OBS}(t)dt$.

It increases asymptotically to a maximum, as $\frac{1}{f}\int P_{OBS}(t)dt$ increases with t.

The limit of this ratio as $\frac{1}{f}\int P_{OBS}(t)dt \to \infty$ is a constant for given values of F_1 and F_0.

$$\frac{INC_1(t)}{INC_0(t)} \approx \frac{F_1}{F_0} \cdot \frac{(1-F_0)\cdot e^{\frac{1}{f}\int P_{OBS_0}(t)dt}}{(1-F_1)\cdot e^{\frac{1}{f}\int P_{OBS_1}(t)dt}} \to \frac{F_1}{F_0} \cdot \frac{(1-F_0)}{(1-F_1)}$$

For example, with $F_0 = 0.10$, and $F_1 = 0.25$, whereby $F_1 / F_0 = 2.5$,

$$\frac{INC_1(t)}{INC_0(t)} \to \frac{0.25}{0.10} \cdot \frac{(1-0.10)}{(1-0.25)} = 3.0$$

While algebraically interesting, values of $\frac{1}{f}\int P_{OBS}(t)dt$ do not approach infinity in the longest human lifespan, so that the value of $INC_1(t) / INC_0(t)$ is not expected to reach a stable maximum in any father/son or mother/daughter comparisons.

Of more practical interest to us is the behavior of $INC_1(t) / INC_0(t)$ for lower values of t, for which the integral becomes vanishingly small, $\frac{1}{f}\int P_{OBS}(t)dt \to 0$.

If $\int P_{OBS}(t)dt \approx 0$, and $f \approx 1$, then $e^{\frac{1}{f}\int P_{OBS_0}(t)dt} \approx 1$, and

$$\frac{INC_1(t)}{INC_0(t)} \approx \frac{F_1}{F_0} \cdot \frac{F_0 + (1-F_0)\cdot e^{\frac{1}{f_0}\int P_{OBS_0}(t)dt}}{F_1 + (1-F_1)\cdot e^{\frac{1}{f_1}\int P_{OBS_1}(t)dt}} \to \frac{F_1}{F_0} \cdot \frac{F_0 + (1-F_0)}{F_1 + (1-F_1)} = \frac{F_1}{F_0}$$

Note that all stochastic processes embodied in the two-stage carcinogenesis model are embodied in the term $e^{\frac{1}{f}\int P_{OBS_0}(t)dt}$, the probability of initiation in any potential age "window", the competition with mortal diseases with shared environmental and genetic risks, and the probability that any preneoplastic cell at risk of transformation may transform at any age after initiation. Thus, the stochastic processes that might bias the estimate of F_1/F_0 are found in both the denominator and numerator, given that father to son or mother to daughter genetic transmissions are rigorously defined. This important algebraic "canceling out" of these complex terms is clearly dependent on same gender genetic transmission of risk as *INC(t)* differs significantly between males and females presumably because of differences in the values of $e^{\frac{1}{f}\int P_{OBS_0}(t)dt}$.

For colorectal cancer, $f \approx 1$ appears to be a valid assumption with regard to risk from other common cancers. The range of values of f was based on findings of limited risk sharing between CRC and other cancers (Gruber, Ellis et al. 2002; Meijers-Heijboer, Wijnen et al. 2003; Hemminki and Chen 2004; Hemminki and Chen 2004). There is no current knowledge about shared risks with non-cancer fatal diseases, such as vascular diseases, which account for nearly half of all current deaths from persons over 50. We thus assume *pro tempore* that there are no significant shared inherited risks for cardiovascular diseases and CRC, and that the assumption $f \approx 1$ holds.

In an independent study relevant to this assumption Prof. W.G. Thilly has explored possible values of $\frac{1}{f}\int P_{OBS}(t)dt$ for values of $f = 0.8 - 1.0$ using the estimated CRC *mLR* for Swedish males of >0.26 and the Java-based program CancerFit4_1.xls developed and employed by his group. His communication is given here:

" *The observations of Atkin you cite indicate that males with adenomatous polyps have about 1.3 polyps per person and I believe you have correctly estimated the fraction of males at risk of polyps to be about 1- $e^{-0.55}$ = 0.427. This estimate of the fraction of males at risk of tumor initiation that experience tumor initiation fixes the initiation rate*

parameter, C_{init}, at a value of about 0.005. I have calculated that this value would lead to an inactivated mutant fraction for a gene in which CRC initiation mutations occur, such as APC, to be at about 2 x 10^{-4} at maturity. This estimate is in accord with the observation of several thousands of polyps in adults with FAPC syndrome and 6-10 x 10^{6} colonic crypts. Thus I have used this value, C_{init} 0.005, and the estimate of a constant preneoplastic growth rate of ~0.1-0.15 , 0.26<F<1.0 and 0.6<f<1.0 to determine if your statement that in the age intervals 50-54, ..., 70-74 the term $(1/f)\int P_{OBS}(t)dt$ is small, approximating zero for Swedish or European American males. Applying this estimate to the Swedish parents data, we note that values of $P_{OBS}(t)$ for all t are essentially invariant with F and f, and thus that $\int P_{OBS}(t)dt$ may be calculated for all values of t. For $t < 75$, $\int P_{OBS}(t)dt \approx 0.035 \approx 0$. Your assumption in this regard appears to be justified."

The finding that $\int P_{OBS}(t)dt \approx 0.035 \approx 0$ for t<75 was crucial for the ratios $INC_1(t)/INC_0(t)$ at t = 50-54, 55-59, 60-64, 65-69, and 70-74 to each be app. equal to $F_1/F_0 = FRR(t; 50 \le t < 75)$. The validity of the approximation having been tested and delimited, we conclude, subject to uncertainty in the potential interactive role of cardiovascular disease in these age intervals, that the ratio of the CRC incidence rates for quinquennial age intervals within $50 \le t < 75$, is approximately equal to the ratio of the population fractions at risk, *i.e.* the familial relative risk for late-onset disease.

$$\frac{INC_1(t)}{INC_0(t)} \approx \frac{F_1}{F_0} \cdot \frac{F_0 + (1-F_0)\cdot 1}{F_1 + (1-F_1)\cdot 1} = \frac{F_1}{F_0} = FRR(t; 50 \le t < 75)$$

Let's note that $FRR(t; 50 \le t < 75)$ may not be reasonably called 'late-onset' $FRR(t; \text{late-onset})$, insofar as these age intervals are strictly defined, but an early subset of 'late-onset'.

Familial risk may arise from shared genetic background and/or shared environmental experiences. We address etiologic apportionment in the following section.

Accounting for Environmental Risk

The historical increase in CRC rates occurred in narrow historical time windows. For birth cohorts in the United States and Western Europe, including Sweden, after 1890-1900, incidence and mortality data for colorectal cancer indicate that peak age-specific rates have reached a stable plateau, a local maximum in mathematical terms. The changes that led to these increases were likely to be environmental and not genetic, since population genetic changes take longer on the evolutionary scale. From some historical time on, the environment is having a uniform effect, and does not account for risk differences.

These considerations suggest that the persons registered in the Swedish Family Cancer Database (1958-2002) experienced homogeneous environmental risk, throughout the historical period of their birth cohorts. Lung cancer is, of course, exempted from this assumption. In quantitative terms, the fraction of subpopulations at environmental risk is the same, $E_0 = E_1 = E$. In particular, the fraction at environmental risk is the same for those with and those without family history of a particular form of cancer. Since the fraction of a population at overall risk comprises the intersection of the fractions at environmental and genetic risk, $F_1 = EG_1$ and $F_0 = EG_0$, it follows that the late-onset familial relative risk is approximately equal to the genotype or genetic relative risk, $FRR(t; \text{late - onset}) = F_1 / F_0 = G_1 / G_0 = GRR$. Therefore, the differences between the general population and the subpopulation with a family history of common cancer are reasonably attributed to differences in genetic makeup.

$$FRR(t; \text{late - onset}) = \frac{F_1}{F_0} = \frac{E_1 \cdot G_1}{E_0 \cdot G_0} \approx \frac{E \cdot G_1}{E \cdot G_0} = \frac{G_1}{G_0} = GRR$$

$$\frac{INC_1(t; \text{late - onset})}{INC_0(t; \text{late - onset})} \approx FRR(t; \text{late - onset}) \approx GRR$$

Original to this thesis is the recognition of the importance of employing strictly gender-specific data and age-specific data comprising ages 50-54,...,70-74, in calculating the familial risk for late-onset common diseases such as CRC. Combination of this advance with the observations of Dr. J.A. Vatland (no difference in CRC or any other cancer risk among communities in any of several U.S. states), the Hemminki group (no spousal risk for CRC or most other cancers) and the observation that age-specific risk for CRC and most other cancers has remained invariant in the U.S. and Sweden for birth year cohorts since 1900, have permitted us to conclude that there are no generational differences in environmental risk in the Swedish Family Cancer Database, and that $FRR(t; \text{late - onset})$ for father-to-son and mother-to-daughter transmission of risk, yields the desired estimates of genetic relative risk, *GRR*.

Comments: Age-specific Familial Relative Risk, *FRR(t)*, in Sweden

The Swedish Family-Cancer Database (1958-2002) provides age-specific incidence rates for individual forms of cancers in the general population of all children born since January 1, 1932, and their biological parents. These two age-specific cancer incidence rates are practically identical for age intervals up to 65-69, the oldest interval for which data from "children" born 1932-1936 are available.

For colorectal cancer and many other cancers, these age-specific incidence curves rise exponentially from teenage years into the 70s, reach a maximum around 90, and decline significantly by the age interval 100-104. The decline in incidence in old age was clinically confirmed by serial autopsies of deaths in the Swedish state of Malmö, for about twenty to twenty-five years.

As noted above the minimum lifetime risk, *mLR*, may be calculated from the age-specific incidence data for any disease as the area under the incidence curve from age zero to a theoretical age when incidence rates would have declined to zero (Herrero-Jimenez, Thilly et al. 1998; Herrero-Jimenez, Tomita-Mitchell et al. 2000). It is a composite function of the fractions of the population at genetic (*G*), environmental (*E*), and stochastic risk, as well as any fatal condition that shares the risk factors of the observed disease – the relevant parameter here being *f*, which accounts for mortality from other diseases sharing the risk factors of the observed disease. Multistage models of carcinogenesis allow estimation of *F*, but absent independent determinations of *E* and *f*, values of the fraction at genetic risk (*G*), cannot be determined from the age-specific incidence data alone.

For colorectal cancer, the area under the curve for combined data of Swedish males and females is about 0.26 as calculated from the area under the curve of Figure 6 above. The fraction at environmental risk (*E*), is as noted, apparently uniform among successive birth cohorts in the present-day Swedish population. Supporting evidence for this assertion comes from (i) *mLR* having reached a stable maximum since the birth cohort of 1900, (ii) lack of spousal risk for colorectal cancer in Sweden and (iii) no apparent risk sharing between colorectal cancer and many other common forms of mortality. *E* may or may not be close to 1. In the case of $E < 1$, familial genetic risk as defined above can still be estimated as the values of *FRR(t)* for age intervals t = 50-54,..., 70-74, since *E* is found as a factor in both the numerator and denominator in the definition of *FRR(t)*. (N.B. Spousal concordant risks have been found for late-onset stomach cancer, lung cancer, genital cancers and melanoma.)

Considered ranges for $G_0 \geq mLR$, q, and FRR (t = 50-54,...,70-74)

The values of *mLR* are higher than that of CRC for prostate, breast, and lung cancer, and lower than CRC for most other cancers. Thus, it is necessary to account for values of $G_0 \geq$ *mLR* from 0.01 to 1.00, with the corresponding values of q and *FRR* (t = 50-54,...,70-74),

explored and summarized in tables. The values of q have been considered over the entire possible range, 0 to 1.0. The values of *FRR (t = 50-54,...,70-74)* have been considered at 1.5, 2.0, 2.5, and 3.0, encompassing estimates from the Hemminki group for a wide variety of late-onset cancer forms. Acceptable values for any cancer type are those for which (i) G_0 is equal to or greater than *mLR*, and (ii) *GRR* is approximately equal to $FRR(t; \text{late-onset})$. Limits on q are not applied even though the 99% ucl on q for genes carrying non-deleterious mutations is about 0.22 (Morgenthaler and Thilly 2006). This is because with ~5000 genes that could carry non-deleterious gene-inactivating mutations some 50 would be expected to have $q>0.22$, sufficient to force consideration of monogenic risk-conferring values of q from 0.22 to 1.0.

Age-specific familial risk, *FRR(t),* for CRC and other cancers has been defined variously in the scientific literature. In some publications, it is meant to be equivalent to the integral of the age-specific incidence in parents of children or in children of parents with a particular form of cancer. For CRC in the Swedish population, the age-specific CRC incidence in children with parents diagnosed with CRC, and that in parents with children diagnosed with CRC, are identical for age intervals up to 65-69, the oldest ages currently available in the 1958-2002 version of the database, for cohorts born in or after 1932. It is clear for the available age intervals that the CRC incidence in children whose parents are diagnosed with CRC is higher than the incidence in the population of all children, creating a *prima facie* case for CRC familial risk. Given the indications that E is identical in the parental (male or female) and offspring (male or female) cohorts for CRC in Sweden, these data suggest that CRC familial risk is attributable to mainly genetic causes. A more extensive discussion can be found in Hemminki and Chen (2004).

Defining the actual numerical value of *FRR(t)* for CRC or any other disease presents a challenge, and is the focus of renewed effort in the collaboration within and between the Hemminki and Thilly groups.

A main problem is that except for sporadic efforts, *e.g.* for endometrial cancer (Hemminki, Vaittinen et al. 1999), for gastric cancer (Hemminki and Jiang 2002), and for

colorectal adenocarcinoma (Hemminki and Chen 2004), cases of early-onset CRC are admixed with late-onset CRC in calculating overall and age-specific familial risk. However, as there are known genetic conditions conferring early risk for CRC, such as familial adenomatous polyposis (FAP) with a mean age of onset of 39, and hereditary non-polyposis colorectal cancer (HNPCC, Lynch syndrome) with a mean age of onset of 44, the definition of late-onset CRC familial risk requires that parent-child shared risks be calculated for late-onset CRC alone. In the 1998 version of the Swedish Family Cancer Database, some 40% of offspring CRC cases were under the age of 50, creating potential bias in estimating familial risk for late-onset CRC. The process of sorting out the age-specific incidence in cases where CRC is detected in parents and children of age 50 or greater is in progress. When this is completed, a reasonable estimate may be obtained for values of *FRR(t)* in male offspring of male parents with late-onset CRC, or female offspring of female parents with late-onset CRC. *i.e.* diagnosed at age 50 or older.

As there are no calculated values available at this time point we have made an educated guess of 1.5-3.0 for *GRR* in colorectal cancer and calculate the values of *GRR* over a wide set (1.5, 2.0, 2.5, 3.0), in order to encompass a broad set of possibilities for various common diseases.

Considered Modes of Inherited Risk

The possibilities of genetic risk considered are heterozygosity and nullizygosity (the latter usually referred to as recessive inheritance in the literature), for one or more genes acting independently, each in a sufficient but not necessary mode (multigenically, *i.e.* in series), or coordinately, in a necessary but not sufficient mode (polygenically, *i.e.* in parallel), to confer risk for the disease under study.

Formal expressions for genetic relative risk, for autosomal and sex-linked monogenic disorders, were previously derived for any kind of relatives, using stochastic matrices (Li and Sacks 1954). Another method to obtain such expressions is using path analysis. Yet

another approach is via mating tables, where parental and offspring combinations are enumerated, probabilities are calculated, and values are summated; this approach allows a more tractable extension to multigenic and polygenic modes of inheritance.

First, the basic algebra for monogenic disorders is presented for risk conferred by heterozygous or nullizygous conditions. Derivations, ranges of values, figures, and tables are given, emphasizing the relationship between and q and G_0, G_0 and GRR, and q and GRR, for each possible mode of inheritance.

Second, in a similar but more elaborate manner, the basic algebra for multigenic disorders is presented. Derivations are lengthier, closed formulæ are given, and more refined formulæ are calculated for levels of muligenicity from 2 to 5 genes.

Third, the basic algebra for polygenic disorders is presented. General formulæ are obtained, that can yield values for any number of genes.

Given these several but distinct formal relationships, we consider how they may be used to delimit the modes of inheritance of risk for a particular form of late-onset cancer, using CRC as an example.

FAMILIAL RELATIVE RISK

As derived above, the genetic relative risk, *GRR*, is observed as the ratios of incidence at ages 50-54, 55-59,…,70-74, among the fathers of sons, or mothers of daughters with a particular common disease at age $t \geq 50$, to the incidence of that disease in those age intervals for all fathers and mothers respectively. Environmental risks appear to be equal in both generations, and stochastic risks are assumed to be the same among persons of the same gender in both generations compared. That is:

$$GRR \approx FRR(t = 50-54,...,70-74) \approx \frac{INC_1(t = 50-54,...,70-74)}{INC_0(t = 50-54,...,70-74)}$$

Formally, *GRR* is also the ratio of genetic risk, G_1, in fathers of sons, or mothers of daughters with incidence at $t \geq 50$, to the general genetic risk, G_0. That is:

$$GRR = G_1 / G_0$$

These two relationships for *GRR* permit us to calculate expected values of *GRR* for any estimated value of G_0. Recall that:

$$G_0 = F_0 / E \geq F_0 \geq mLR$$

Thus, for any observed value of *mLR*, which is the area under the curve *INC(t)* versus *t* in the general population of male or female parents, we may posit that all values of G_0 equal to or greater than *mLR* are possible, and that for an observed value of *GRR* the set of all values of G_1 are thus approximated.

Formal forms of Mendelian risk may thus be evaluated to discover if any of the possible values of *GRR* are consistent with the specifically posited form of genetic risk.

Risk may vary with regard to the number of genes conferring risk for a particular disease independently or in concert. Risk conferred by a mutation or mutations in that gene distributed over the general population is termed "monogenic risk".

Risk conferred by any of *m* genes each independently carrying risk-conferring mutations is termed "multigenic risk". Risk conferred by any of *n* genes carrying risk-conferring mutations but acting together to create risk is termed "polygenic risk".

The formal Mendelian definitions of the "zygosity" of the genetic risk apply. If an individual to be at genetic risk must carry gene-inactivating, risk-conferring mutations in both copies of an autosomal gene then the risk is termed "nullizygous". If two active copies are required for risk, it is termed "homozygous risk". If one active and one inactivated copy is required to define risk, it is termed "heterozygous risk".

There are, of course, a very large number of possible combinations of genetic risk, such as polygenic risks, each element of which consists of several multigenic risks.

To present a careful analysis of these possibilities, we have restricted our analyses to monogenic and simplified sets of multigenic and polygenic risk. For each condition, we have derived the explicit equation for G_0 for any values of q (fraction of inactivated, risk-conferring alleles) for any number of genes considered, but have also calculated explicit values for q, under the simplifying assumption that q is the same for all genes.

These points having been explained, we now present the formulae we have derived for the various forms of genetic risk relating G_0 to q, GRR to q, and GRR to G_0. These formulae are then in turn applied to create tables by which the possible values of $G_0 \geq$ mLR and GRR are used to discover if they exclude one or more of the formal genetic possibilities considered. Among those possibilities that are not excluded, the resulting estimates of q are then considered in terms of the expectation that genes conferring risk for common diseases have a mean value of q of about 0.03 and a 99% u.c.l. of about 0.22.

I. MONOGENIC RISK

1. Monogenic Heterozygous Risk: $A^{+/-}$

[Figures 8.1-8.4 and Table 1]

General genetic risk:

$G_0 = 2pq = 2q(1-q)$

$2q^2 - 2q + G_0 = 0$

$\boxed{q = (1 \pm \sqrt{1-2G_0})/2}$

Transmission of risk:

Mating: $A^{+/-} \times$ ($A^{+/+}$, $A^{+/-}$, $A^{-/-}$).

G_1 = P(any child is $A^{+/-}$) = 0.5

G_1 is the same for either one or two afflicted parents

$\boxed{GRR = \frac{0.5}{G_0} = \frac{0.5}{2pq}}$ with either 1 or 2 affected parents

Limitations on value of G_0:

$G_0 \leq 0.5 = 1/2$ as the maximum fraction of heterozygotes is 1/2 for all values of q

2. Monogenic Nullizygous Risk: $A^{-/-}$

General genetic risk:

$G_0 = q^2$

$\boxed{q = \sqrt{G_0}}$

Transmission of risk:

Mating: $A^{-/-} \times (A^{+/+}, A^{+/-}, A^{-/-})$

G_1 differs for one vs. two afflicted parents

i. One affected parent [Figures 9.1-9.4 and Table 2]

Mating: $A^{-/-} \times (A^{+/+},\ A^{+/-},\ A^{-/-})$

$G_1 = \text{P(any child is } A^{-/-}) = 0 \cdot p^2 + \frac{1}{2} \cdot 2pq + 1 \cdot q^2 = pq + q^2 = q(p+q) = q$

$$GRR = \frac{G_1}{G_0} = \frac{q}{q^2} = \frac{1}{q} = \frac{1}{\sqrt{G_0}}$$

$$\boxed{GRR = \frac{1}{\sqrt{G_0}} = \frac{1}{q}}$$

ii. Two affected parents

Mating: $A^{-/-} \times A^{-/-}$

$G_1 = \text{P(any child is } A^{-/-}) = 1$

$$\boxed{GRR = \frac{1}{G_0} = \frac{1}{q^2}}$$

No limitations on value of G_0 for monogenic nullizygous risk:

G_0 may have any value from 0 to 1

II. MULTIGENIC RISK (OR)

The inclusion-exclusion formula of probability has been used to define the expected values of GRR for the specific examples. Risk may be conferred by carrying mutations in one or more of $n > 1$ genes but mutation in any one of n genes is sufficient to create risk.

1. Multigenic Heterozygous Risk: $A_1^{+/-}$ OR $A_2^{+/-}$ OR ... OR $A_m^{+/-}$

i. m = 2 [Figures 10.1-10.4 and Table 3]

General genetic risk:

$G_0 = 2p_1q_1 + 2p_2q_2 - (2p_1q_1)(2p_2q_2)$

Transmission of risk:

$G_1 = [0.5 + 2p_2q_2 - (0.5)(2p_2q_2)][2p_1q_1/(2p_1q_1+2p_2q_2)] +$

$[0.5 + 2p_1q_1 - (0.5)(2p_1q_1)][2p_2q_2/(2p_1q_1+2p_2q_2)]$

$GRR = \{[0.5 + 2p_2q_2 - (0.5)(2p_2q_2)][2p_1q_1/(2p_1q_1+2p_2q_2)] +$

$[0.5 + 2p_1q_1 - (0.5)(2p_1q_1)][2p_2q_2/(2p_1q_1+2p_2q_2)]\} /$

$[2p_1q_1 + 2p_2q_2 - (2p_1q_1)(2p_2q_2)]$

Simplification for calculation: $q_i \approx q$

Then,

$$GRR = \frac{G_1}{G_0} \approx \frac{0.5 + pq}{2(2pq) - (2pq)^2} = \frac{0.5 + pq}{1 - (1 - 2pq)^2}$$

$$q = [1 \pm \sqrt{1 - 2(1 - \sqrt{1 - G_0})}]/2$$

Limitations on value of G_0:

As the maximum fraction of heterozygotes is 1/2 for each of the two genes conferring risk, a limit exists for the maximum genetic risk that may be created by two genes independently conferring heterozygous risk:

$G_0 \leq 0.75 = 3/4$

ii. m = 3 [Figures 11.1-11.4 and Table 4]

General genetic risk:

$$G_0 = \sum_{i=1}^{3}(2p_iq_i) - \sum_{i<j}^{3}(2p_iq_i)(2p_jq_j) + (2p_1q_1)(2p_2q_2)(2p_3q_3) = 2p_1q_1 + 2p_2q_2 + 2p_3q_3$$
$$- (2p_1q_1)(2p_2q_2) - (2p_1q_1)(2p_3q_3) - (2p_2q_2)(2p_3q_3) + (2p_1q_1)(2p_2q_2)(2p_3q_3)$$

Transmission of risk:

$G_1 = [2p_1q_1/(2p_1q_1+2p_2q_2+2p_3q_3)]$ x

$[0.5 + 2p_2q_2 + 2p_3q_3 - (0.5)(2p_2q_2 + 2p_3q_3) - (2p_2q_2)(2p_3q_3) + (0.5)(2p_2q_2)(2p_3q_3)] +$

$[2p_2q_2/(2p_1q_1+2p_2q_2+2p_3q_3)]$ x

$$[0.5+2p_1q_1+2p_3q_3-(0.5)(2p_1q_1+2p_3q_3)-(2p_1q_2)(2p_1q_3)+(0.5)(2p_1q_1)(2p_3q_3)]+$$

$[2p_3q_3/(2p_1q_1+2p_2q_2+2p_3q_3)]$ x

$$[0.5+2p_1q_1+2p_2q_2-(0.5)(2p_1q_1+2p_2q_2)-(2p_1q_1)(2p_2q_2)+(0.5)(2p_1q_1)(2p_2q_2)]$$

Simplification for calculation: $q_i \approx q$

Then,

$$GRR=\frac{G_1}{G_0}\approx\frac{0.5+2pq-2p^2q^2}{3(2pq)-3(2pq)^2+(2pq)^3}=\frac{1-0.5(1-2pq)^2}{1-(1-2pq)^3}$$

$$q=[1\pm\sqrt{1-2(1-\sqrt[3]{1-G_0})}]/2$$

Limitations on value of G_0:

As the maximum fraction of heterozygotes is 1/2 for each of the three genes conferring risk, a limit exists for the maximum genetic risk that may be created by three genes independently conferring heterozygous risk:

$G_0 \leq 0.875 = 7/8$

iii. m = 4 [Figures 12.1-12.4 and Table 5]

Similar terms may be defined for the exact expected values of G_0 and G_1 for all values of m but absent values of q for any of the genes, we here summarize the results using $q_i \approx q$ for m = 4 and m = 5 that are used in my calculations:

$$GRR=\frac{G_1}{G_0}\approx\frac{1-0.5(1-2pq)^3}{1-(1-2pq)^4}$$

$$q=[1\pm\sqrt{1-2(1-\sqrt[4]{1-G_0})}]/2$$

Limitations on value of G_0:

As the maximum fraction of heterozygotes is 1/2 for each of the four genes conferring risk, a limit exists for the maximum genetic risk that may be created by four genes independently conferring heterozygous risk:

$G_0 \leq 0.9375 = 15/16$

iv. m = 5 [Figures 13.1-13.4 and Table 6]

Assuming $q_i \approx q$:

$$GRR = \frac{G_1}{G_0} \approx \frac{1-0.5(1-2pq)^4}{1-(1-2pq)^5}$$

$$q = [1 \pm \sqrt{1-2(1-\sqrt[5]{1-G_0})}]/2$$

Limitations on value of G_0:

As the maximum fraction of heterozygotes is 1/2 for each of the five genes conferring risk, a limit exists for the maximum genetic risk that may be created by five genes independently conferring heterozygous risk:

$G_0 \leq 0.96875 = 31/32$

v. m>5

General calculations for multigenic heterozygous risk

Set $0.5 = 2p_{1*}q_{1*}$ for indexing in the numerator ($*$) - no biological interpretation

$$GRR = \frac{0.5 + \sum_{i>1}(2p_iq_i) - \sum_{i*<j*}^{\binom{m}{2}}(2p_iq_i)(2p_jq_j) + \ldots + (-1)^m(0.5)(2p_2q_2)\ldots(2p_mq_m)}{\sum_{i=1}^{m}(2p_iq_i) - \sum_{i<j}^{\binom{m}{2}}(2p_iq_i)(2p_jq_j) + \ldots + (-1)^{m-1}(2p_1q_1)(2p_2q_2)\ldots(2p_mq_m)}$$

$\boxed{q_i \approx q}$ for the following calculations for GRR = G_1/G_0

$\boxed{G_1}$: For the numerator, there are $\binom{m}{k}=\binom{m-1}{k-1}+\binom{m-1}{k}$ combinatorial terms

$$G_1=[0.5+(m-1)(2pq)]-\left[(m-1)(0.5)(2pq)+\binom{m-1}{2}(2pq)^2\right]$$
$$+\left[\binom{m-1}{2}(0.5)(2pq)^2+\binom{m-1}{3}(2pq)^3\right]-\left[\binom{m-1}{3}(0.5)(2pq)^3+\binom{m-1}{4}(2pq)^4\right]$$
$$+\ldots+(-1)^{k-1}\left[\binom{m-1}{k-1}(0.5)(2pq)^{k-1}+\binom{m-1}{k}(2pq)^k\right]+\ldots+(-1)^m(0.5)(2pq)^{m-1}$$

$$G_1=1-0.5+(m-1)(0.5)(2pq)-\binom{m-1}{2}(0.5)(2pq)^2$$
$$+\binom{m-1}{3}(0.5)(2pq)^3+\ldots+(-1)^{k+1}\binom{m-1}{k}(0.5)(2pq)^k+\ldots+(-1)^m(0.5)(2pq)^{m-1}$$

$$\Rightarrow\quad G_1=1-0.5\left[\sum_{i=0}^{m-1}\binom{m-1}{i}(-2pq)^i\right]=1-0.5(1-2pq)^{m-1}$$

$\boxed{G_0}$: For the denominator,

$$G_0=m(2pq)-\binom{m}{2}(2pq)^2+\binom{m}{3}(2pq)^3+\ldots+(-1)^{k-1}\binom{m}{k}(2pq)^k+\ldots+(-1)^{m-1}(2pq)^m$$

$$G_0=-\sum_{j=1}^{m}\binom{m}{j}(-2pq)^j=1-\sum_{j=0}^{m}\binom{m}{j}(-2pq)^j=1-(1-2pq)^m$$

The ratio, given the assumption $q_i\approx q$, reduces to:

$$\boxed{GRR=\frac{G_1}{G_0}\approx\frac{1-0.5(1-2pq)^{m-1}}{1-(1-2pq)^m}\approx\frac{1-0.5(1-G_0)^{1-1/m}}{G_0}}$$

Comments on multigenic heterozygous risk

GRR is a decreasing function of G_0, for a given m.

$$GRR = \frac{1-0.5(1-2pq)^{m-1}}{1-(1-2pq)(1-2pq)^{m-1}}$$

GRR is a symmetric bathtub function of q, for a given m.

It is symmetric about $q = 0.5$, where it attains the minimum value 1.

$0 \le q \le 1 \quad \Rightarrow \quad 2pq \le 0.5 \quad \Rightarrow \quad 1-2pq \ge 0.5 \quad \Rightarrow \quad GRR \ge 1$

$GRR = 1 \quad \Leftrightarrow \quad q = 0.5$

Range of values

$q \downarrow 0$, $q \uparrow 1$: $GRR \to \infty$

$0 < q \le 0.5 \qquad GRR \downarrow$

$0.5 \le q < 1 \qquad GRR \uparrow$

Expression for q in terms of G_0

$G_0 = 1-(1-2pq)^m$

$(1-2pq)^m = 1-G_0$

$1-2pq = \sqrt[m]{1-G_0}$

$2pq = 1-\sqrt[m]{1-G_0}$

$2q(1-q) = 1-\sqrt[m]{1-G_0}$

$2q^2 - 2q + (1-\sqrt[m]{1-G_0}) = 0$

$$\boxed{q = [1 \pm \sqrt{1-2(1-\sqrt[m]{1-G_0})}]/2}$$

Limitations on value of G_0:

As the maximum fraction of heterozygotes is 1/2 for each of the *m* genes conferring risk, a limit exists for the maximum genetic risk that may be created by *m* genes independently conferring heterozygous risk:

$$\boxed{G_0 \leq 1-0.5^m}$$

2. Multigenic Nullizygous Risk: $A_1^{-/-}$ OR $A_2^{-/-}$ OR ... OR $A_m^{-/-}$

Only calculations for one affected parent have been made. A briefer text format is now used to avoid even more tedious repetitions.

i. m = 2 [Figures 14.1-14.4 and Table 7]

$G_0 = q_1^2 + q_2^2 - q_1^2 q_2^2$

$G_1 = q_1 + q_2^2 - q_1 q_2^2$

$$GRR = \frac{q_1 + q_2^2 - q_1 q_2^2}{q_1^2 + q_2^2 - q_1^2 q_2^2}$$

For the simplifying condition, $q_i \approx q$:

$$GRR = \frac{G_1}{G_0} \approx \frac{q + q^2 - q^3}{2q^2 - q^4} = \frac{1-(1-q)(1-q^2)}{1-(1-q^2)^2}$$

and

$$q = \sqrt{1-\sqrt{1-G_0}}$$

ii. m = 3 [Figures 15.1-15.4 and Table 8]

$G_0 = q_1^2 + q_2^2 + q_3^2 - q_1^2 q_2^2 - q_1^2 q_3^2 - q_2^2 q_3^2 + q_1^2 q_2^2 q_3^2$

$G_1 = q_1 + q_2^2 + q_3^2 - q_1 q_2^2 - q_1 q_3^2 - q_2^2 q_3^2 + q_1 q_2^2 q_3^2$

$$GRR = \frac{q_1 + q_2^2 + q_3^2 - q_1 q_2^2 - q_1 q_3^2 - q_2^2 q_3^2 + q_1 q_2^2 q_3^2}{q_1^2 + q_2^2 + q_3^2 - q_1^2 q_2^2 - q_1^2 q_3^2 - q_2^2 q_3^2 + q_1^2 q_2^2 q_3^2}$$

For the simplifying condition, $q_i \approx q$:

$$GRR = \frac{G_1}{G_0} \approx \frac{q + 2q^2 - 2q^3 - q^4 + q^5}{3q^2 - 3q^4 + q^6} = \frac{1-(1-q)(1-q^2)^2}{1-(1-q^2)^3}$$

and

$$q = \sqrt{1-\sqrt[3]{1-G_0}}\ .$$

iii. m = 4 [Figures 16.1-16.4 and Table 9]

For the simplifying condition, $q_i \approx q$:

$$GRR = \frac{G_1}{G_0} \approx \frac{1-(1-q)(1-q^2)^3}{1-(1-q^2)^4}$$

and

$$q = \sqrt{1-\sqrt[4]{1-G_0}}\ .$$

iv. m = 5 [Figures 17.1-17.4 and Table 10]

For the simplifying condition, $q_i \approx q$:

$$GRR = \frac{G_1}{G_0} \approx \frac{1-(1-q)(1-q^2)^4}{1-(1-q^2)^5}$$

and

$$q = \sqrt{1-\sqrt[5]{1-G_0}}\ .$$

v. m>5

General calculations for multigenic nullizygous risk

Setting $q_1 = q_{1*}^2$ just for indexing in the numerator ($*$) - no biological interpretation

$$GRR = \frac{q_1 + \sum_{i>1} q_i^2 - \sum_{i*<j*}^{\binom{m}{2}} (q_i^2 q_j^2) + \sum_{i*<j*<k*}^{\binom{m}{3}} (q_i^2 q_j^2 q_k^2) + \ldots + (-1)^{m-1} q_1 q_2^2 q_3^2 \ldots q_m^2}{\sum_{i=1}^{m} q_i^2 - \sum_{i<j}^{\binom{m}{2}} (q_i^2 q_j^2) + \sum_{i<j<k}^{\binom{m}{3}} (q_i^2 q_j^2 q_k^2) + \ldots + (-1)^{m-1} q_1^2 q_2^2 q_3^2 \ldots q_m^2}$$

$\boxed{q_i \approx q}$ for the following calculations

$\boxed{G_1}$: For the numerator,

$$G_1 = \left[q + (m-1)q^2\right] - \left[(m-1)q^3 + \binom{m-1}{2}q^4\right] + \left[\binom{m-1}{2}q^5 + \binom{m-1}{3}q^6\right]$$
$$+ \ldots + (-1)^k\left[\binom{m-1}{k}q^{2k+1} + \binom{m-1}{k+1}q^{2k+2}\right] + \ldots + (-1)^{m-1}q^{2m-1}$$

$$G_1 = q + (m-1)(q^2 - q^3) - \binom{m-1}{2}(q^4 - q^5) + \binom{m-1}{3}(q^6 - q^7)$$
$$+ \ldots + (-1)^{k+1}\binom{m-1}{k}(q^{2k} - q^{2k+1}) + \ldots + (-1)^{m-1}q^{2m-1}$$

$$G_1 = 1 - (1-q) + (m-1)q^2(1-q) - \binom{m-1}{2}q^4(1-q)$$
$$+ \ldots + (-1)^{k+1}\binom{m-1}{k}q^{2k}(1-q) + \ldots + (-1)^{m-1}q^{2m-1}$$

$$\Rightarrow \quad G_1 = 1 - (1-q)\sum_{i=0}^{m-1}\binom{m-1}{i}(-q^2)^i = 1 - (1-q)(1-q^2)^{m-1}$$

$\boxed{G_0}$: For the denominator,

$$G_0 = mq^2 - \binom{m}{2}q^4 + \binom{m}{3}q^6 + \ldots + (-1)^{m-1}q^{2m} = -\sum_{j=1}^{m}\binom{m}{j}(-q^2)^j = 1 - (1-q^2)^m$$

$$\boxed{GRR = \frac{G_1}{G_0} \approx \frac{1 - (1-q)(1-q^2)^{m-1}}{1 - (1-q^2)^m} \approx \frac{1 - (1 - \sqrt{1 - \sqrt[m]{1-G_0}})\sqrt[m]{(1-G_0)^{m-1}}}{G_0}}$$

Comments on multigenic nullizygous risk

GRR is a decreasing function of G_0, for a given m

$$GRR = \frac{1-(1-q)(1-q^2)^{m-1}}{1-(1-q^2)(1-q^2)^{m-1}}$$

GRR is a decreasing function of q, for a given m.

$0 < q < 1 \quad \Rightarrow \quad q > q^2 \quad \Rightarrow \quad 1-q < 1-q^2 \quad \Rightarrow \quad GRR > 1$

Range of values for GRR

$q \downarrow 0: \; GRR \to \infty$

$q \uparrow 1: \; GRR \to 1$

Expression for q in terms of G_0

$G_0 = 1-(1-q^2)^m$

$(1-q^2)^m = 1-G_0$

$1-q^2 = \sqrt[m]{1-G_0}$

$q^2 = 1-\sqrt[m]{1-G_0}$

$$\boxed{q = \sqrt{1-\sqrt[m]{1-G_0}}}$$

Given G_0,

$m \uparrow \quad q \downarrow \quad GRR \downarrow$

III. POLYGENIC RISK (AND)

1. Polygenic Heterozygous Risk: $A_1^{+/-}$ AND $A_2^{+/-}$ AND ... AND $A_n^{+/-}$

$$G_0 = \prod_1^n (2p_i q_i)$$

$G_1 = 0.5^n$

$$GRR = \frac{0.5^n}{G_0} = \frac{0.5^n}{(2p_1q_1)(2p_2q_2)...(2p_nq_n)} = \frac{0.5^n}{\prod_1^n (2p_i q_i)}$$

For the simplifying assumption $q_i \approx q$,

$$\boxed{GRR = \frac{0.5^n}{G_0} \approx \left(\frac{0.5}{2pq}\right)^n}$$

$$G_0^{1/n} = 2pq = 2q(1-q)$$

$$2q^2 - 2q + G_0^{1/n} = 0$$

$$\boxed{q = (1 \pm \sqrt{1 - 2G_0^{1/n}})/2}$$

Limitations on value of G_0:

$$\boxed{G_0 \le 0.5^n}$$

For the case n = 2,

[Figures 18.1-18.4 and Table 11]

$$GRR = \frac{(0.5)(0.5)}{(2p_1q_1)(2p_2q_2)} = \frac{0.5^2}{(2pq)^2} = \frac{0.25}{G_0}$$

where $G_0(n=2) \le 0.25$.

2. Polygenic Nullizygous Risk: $A_1^{-/-}$ AND $A_2^{-/-}$ OR ... AND $A_n^{-/-}$

i. One affected parent [Figures 19.1-19.4 and Tables 12-13]

Generally,

$$G_0 = q_1^2 q_2^2 q_3^2 ... q_n^2$$

$$G_1 = q_1 \cdot (q_2 q_3 ... q_n)$$

$$GRR = \frac{G_1}{G_0} = \frac{q_1 \cdot (q_2 q_3 .. q_n)}{q_1^2 q_2^2 q_3^2 .. q_n^2} = \frac{1}{q_1 q_2 .. q_n} = \frac{1}{\prod_1^n q_i} = \frac{1}{\sqrt{G_0}}$$

If $q_i \approx q$, then

$$\boxed{GRR = \frac{1}{\sqrt{G_0}} \approx \frac{1}{q^n}}$$

ii. Two affected parents

Generally,

$G_0 = 1$

$$G_1 = \prod_1^n q_i^2$$

$$GRR = \frac{1}{G_0} = \frac{1}{q_1^2 q_2^2 q_3^2 .. q_n^2} = \frac{1}{\prod_1^n q_i^2}$$

If $q_i \approx q$, then

$$\boxed{GRR = \frac{1}{G_0} \approx \frac{1}{q^{2n}}}$$

With either one or two affected parents,

$$G_0 = q^{2n}$$

$$\boxed{q = G_0^{1/2n}}$$

GRAPHIC AND TABULAR PRESENTATION OF CALCULATED RELATIONSHIPS

The values obtained for each mode of genetic transmission addressed above are presented in this section. The observed values corresponding to *FRR (t = 50-54,...,70-74)*, by which *GRR* is approximated, are considered at the specific values of 1.5, 2.0, 2.5 and 3.0 which encompasses the estimates made to date by the Hemminki group for various common late-onset forms of cancer including colorectal cancer, CRC. When the data of the Swedish family cancer register are calculated to yield gender specific values of *FRR (t = 50-54,...,70-74)* for specific cancers these tables may be used with the estimate of $G_0 \geq$ *mLR* to discover for each cancer type if any mode of Mendelian inheritance is excluded.

As an example, we consider the case of colorectal cancer in Swedish male parents from which we derived (Figure 6) a value of *mLR* of 0.26 and from several reports of the Hemminki group of an estimate of *GRR* for colorectal cancer between 1.5 and 3.0. This estimate of *GRR* for colorectal cancer is under refinement now by the collaborating groups but our contribution is to set up the logical structure that would permit arithmetic exclusion of particular Mendelian modes for any disease for which estimates of *mLR* and *GRR* were made.

Please note that for the following figures, the variable G actually indicates G_0.

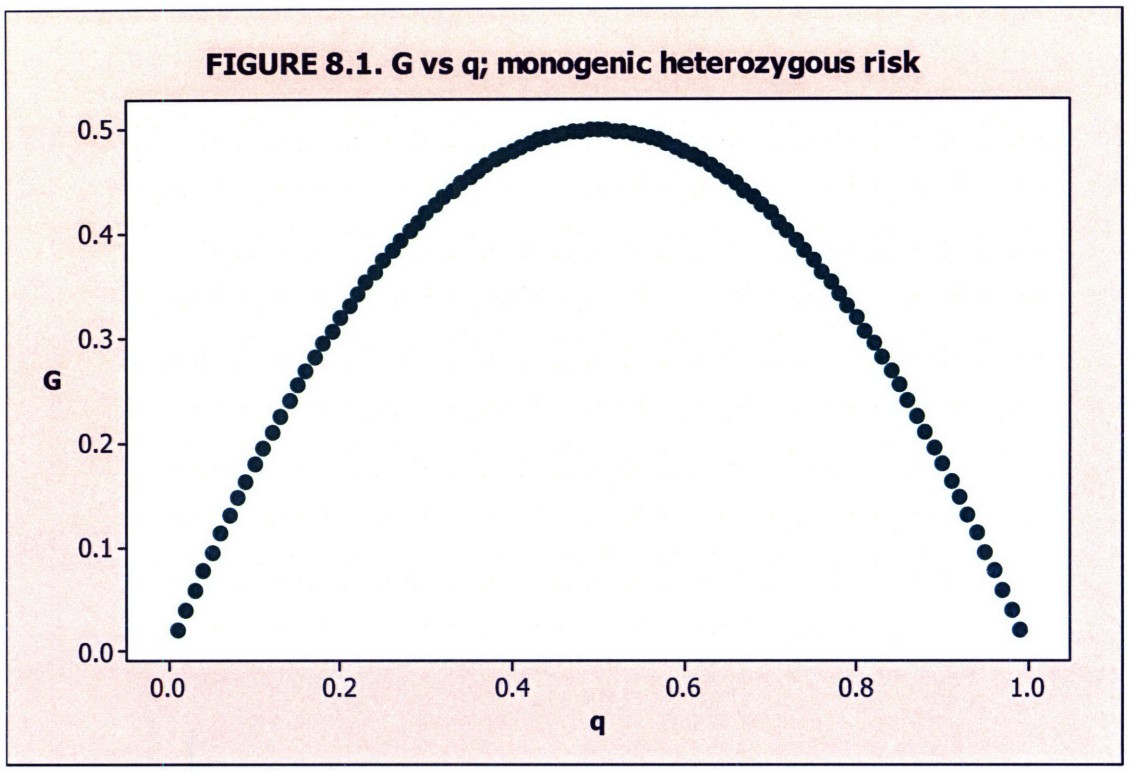


FIGURE 8.1. G vs q; monogenic heterozygous risk
G
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.2
0.4
0.6
0.8
1.0
q

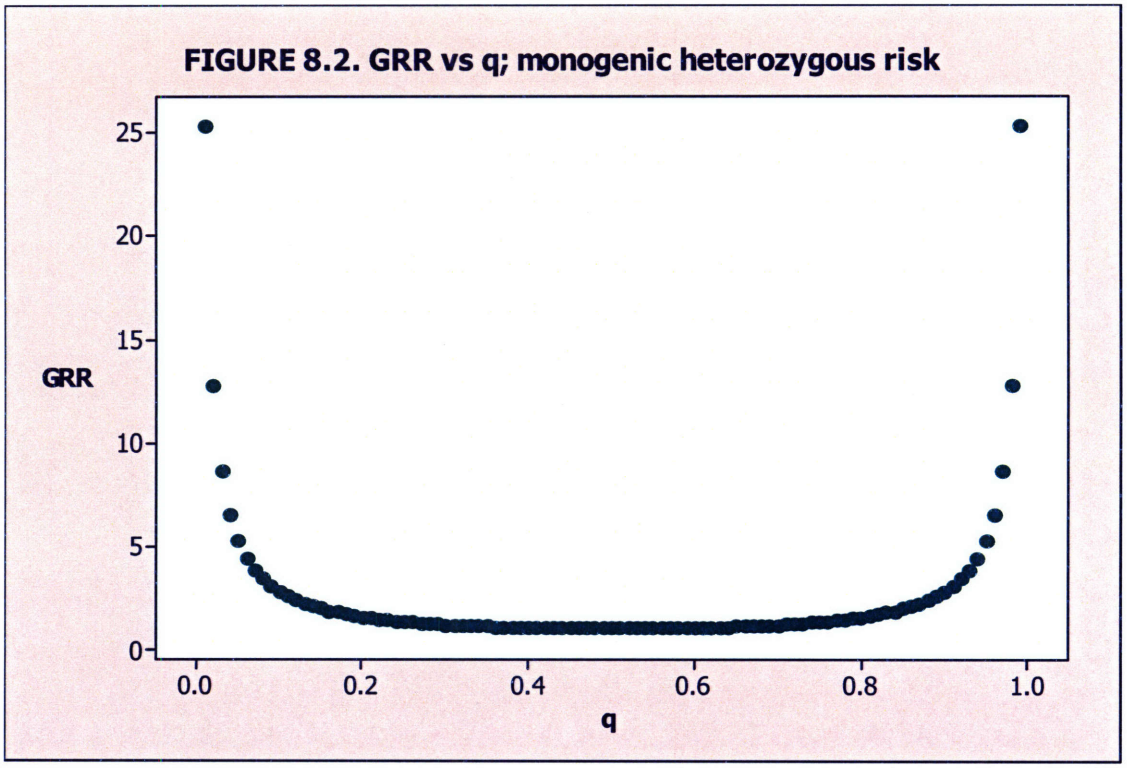


FIGURE 8.2. GRR vs q; monogenic heterozygous risk
GRR
25
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
q

FIGURE 8.3. GRR vs G; monogenic heterozygous risk

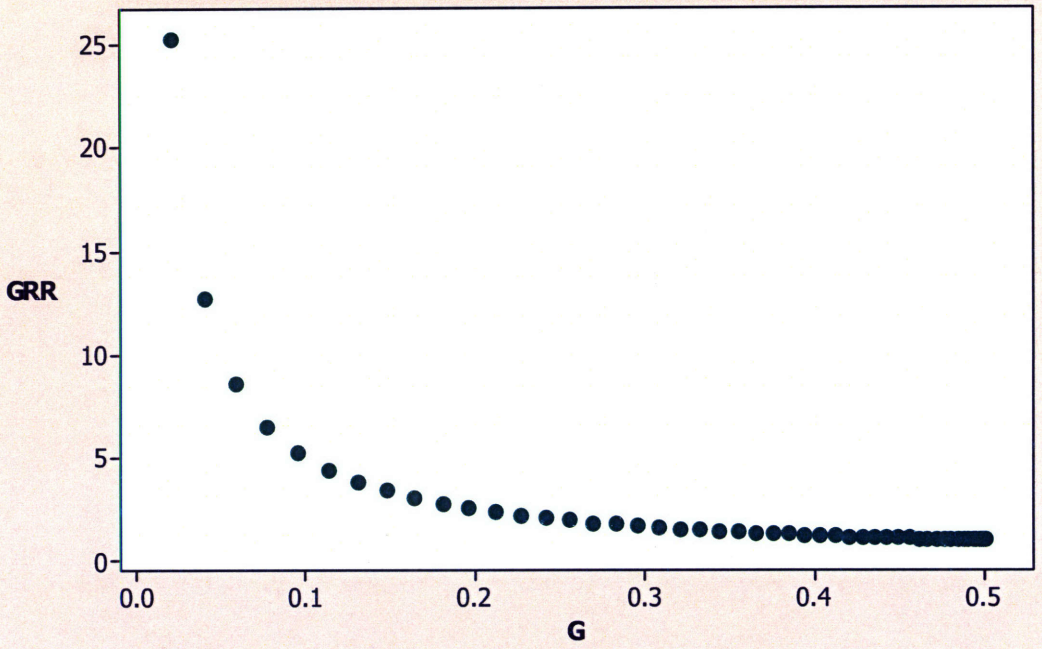
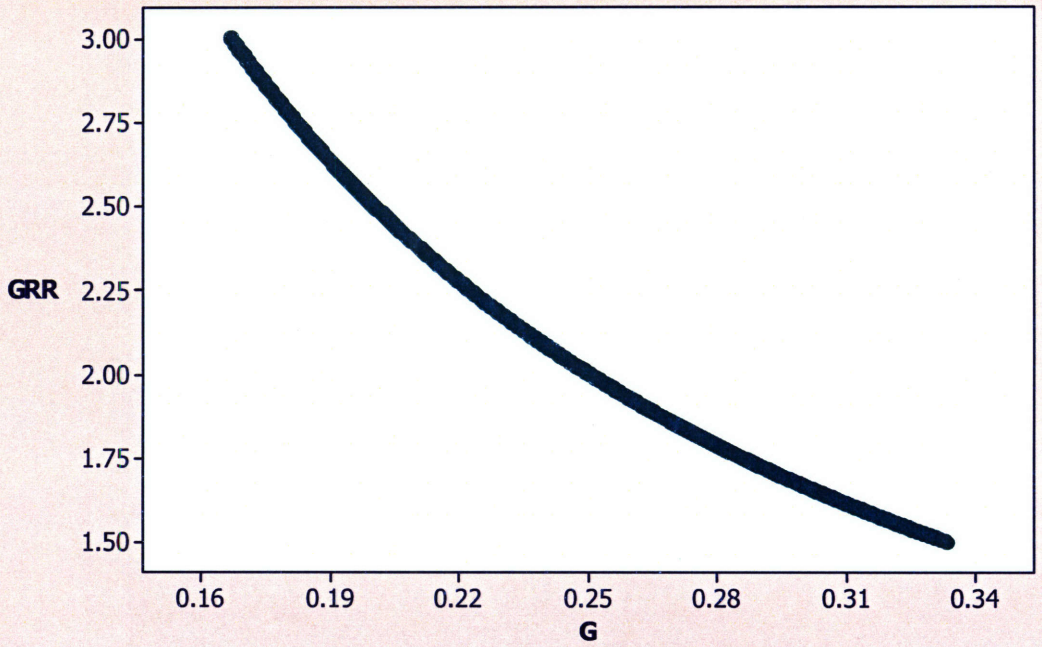


FIGURE 8.4. GRR vs G; monogenic heterozygous risk



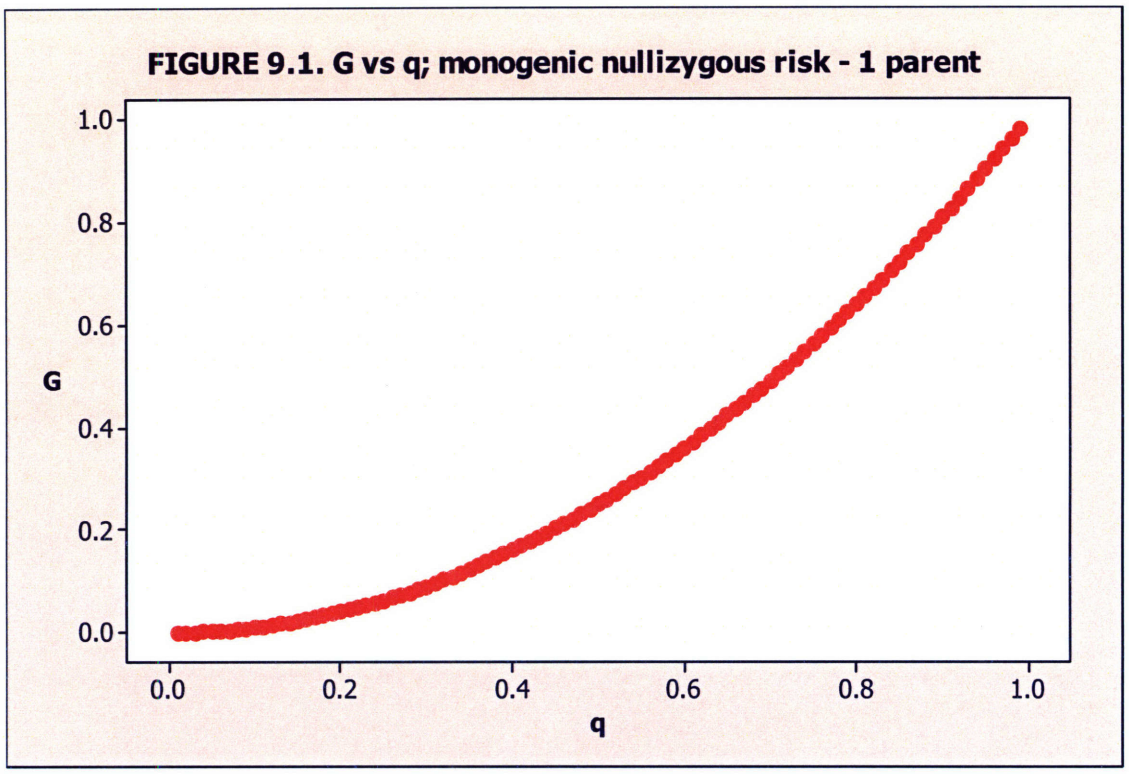


FIGURE 9.1. G vs q; monogenic nullizygous risk - 1 parent
G
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
q

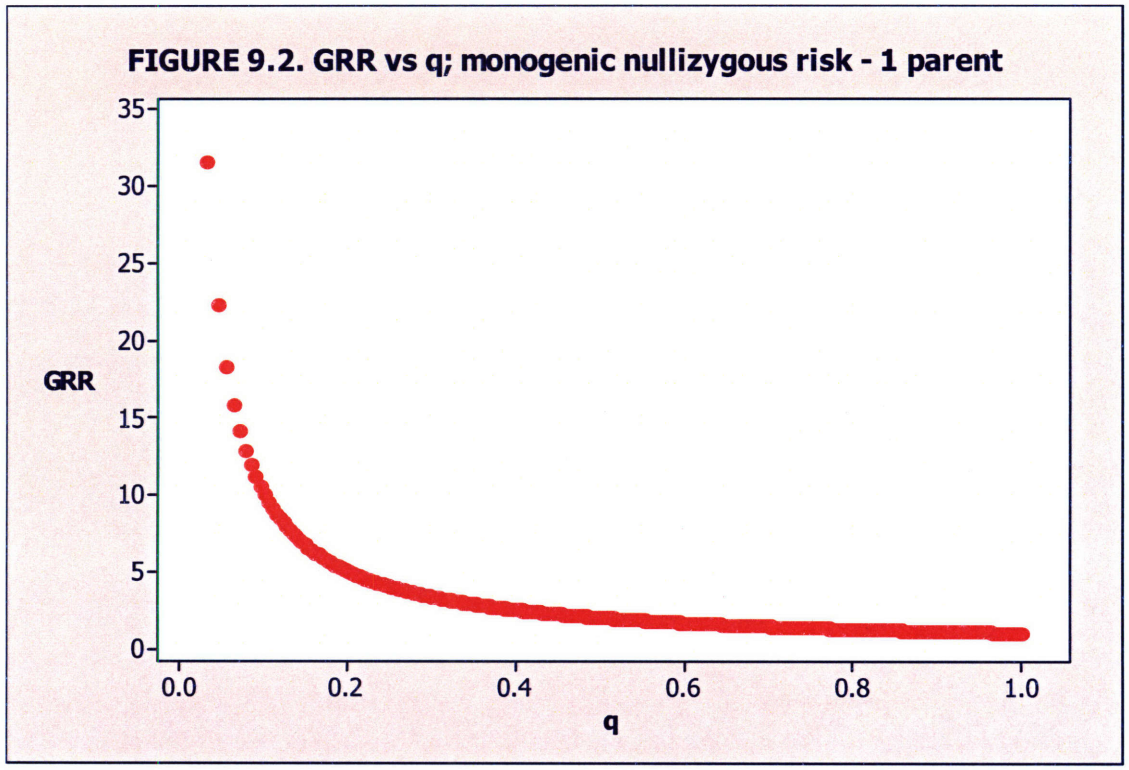


FIGURE 9.2. GRR vs q; monogenic nullizygous risk - 1 parent
GRR
35
30
25
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
q

FIGURE 9.3. GRR vs G; monogenic nullizygous risk - 1 parent

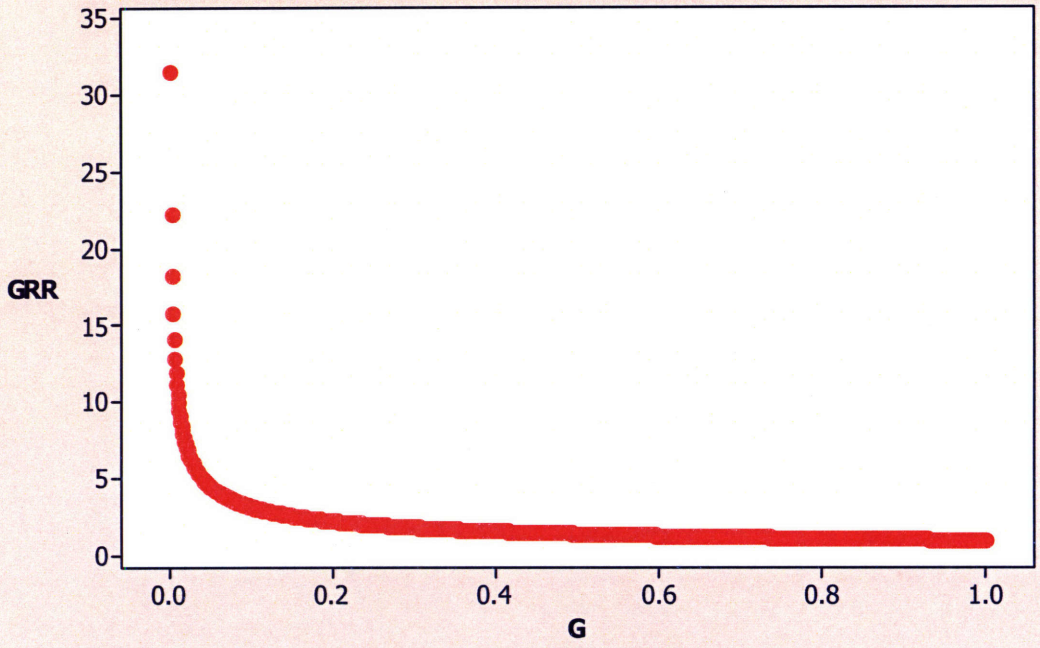


FIGURE 9.4. GRR vs G; monogenic nullizygous risk - 1 parent

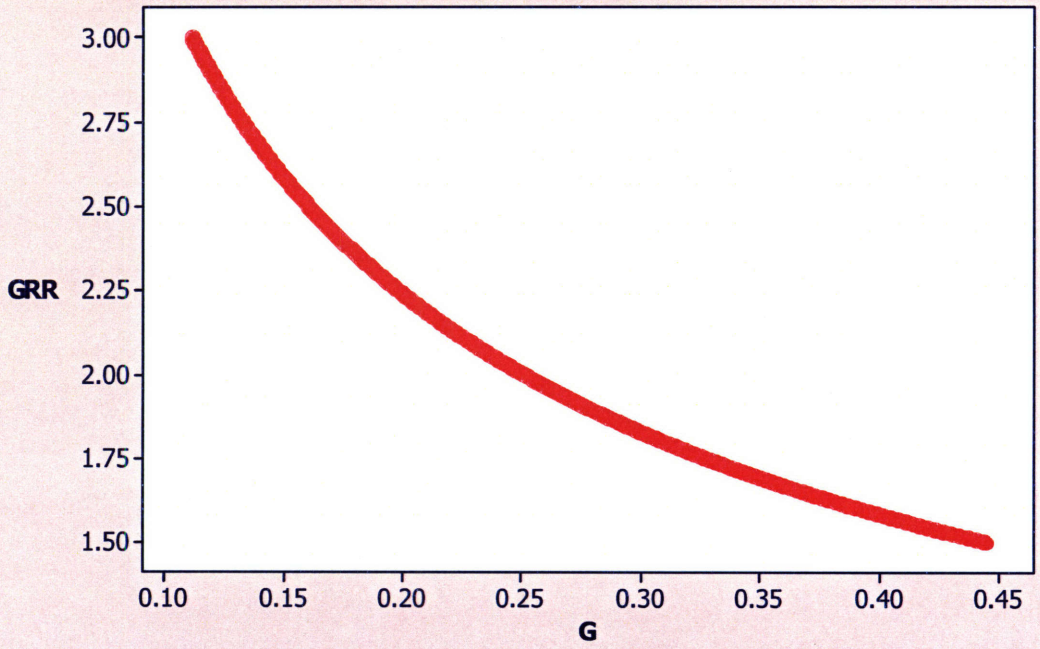


FIGURE 10.1. G vs q; multigenic heterozygous risk, m=2

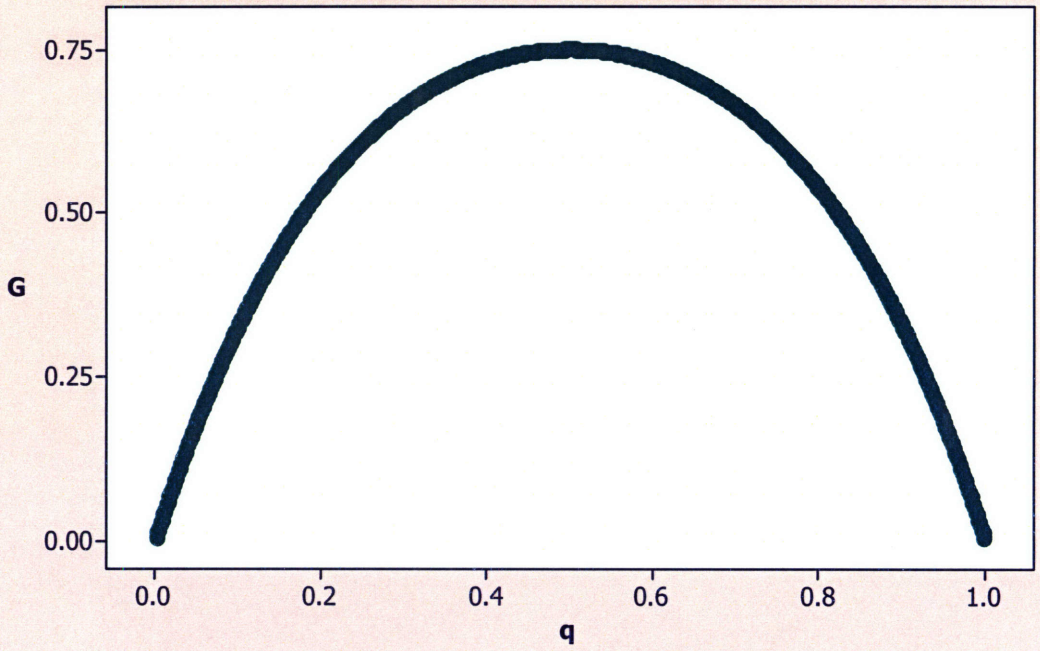


FIGURE 10.2. GRR vs q; multigenic heterozygous risk, m=2

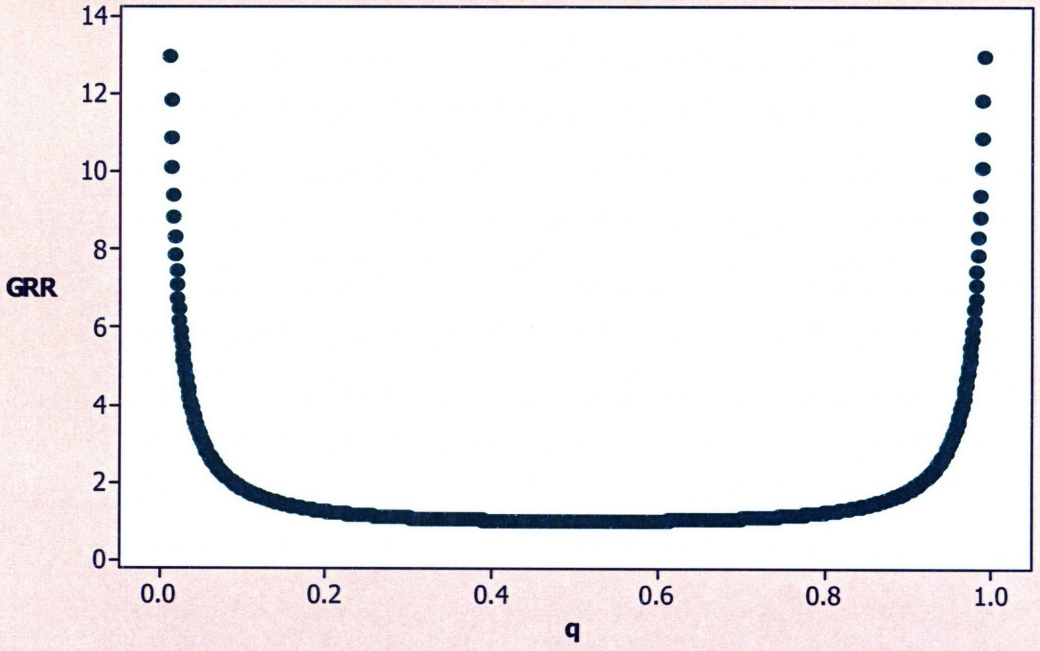


FIGURE 10.3. GRR vs G; multigenic heterozygous risk, m=2

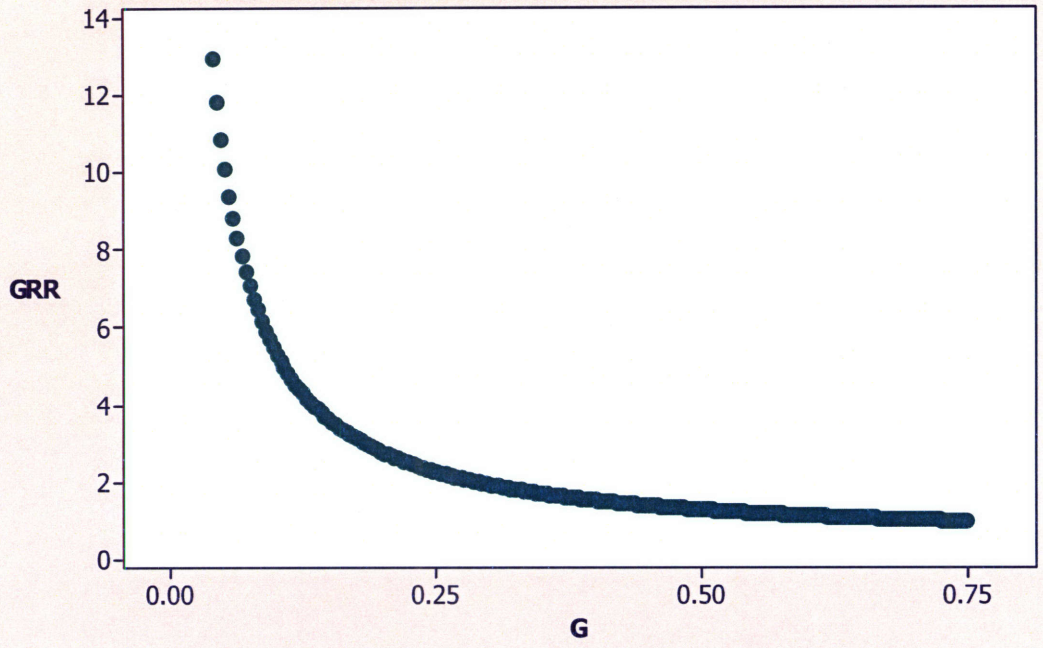


FIGURE 10.4. GRR vs G; multigenic heterozygous risk, m=2

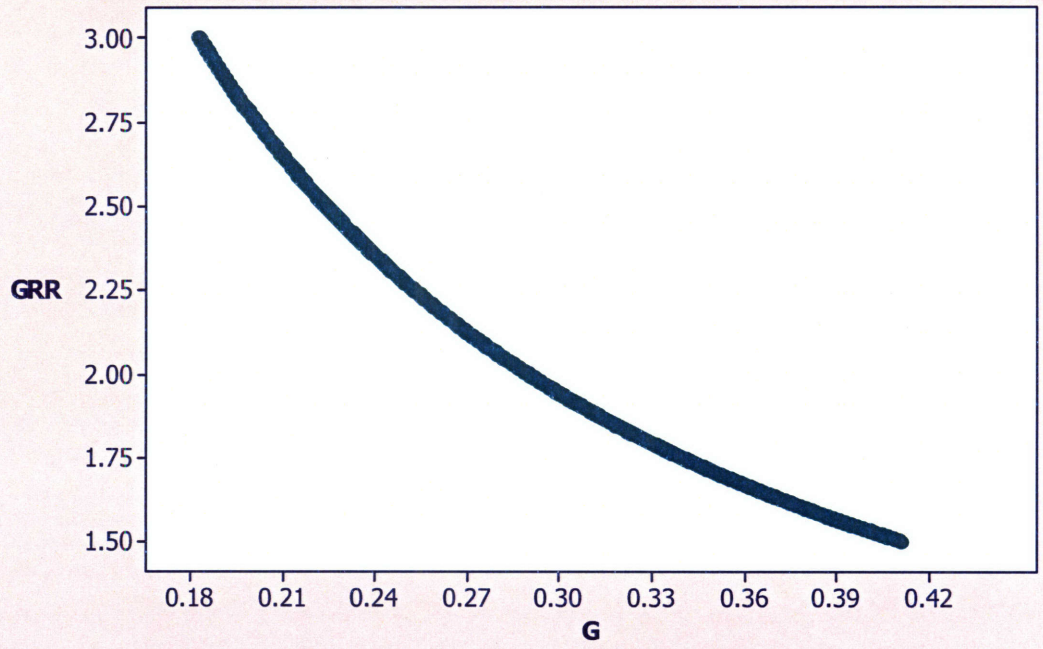


FIGURE 11.1. G vs q; multigenic heterozygous risk, m=3

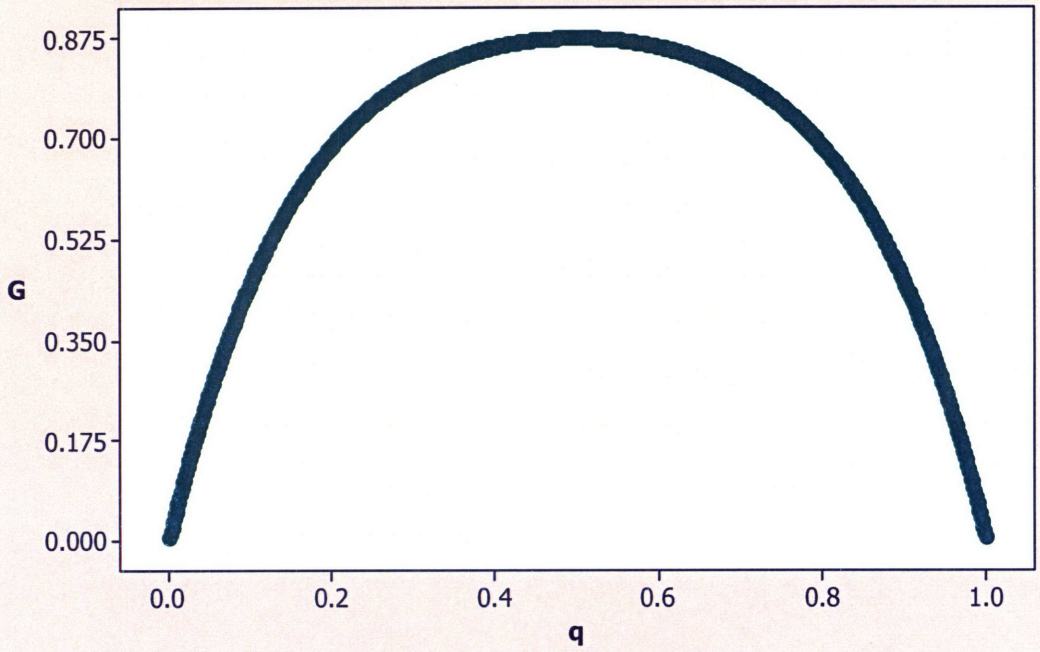
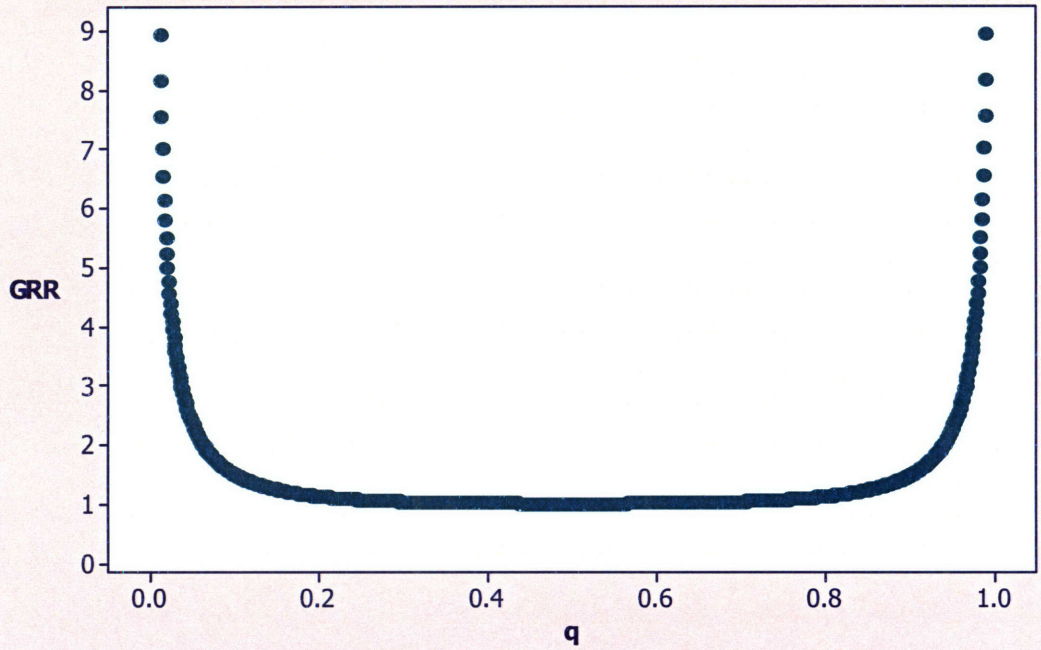


FIGURE 11.2. GRR vs q; multigenic heterozygous risk, m=3



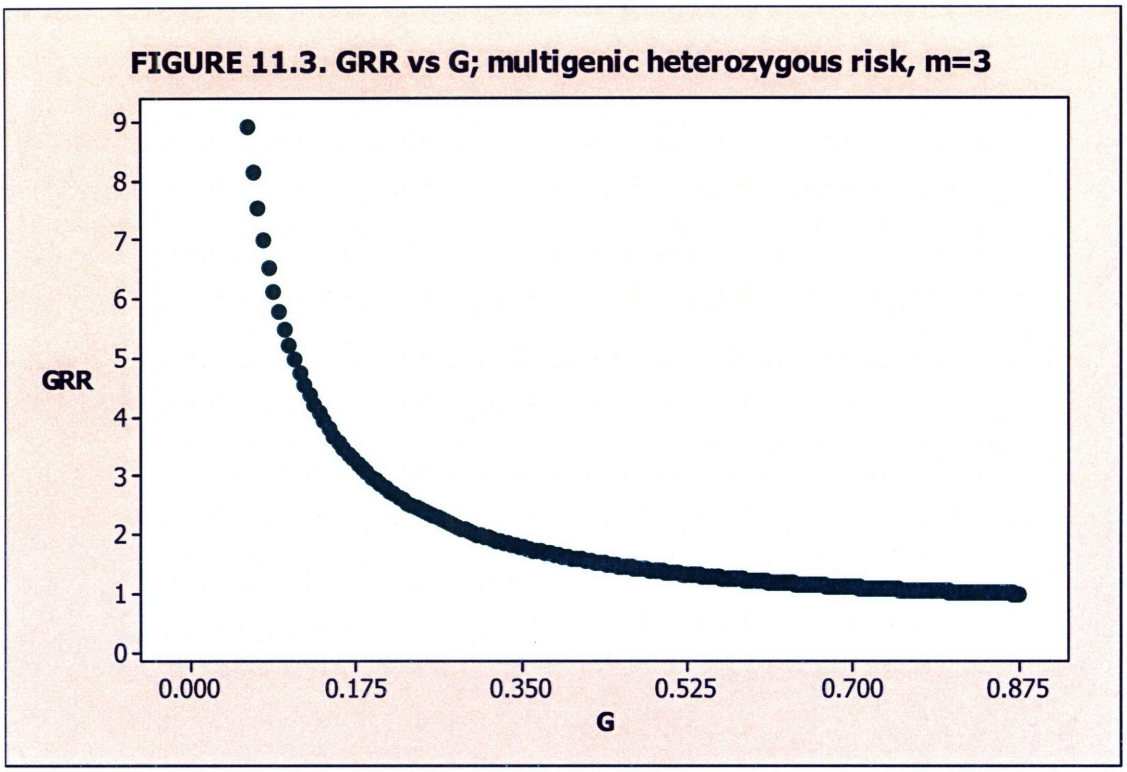


FIGURE 11.3. GRR vs G; multigenic heterozygous risk, m=3
GRR
9
8
7
6
5
4
3
2
1
0
0.000
0.175
0.350
0.525
0.700
0.875
G

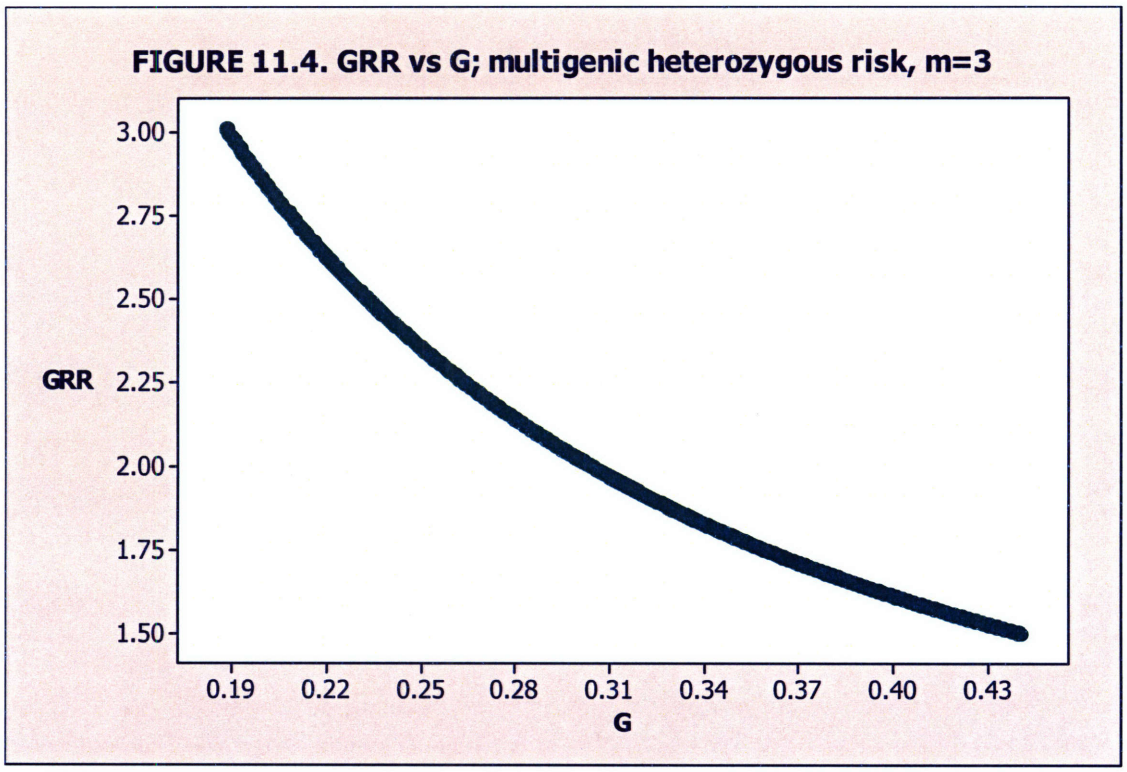


FIGURE 11.4. GRR vs G; multigenic heterozygous risk, m=3
GRR
3.00
2.75
2.50
2.25
2.00
1.75
1.50
0.19
0.22
0.25
0.28
0.31
0.34
0.37
0.40
0.43
G

FIGURE 12.1. G vs q; multigenic heterozygous risk, m=4

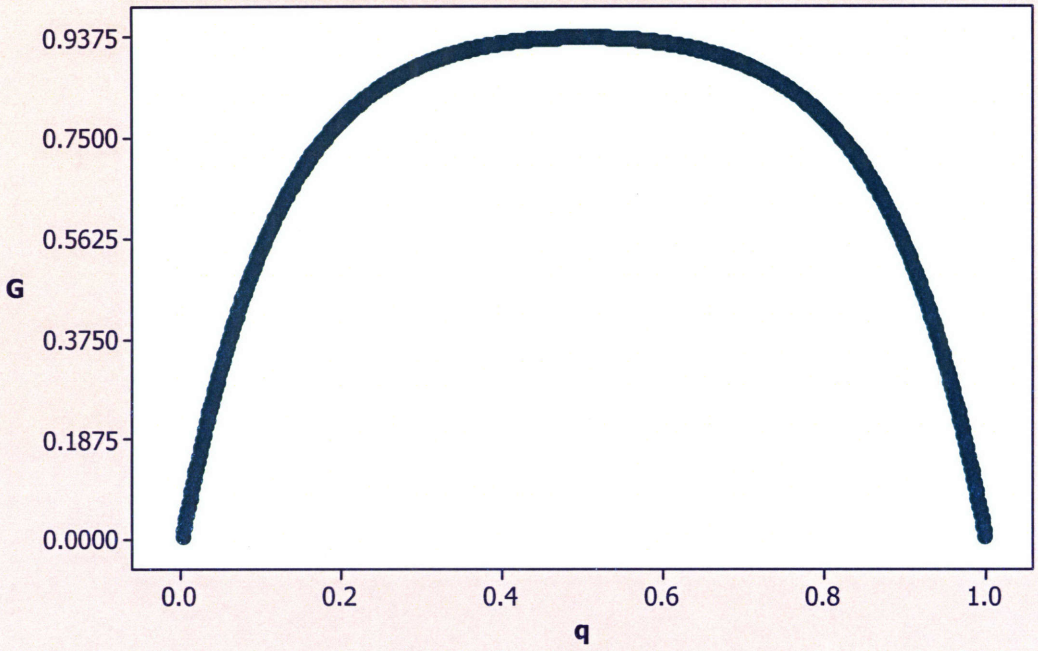


FIGURE 12.2. GRR vs q; multigenic heterozygous risk, m=4

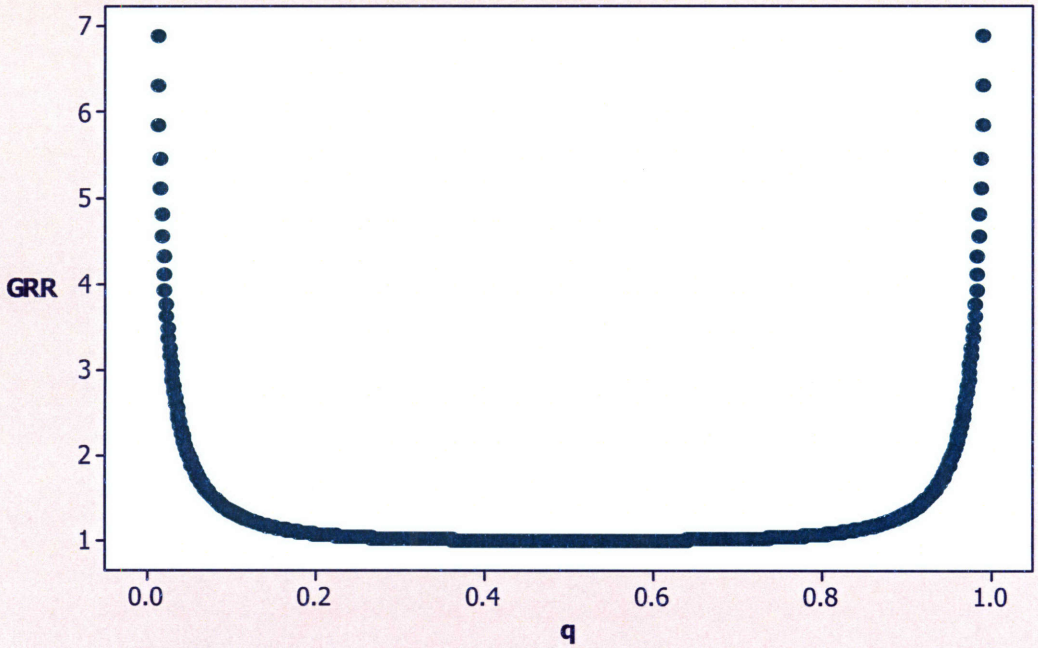


FIGURE 12.3. GRR vs G; multigenic heterozygous risk, m=4

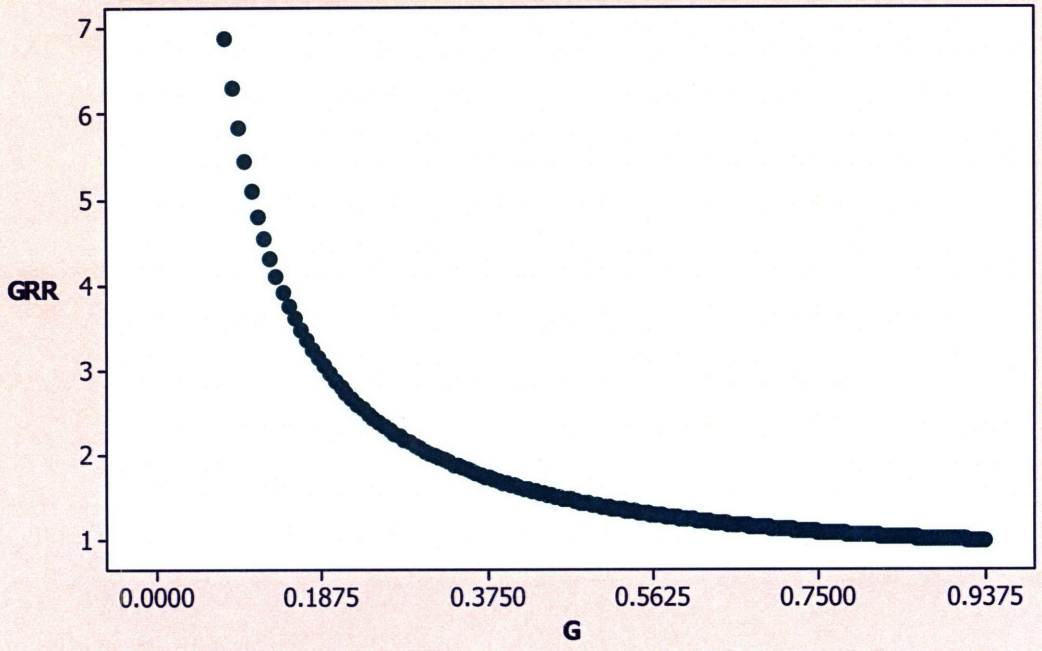


FIGURE 12.4. GRR vs G; multigenic heterozygous risk, m=4

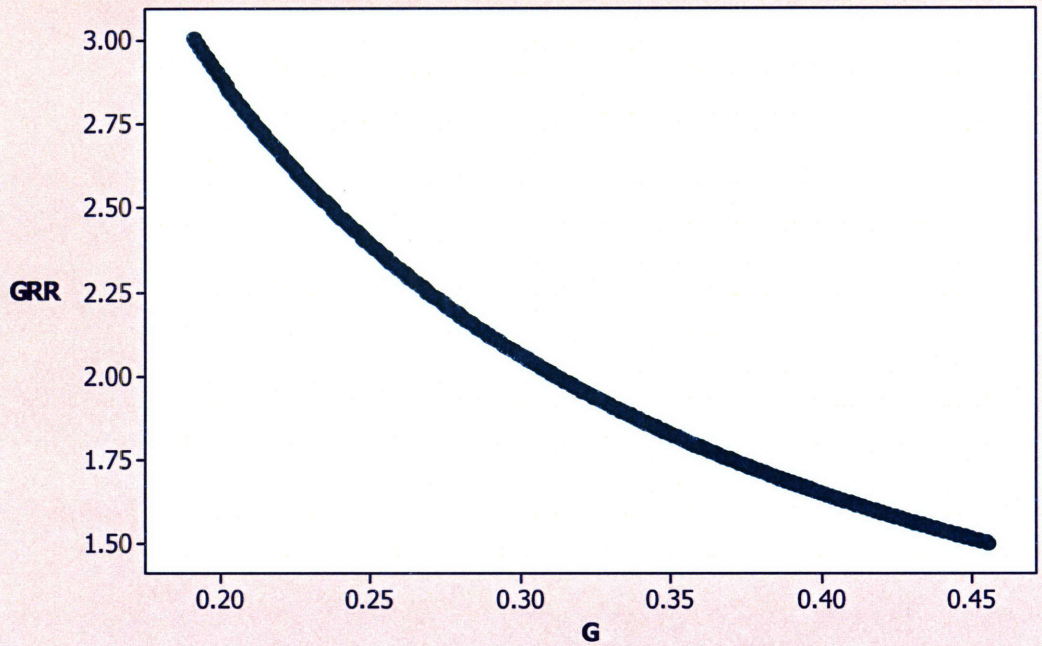


FIGURE 13.1. G vs q; multigenic heterozygous risk, m=5

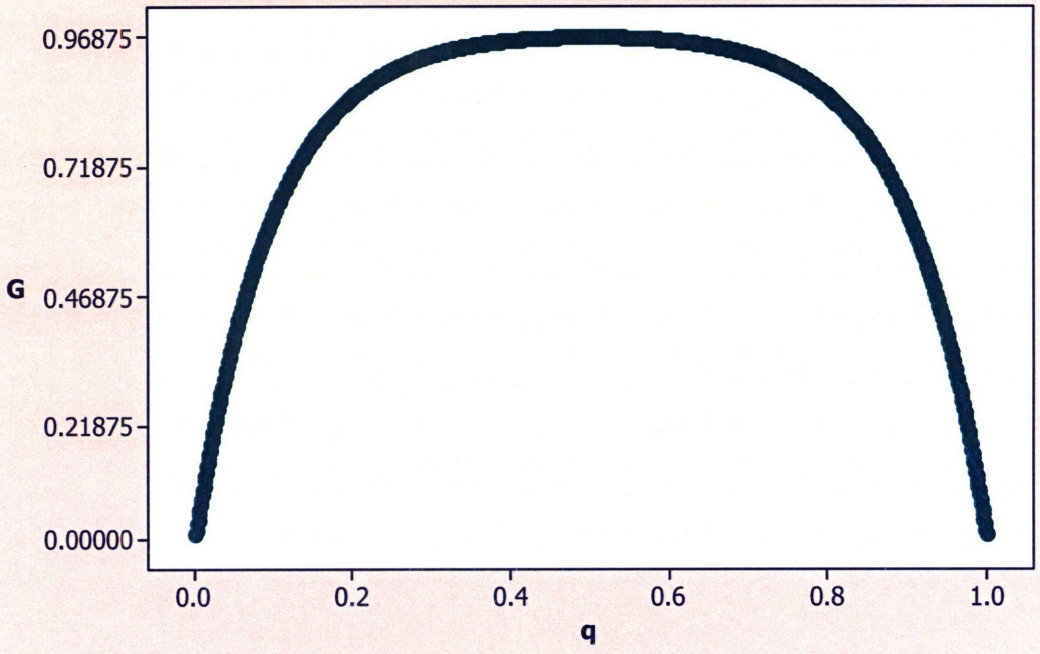


FIGURE 13.2. GRR vs q; multigenic heterozygous risk, m=5

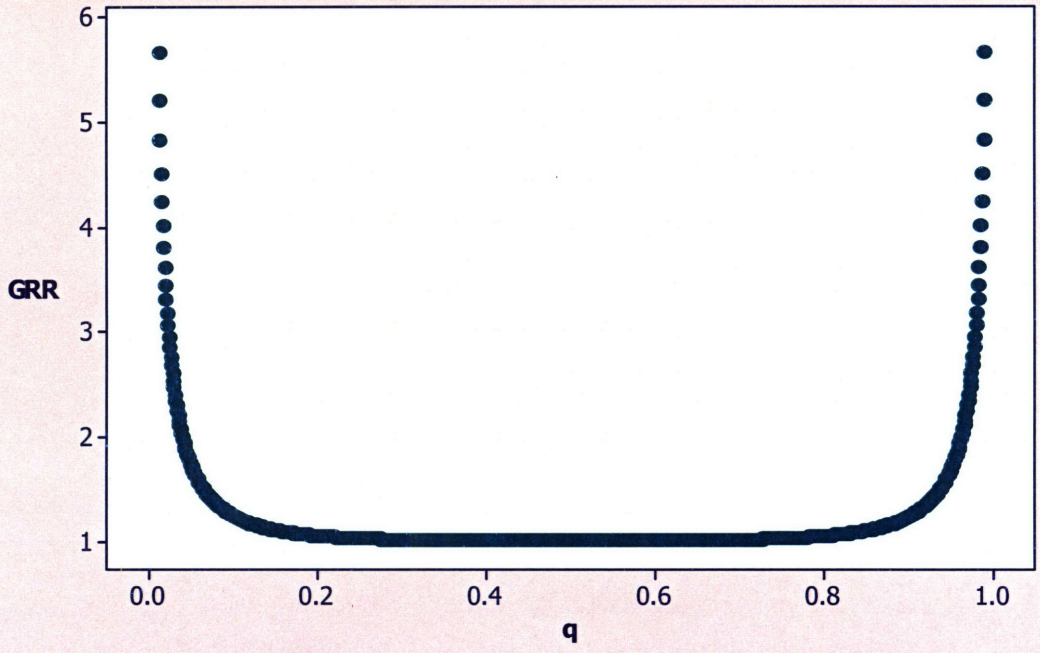


FIGURE 13.3. GRR vs G; multigenic heterozygous risk, m=5

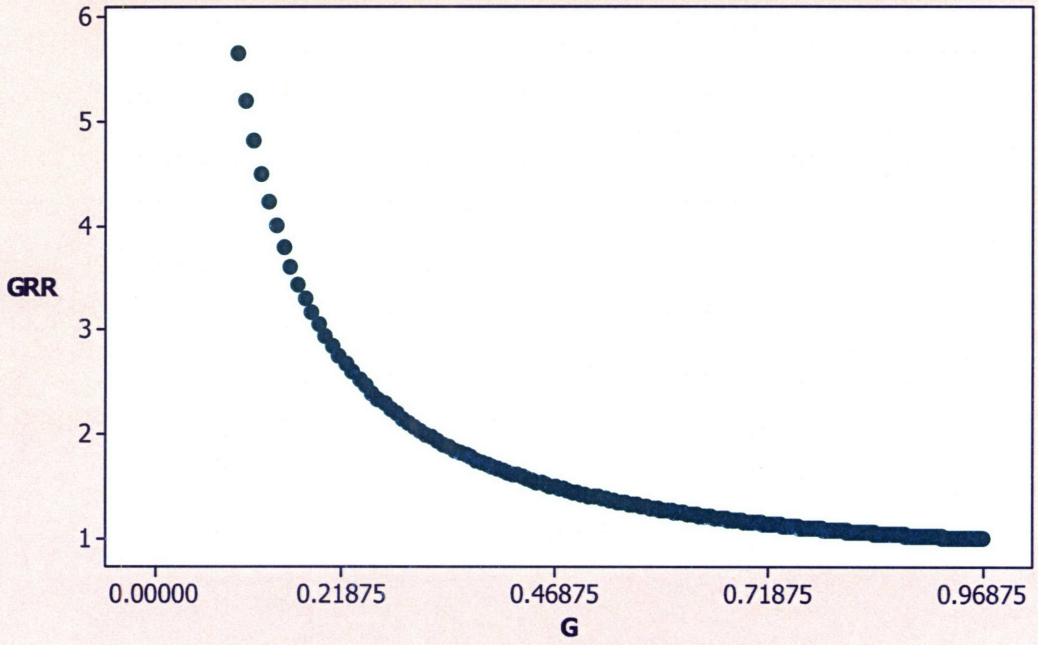
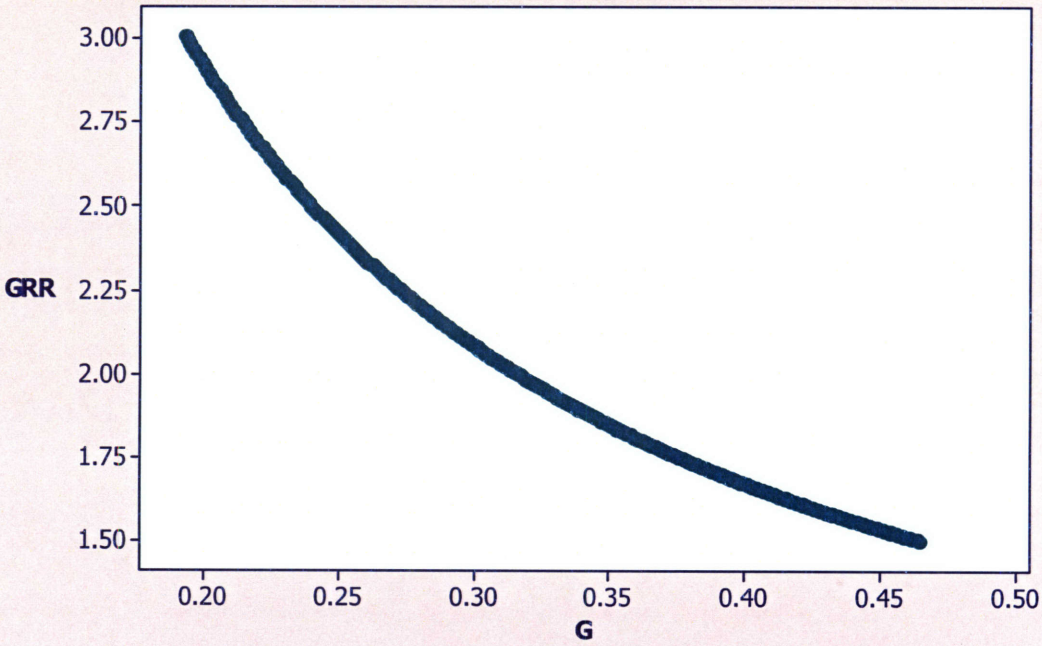


FIGURE 13.4. GRR vs G; multigenic heterozygous risk, m=5



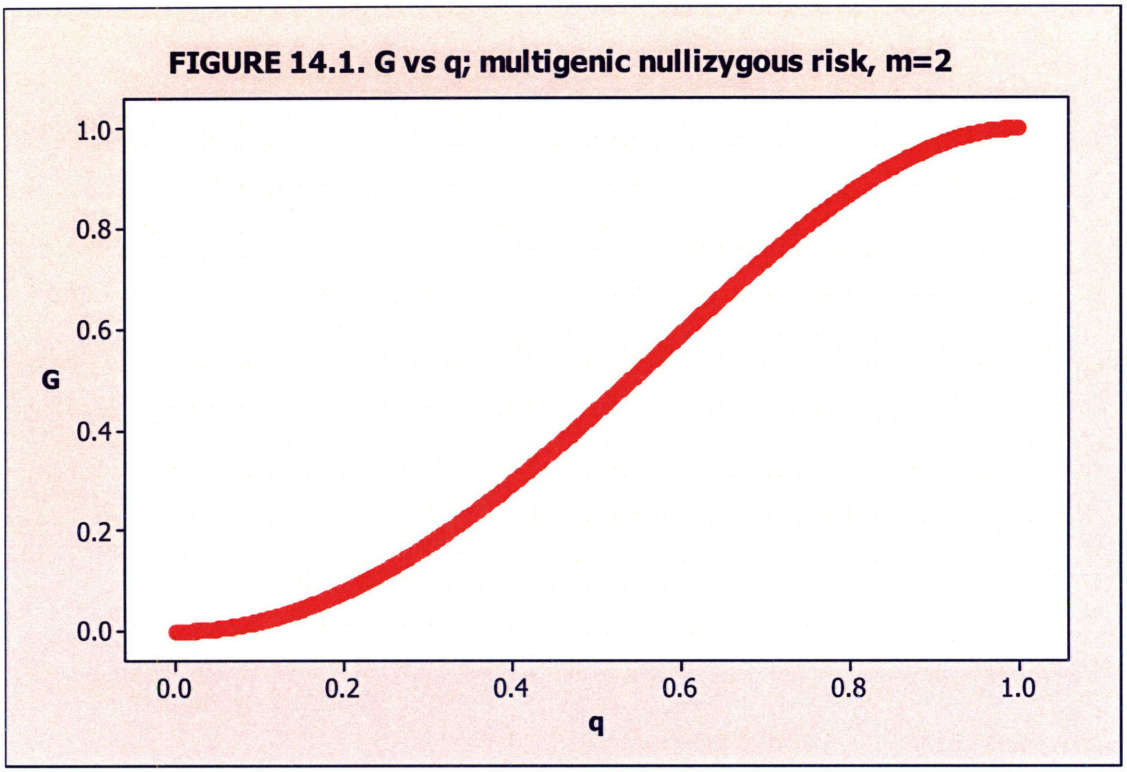


FIGURE 14.1. G vs q; multigenic nullizygous risk, m=2
G
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
q

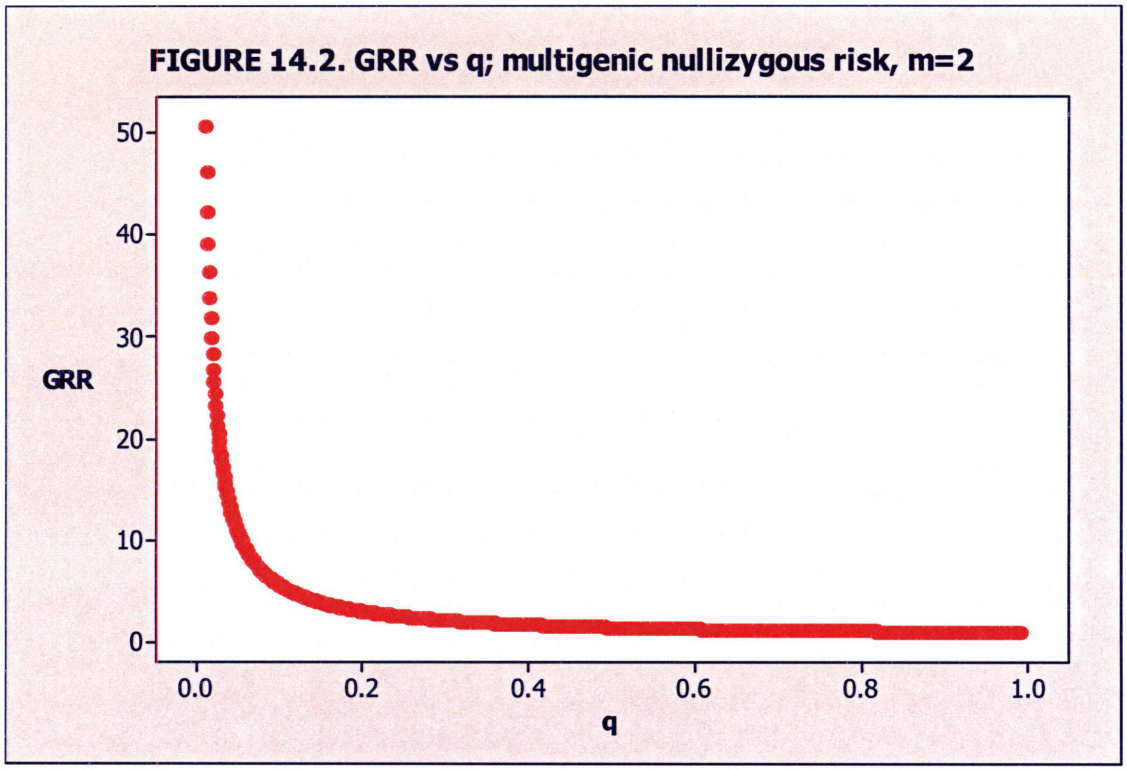


FIGURE 14.2. GRR vs q; multigenic nullizygous risk, m=2
GRR
50
40
30
20
10
0
0.0
0.2
0.4
0.6
0.8
1.0
q

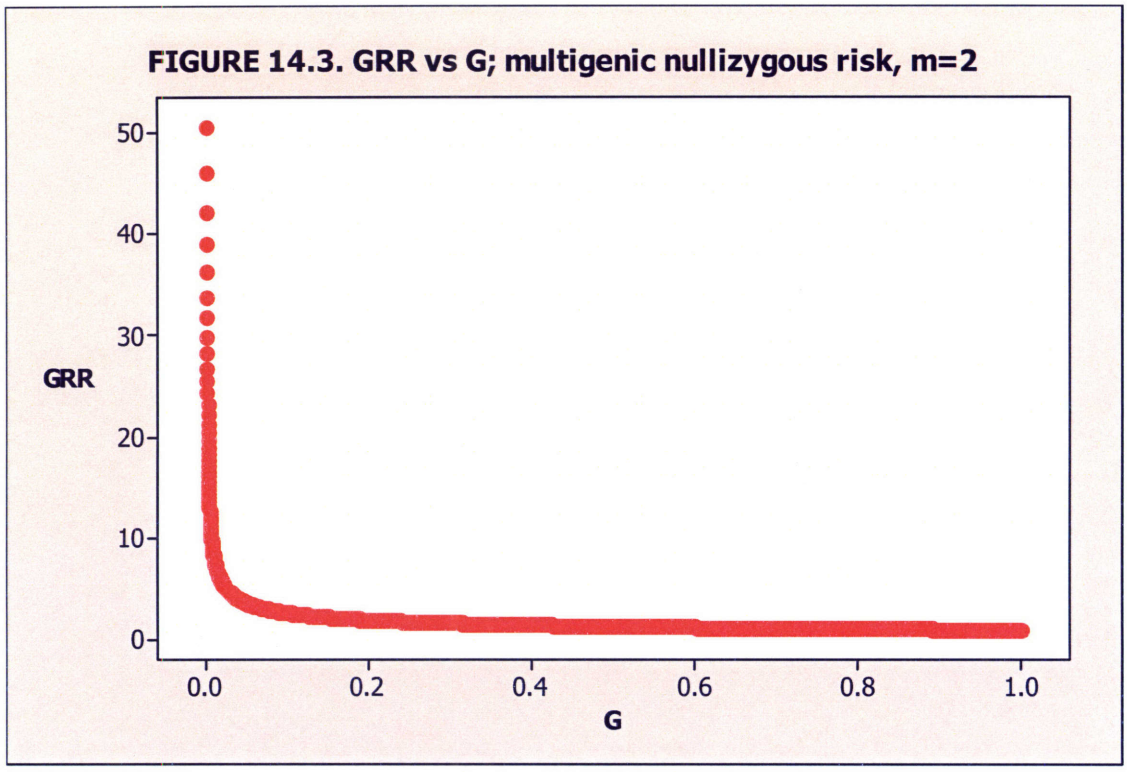


FIGURE 14.3. GRR vs G; multigenic nullizygous risk, m=2
GRR
50
40
30
20
10
0
0.0
0.2
0.4
0.6
0.8
1.0
G

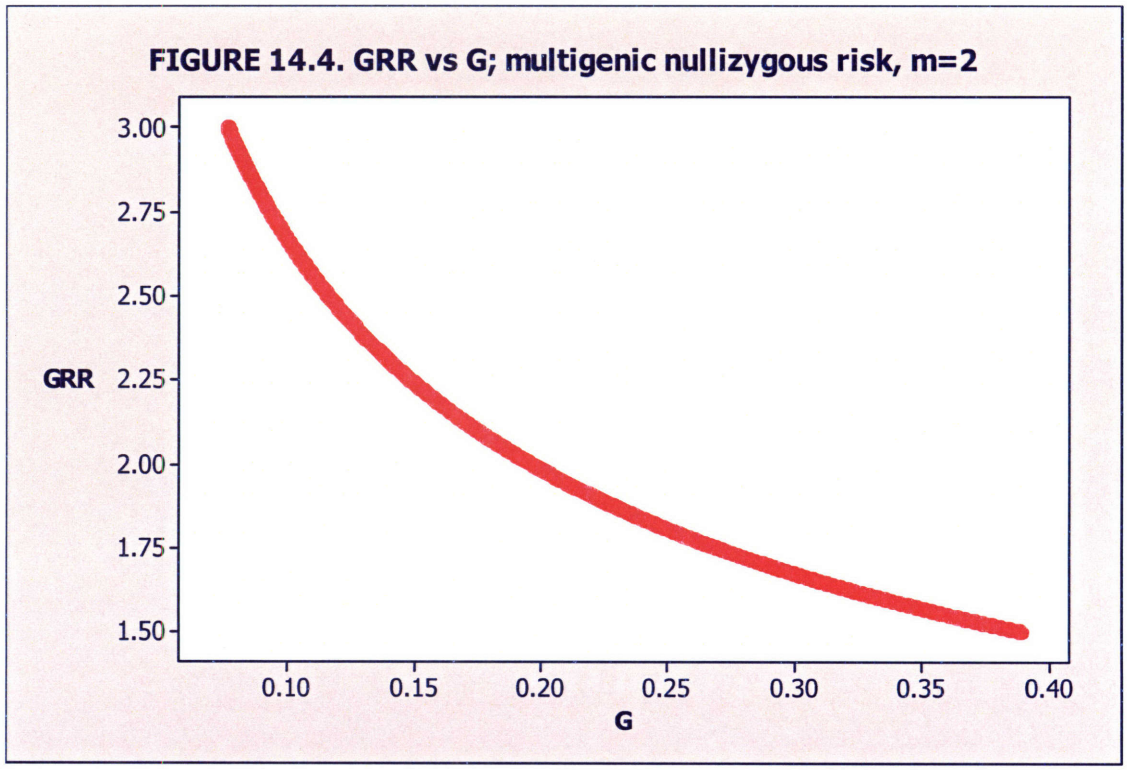


FIGURE 14.4. GRR vs G; multigenic nullizygous risk, m=2
GRR
3.00
2.75
2.50
2.25
2.00
1.75
1.50
0.10
0.15
0.20
0.25
0.30
0.35
0.40
G

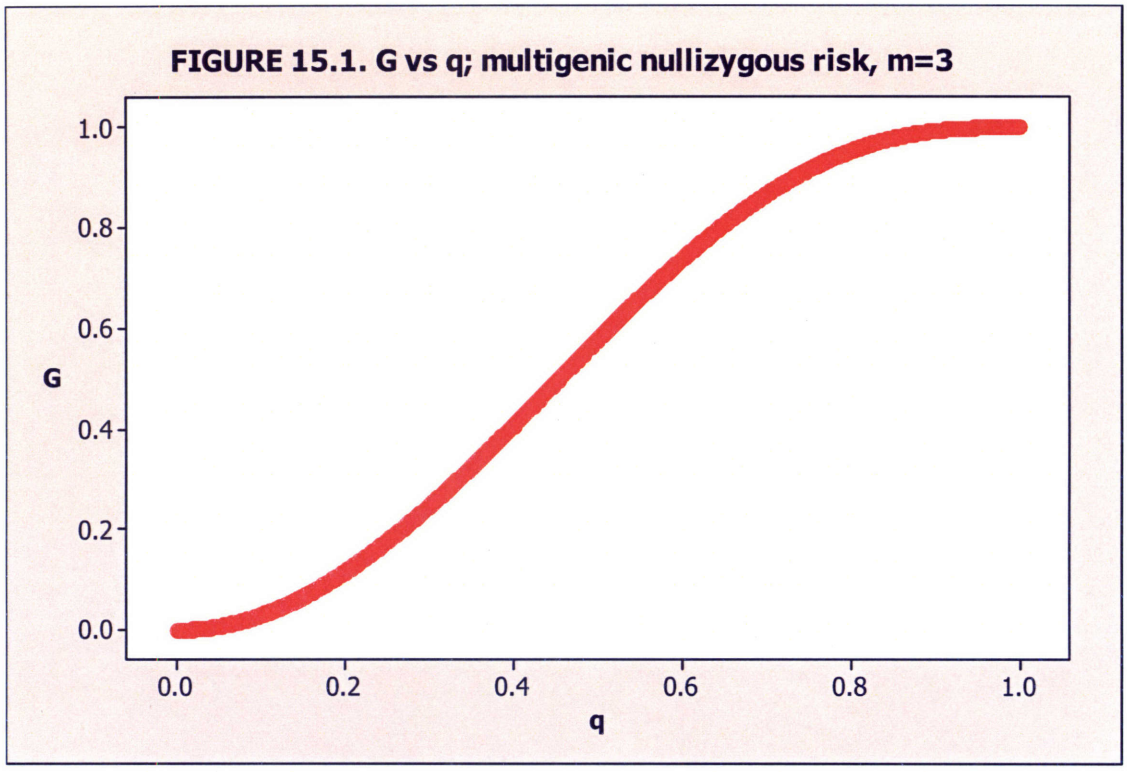


FIGURE 15.1. G vs q; multigenic nullizygous risk, m=3
G
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
q

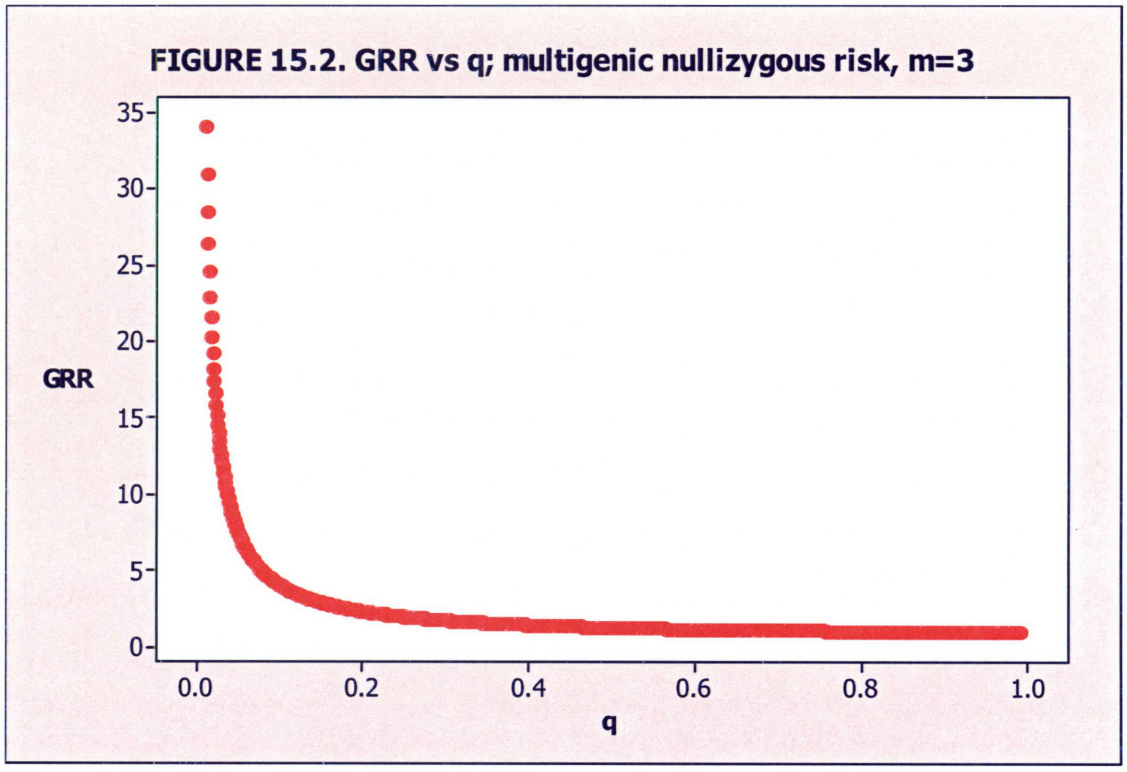


FIGURE 15.2. GRR vs q; multigenic nullizygous risk, m=3
GRR
35
30
25
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
q

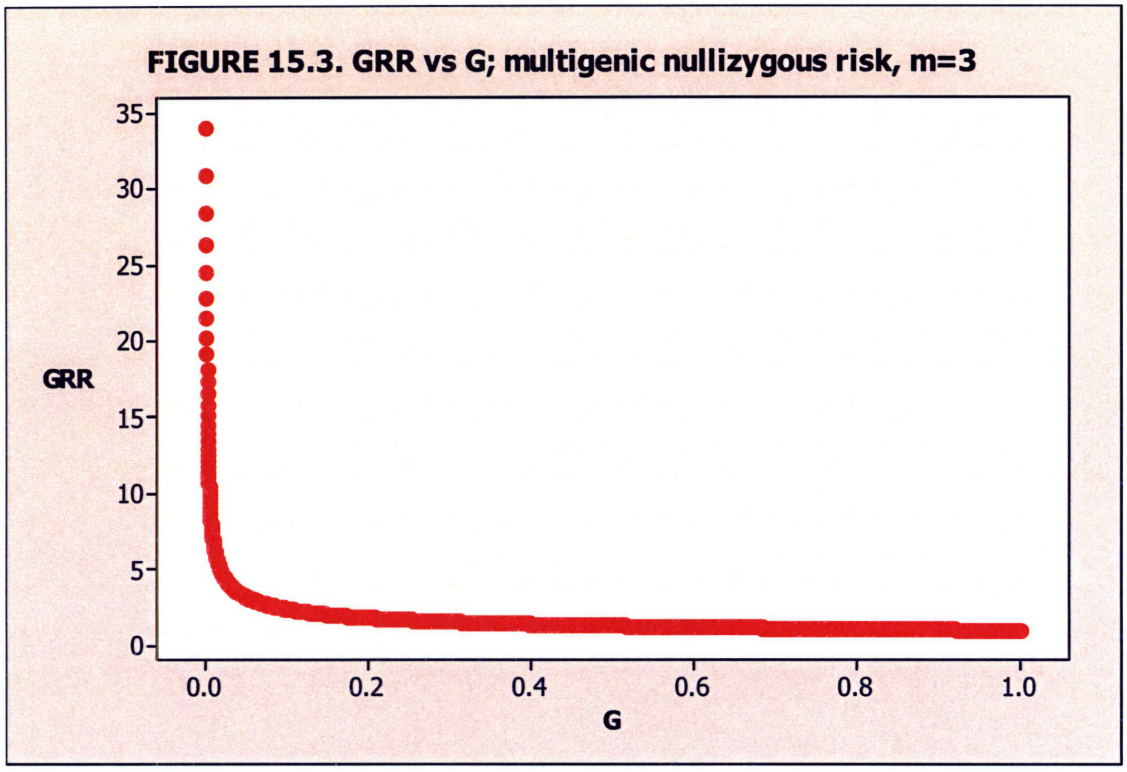


FIGURE 15.3. GRR vs G; multigenic nullizygous risk, m=3
GRR
35
30
25
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
G

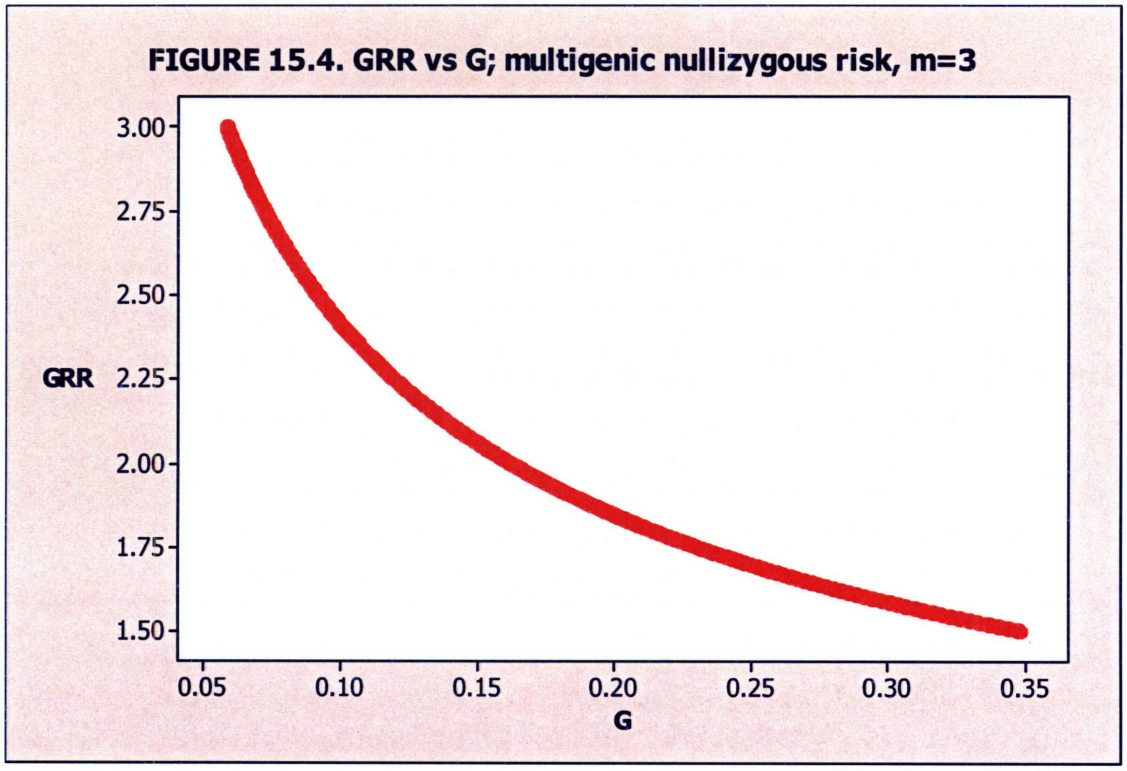


FIGURE 15.4. GRR vs G; multigenic nullizygous risk, m=3
GRR
3.00
2.75
2.50
2.25
2.00
1.75
1.50
0.05
0.10
0.15
0.20
0.25
0.30
0.35
G

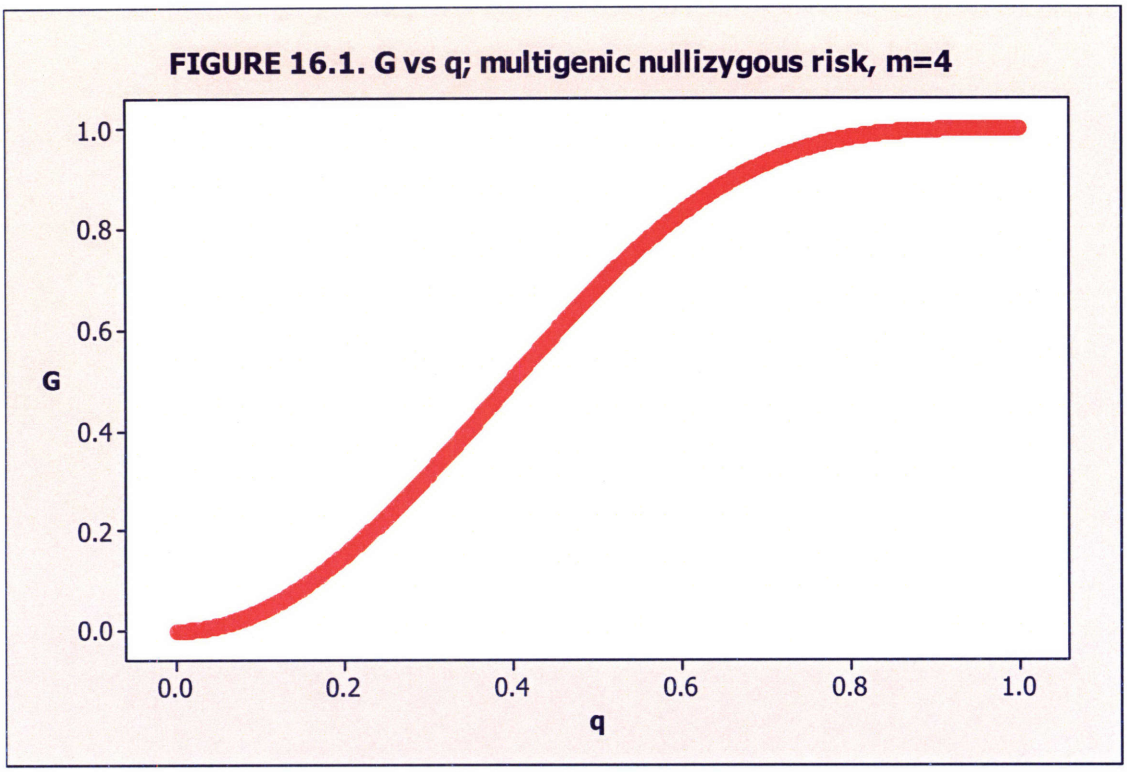


FIGURE 16.1. G vs q; multigenic nullizygous risk, m=4
1.0
0.8
0.6
G
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
q

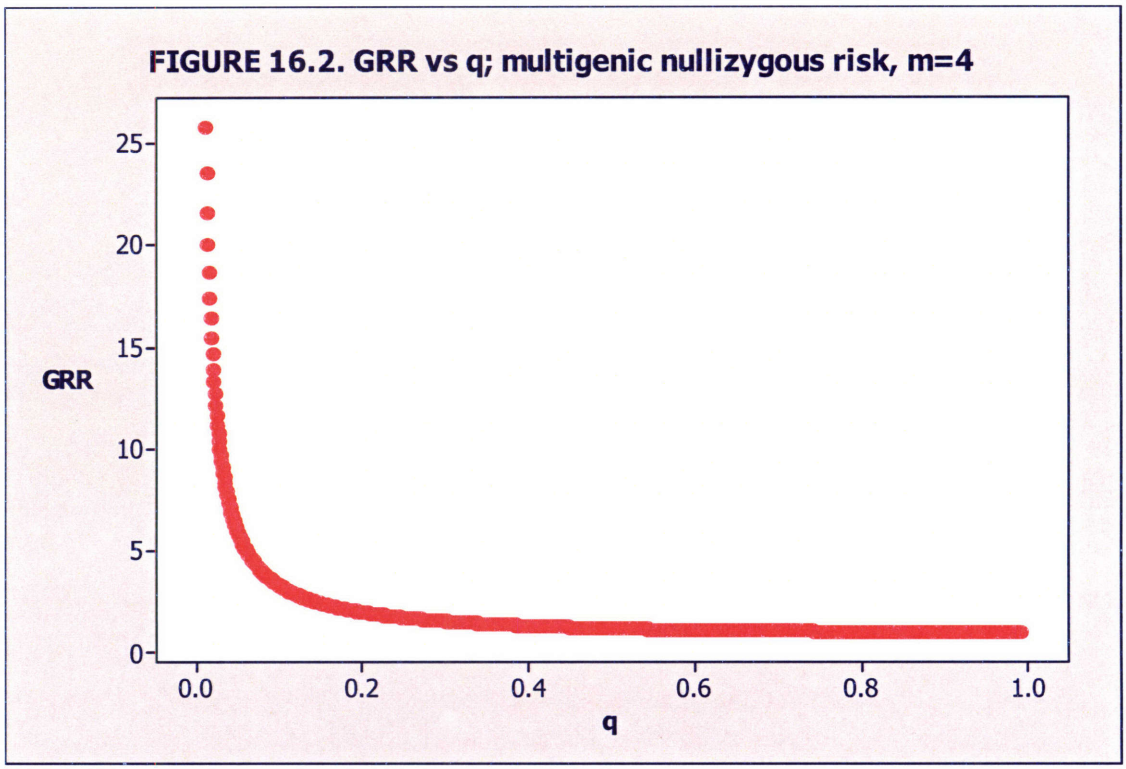


FIGURE 16.2. GRR vs q; multigenic nullizygous risk, m=4
25
20
15
GRR
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
q

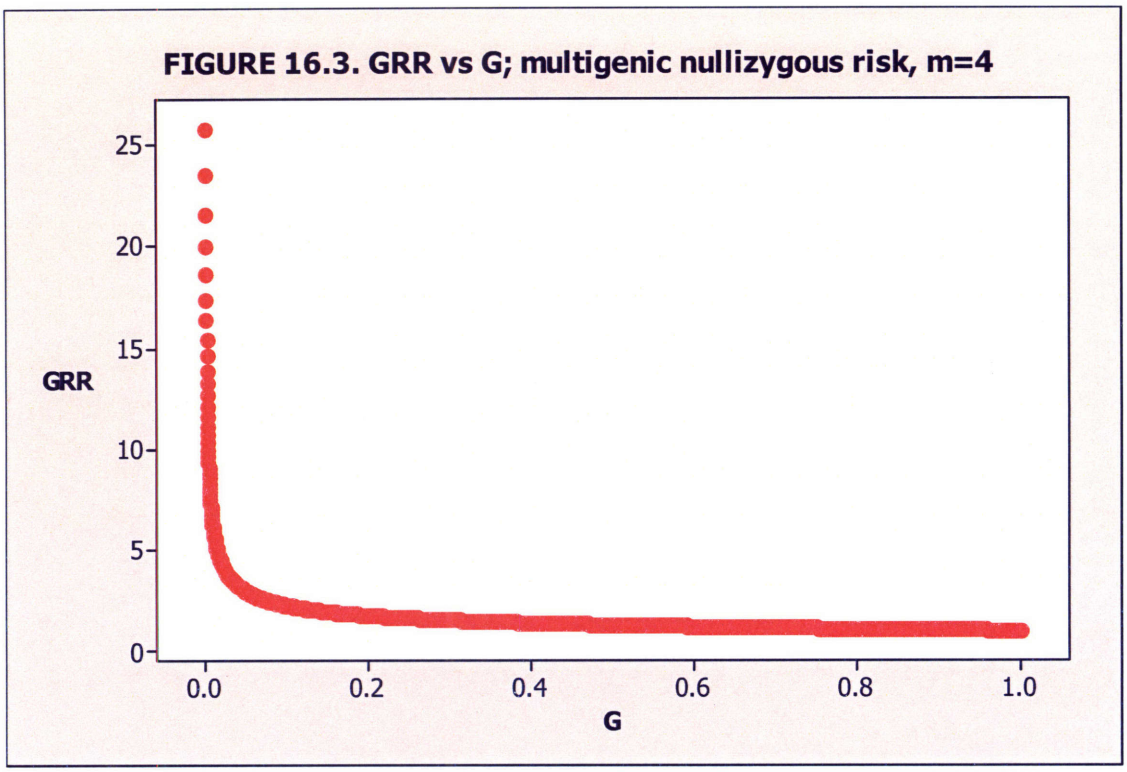


FIGURE 16.3. GRR vs G; multigenic nullizygous risk, m=4
GRR
25
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
G

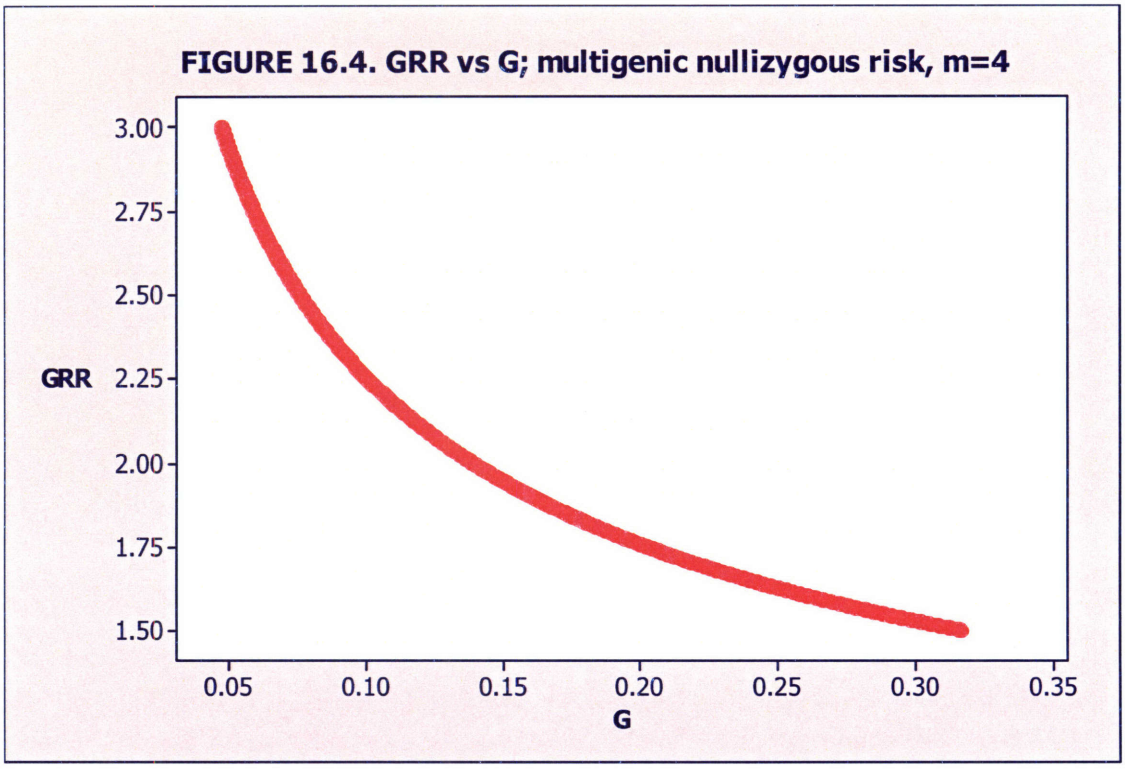


FIGURE 16.4. GRR vs G; multigenic nullizygous risk, m=4
GRR
3.00
2.75
2.50
2.25
2.00
1.75
1.50
0.05
0.10
0.15
0.20
0.25
0.30
0.35
G

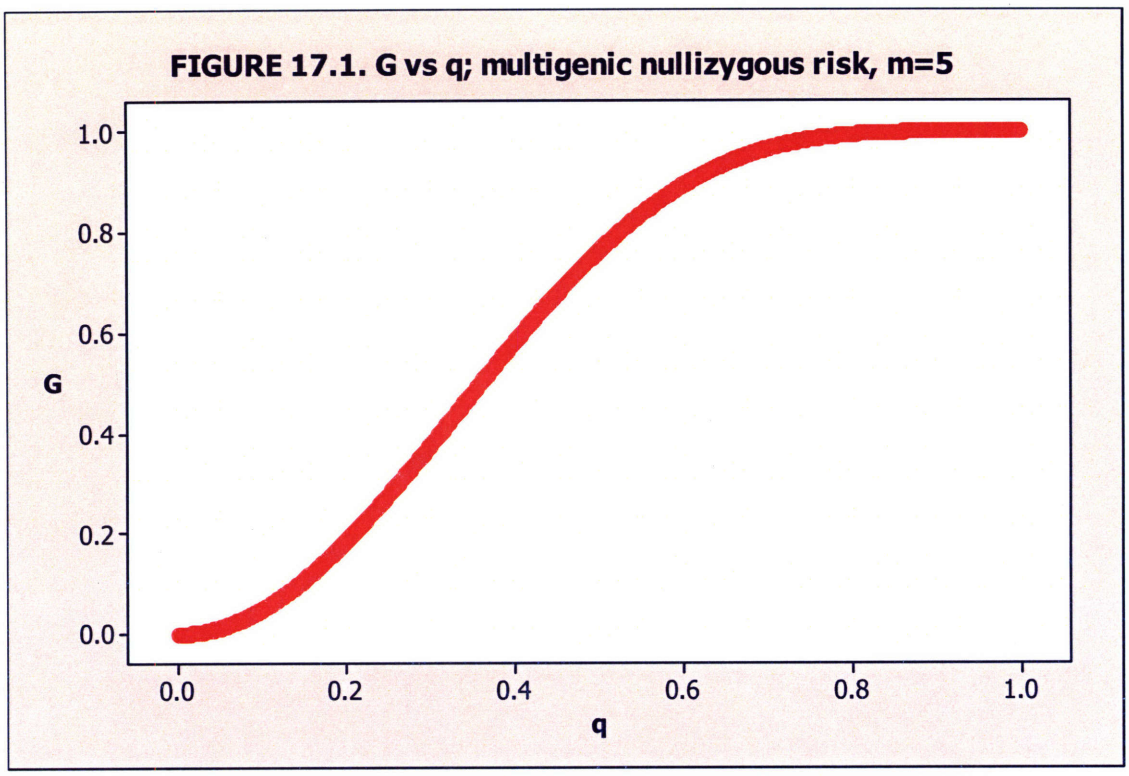


FIGURE 17.1. G vs q; multigenic nullizygous risk, m=5
G
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
q

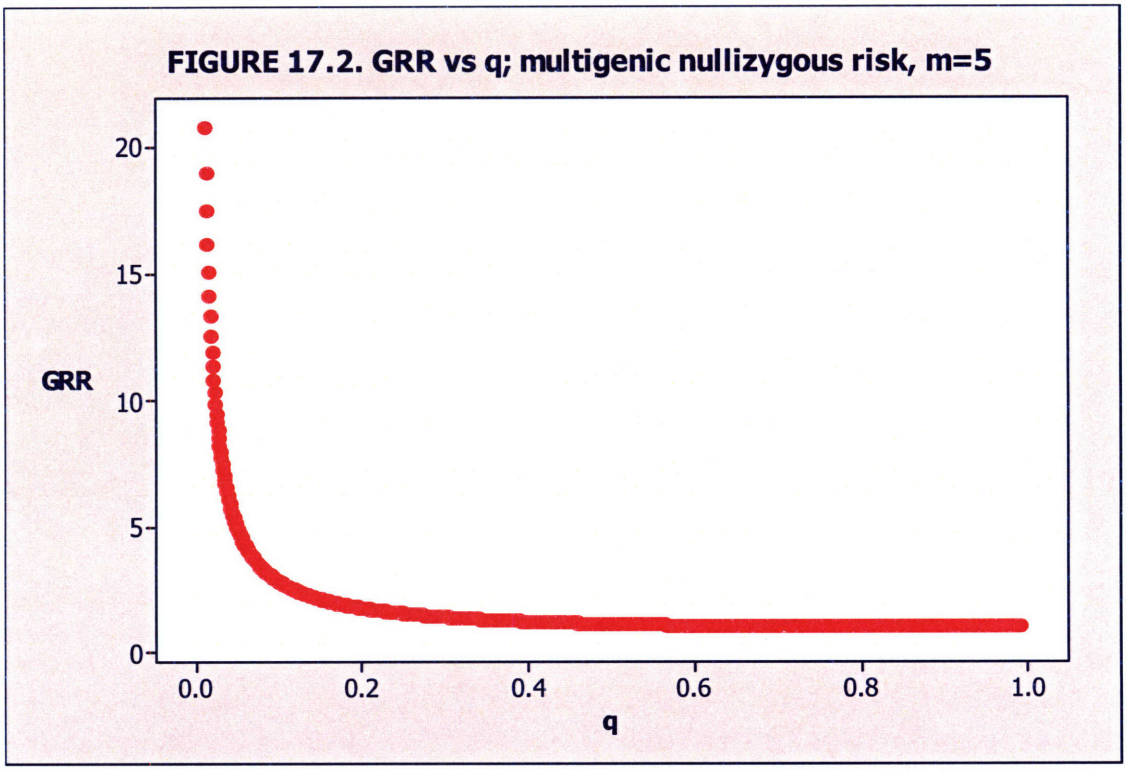


FIGURE 17.2. GRR vs q; multigenic nullizygous risk, m=5
GRR
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
q

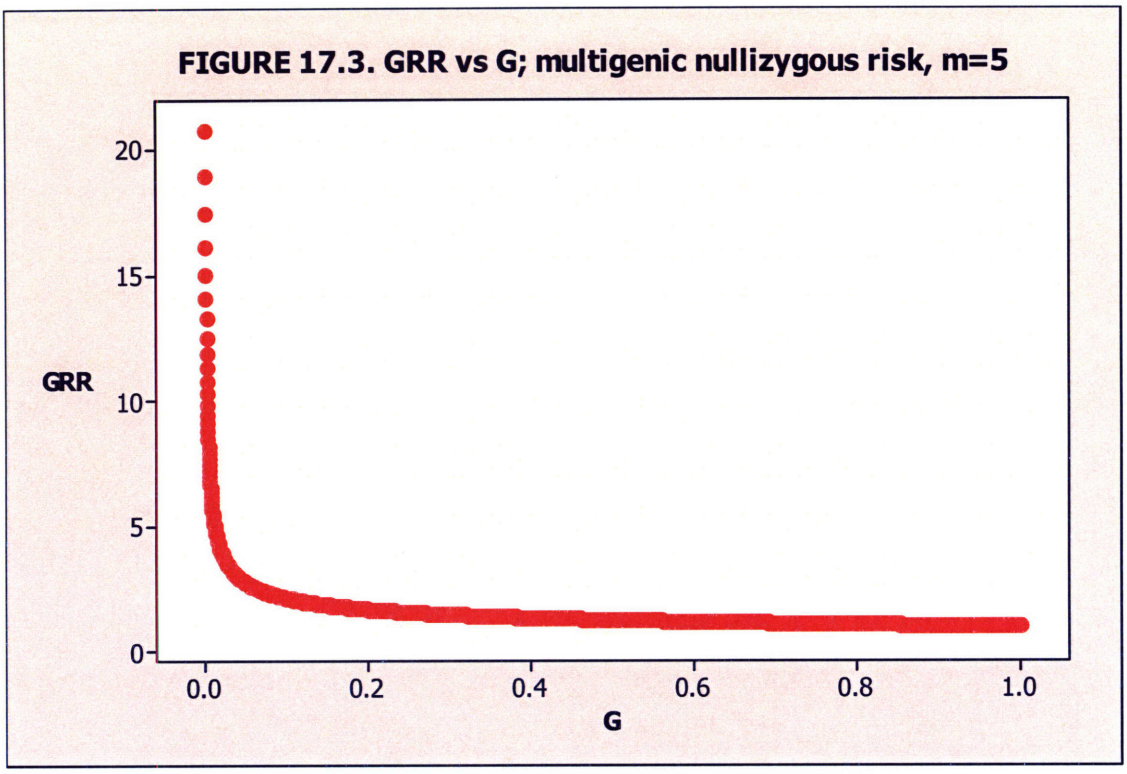


FIGURE 17.3. GRR vs G; multigenic nullizygous risk, m=5
GRR
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
G

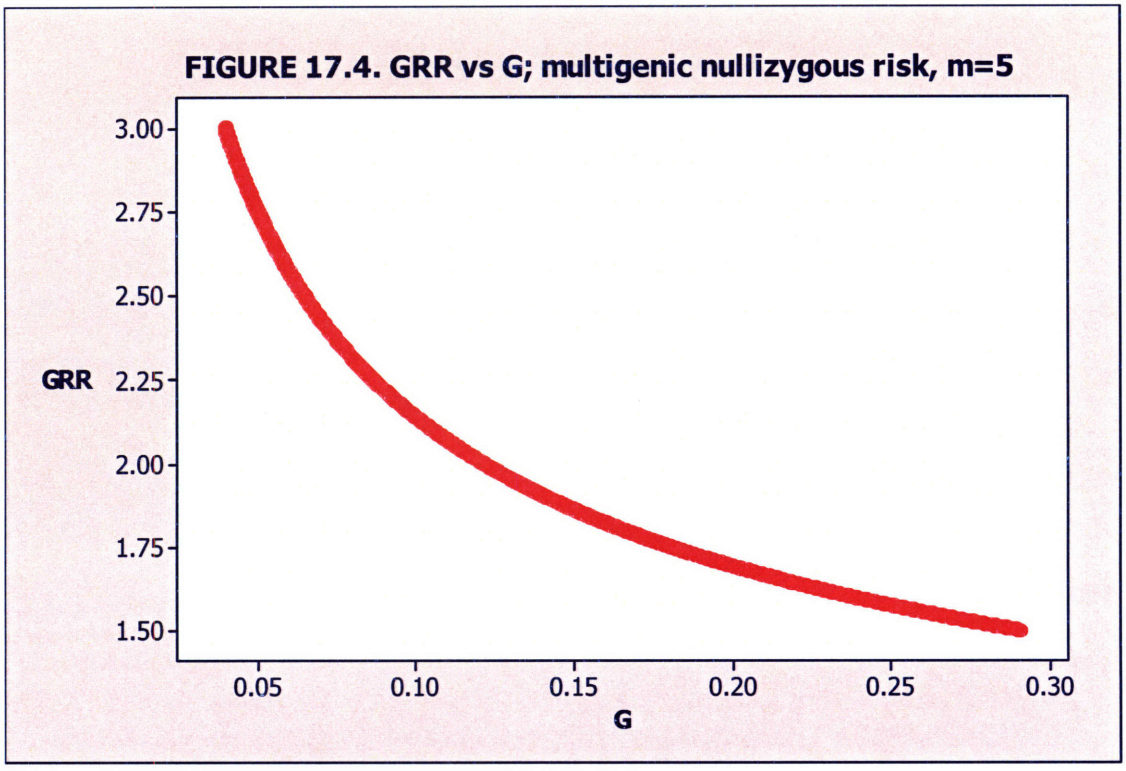


FIGURE 17.4. GRR vs G; multigenic nullizygous risk, m=5
GRR
3.00
2.75
2.50
2.25
2.00
1.75
1.50
0.05
0.10
0.15
0.20
0.25
0.30
G

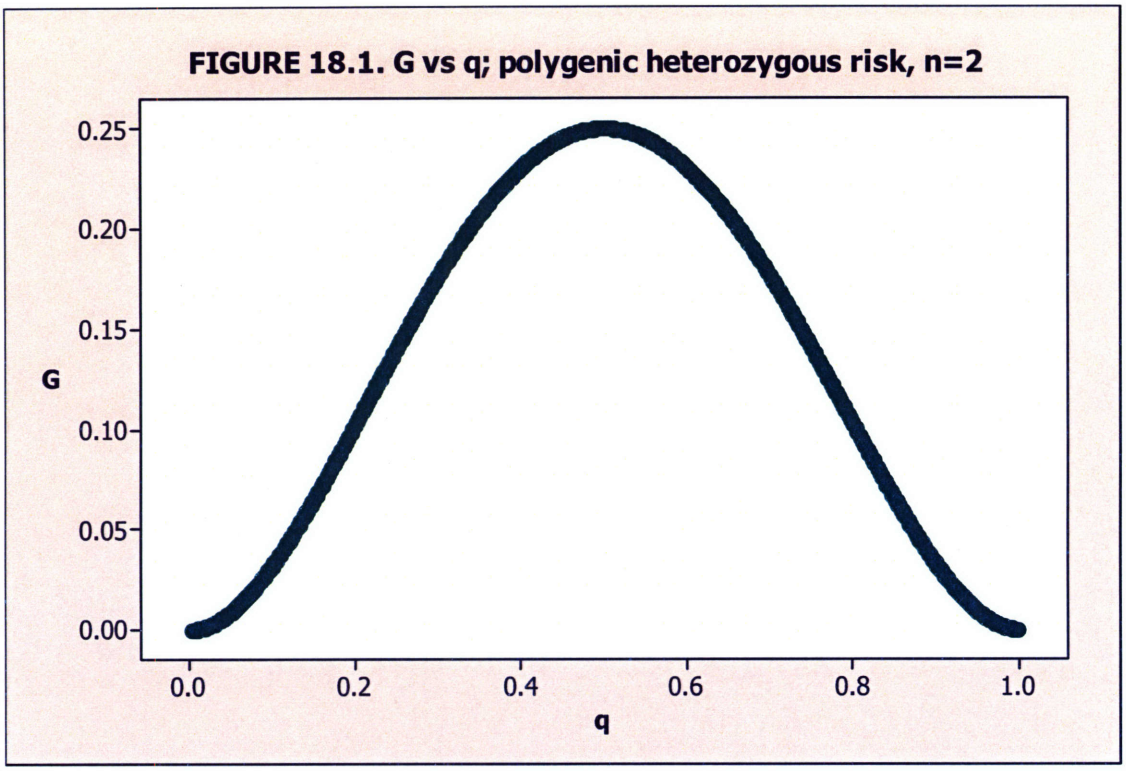


FIGURE 18.1. G vs q; polygenic heterozygous risk, n=2
G
0.25
0.20
0.15
0.10
0.05
0.00
0.0
0.2
0.4
0.6
0.8
1.0
q

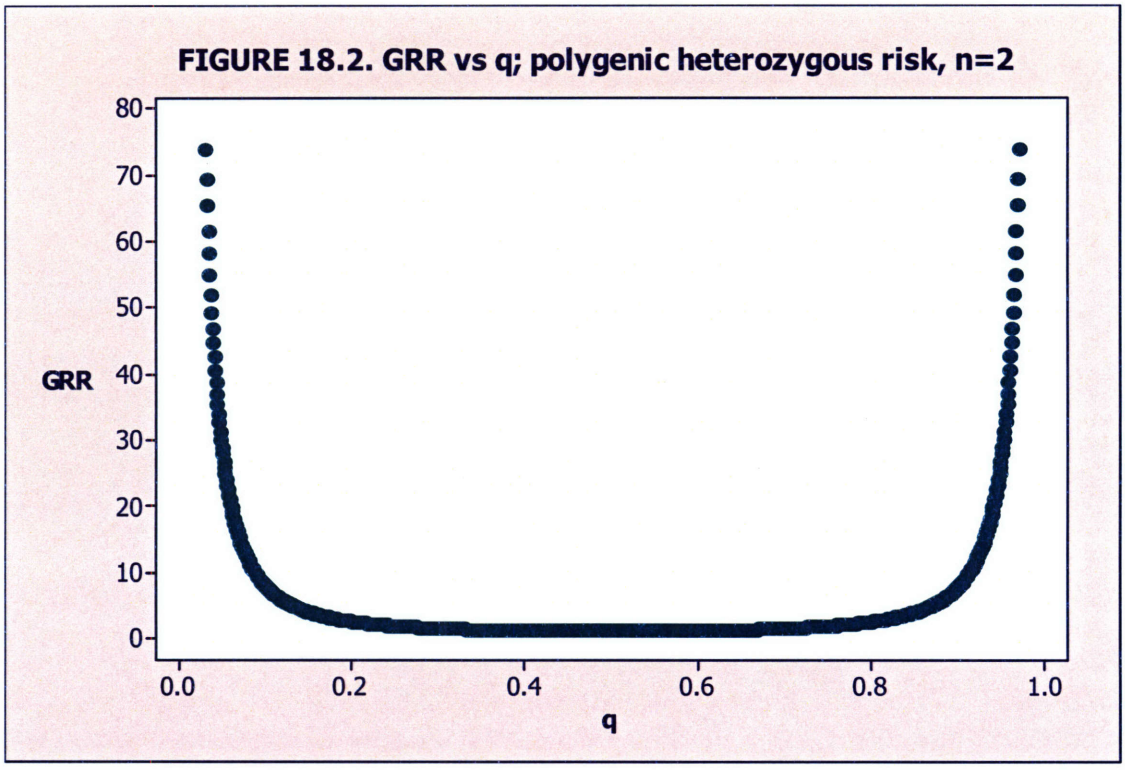


FIGURE 18.2. GRR vs q; polygenic heterozygous risk, n=2
GRR
80
70
60
50
40
30
20
10
0
0.0
0.2
0.4
0.6
0.8
1.0
q

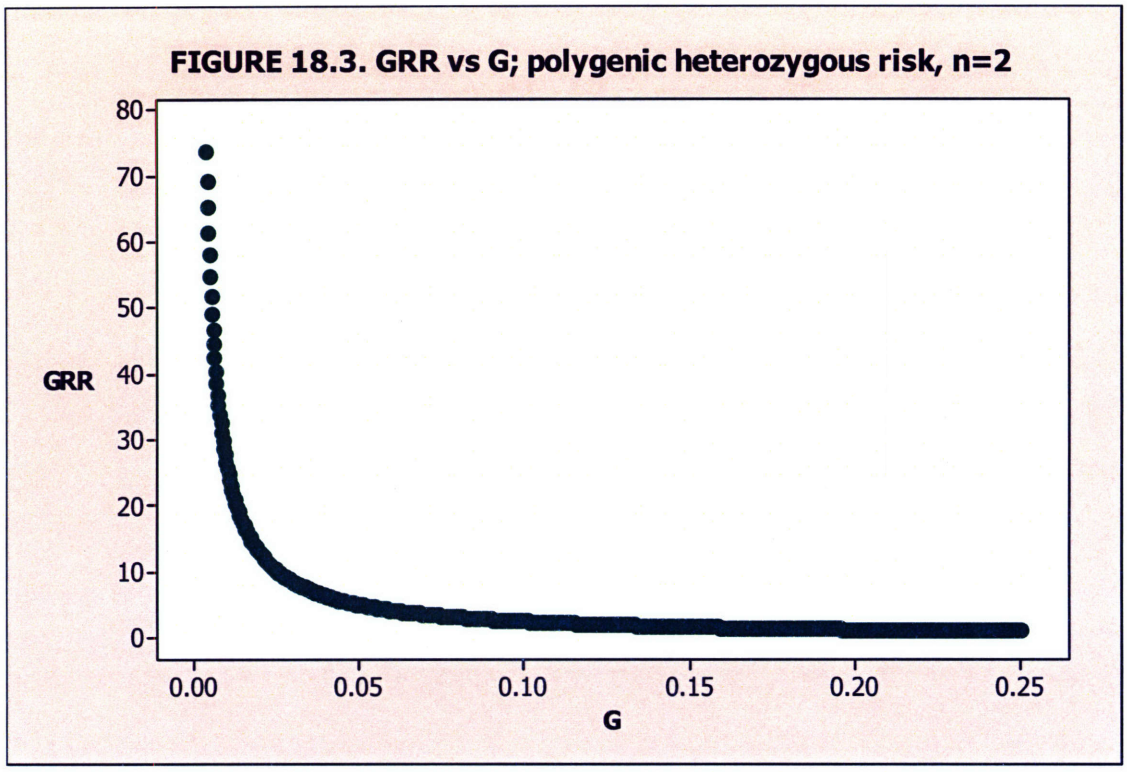


FIGURE 18.3. GRR vs G; polygenic heterozygous risk, n=2
GRR
80
70
60
50
40
30
20
10
0
0.00
0.05
0.10
0.15
0.20
0.25
G

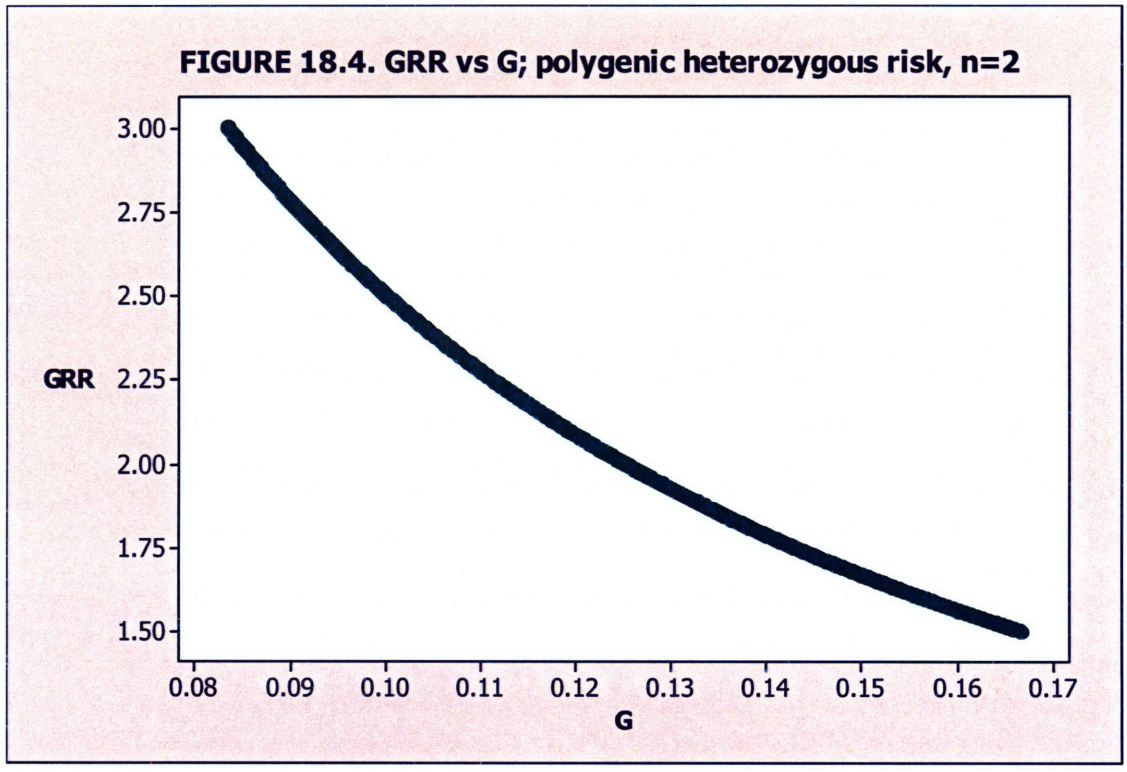


FIGURE 18.4. GRR vs G; polygenic heterozygous risk, n=2
GRR
3.00
2.75
2.50
2.25
2.00
1.75
1.50
0.08
0.09
0.10
0.11
0.12
0.13
0.14
0.15
0.16
0.17
G

FIGURE 19.1. G vs q; polygenic nullizygous risk, n=2, 5, 15, 25

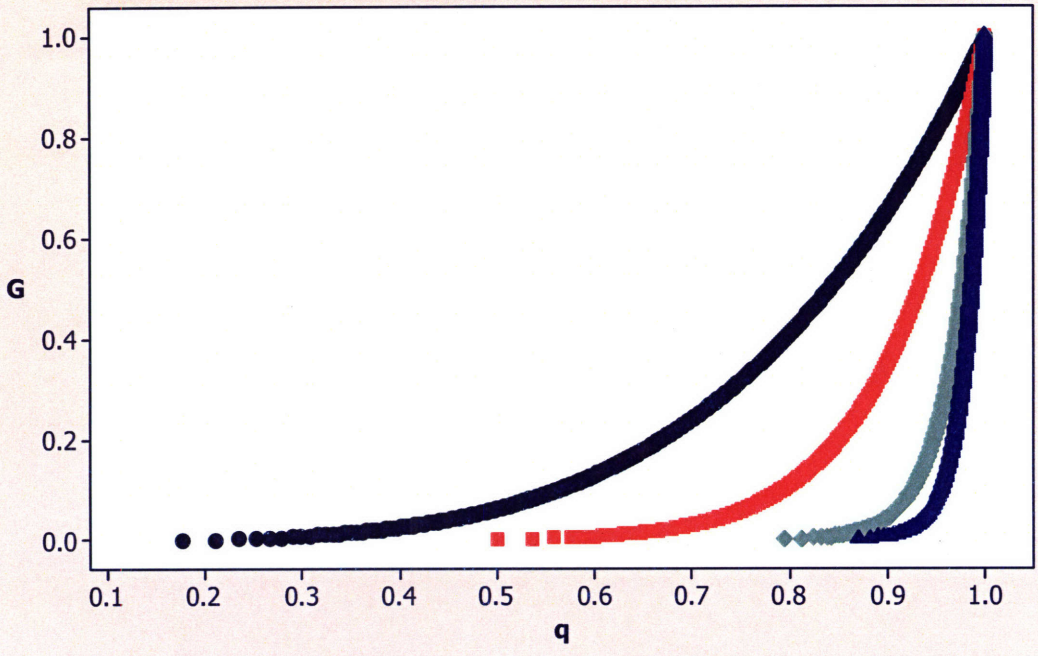
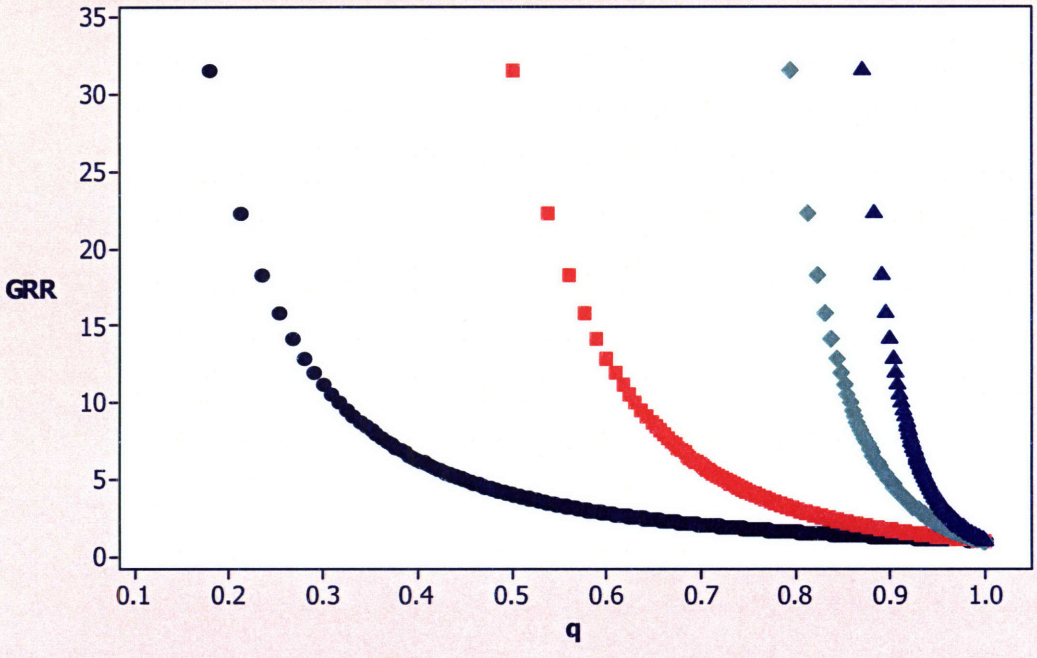


FIGURE 19.2. GRR vs q; polygenic nullizygous risk, n=2, 5, 15, 25



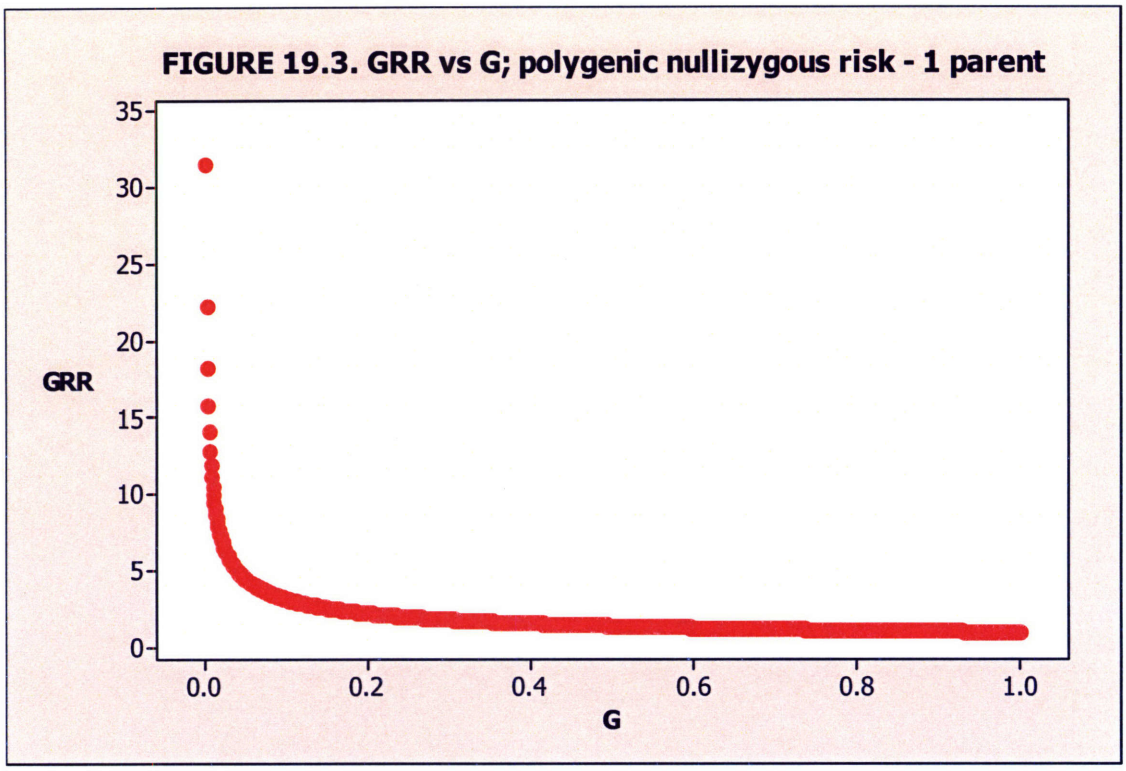


FIGURE 19.3. GRR vs G; polygenic nullizygous risk - 1 parent
GRR
35
30
25
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
G

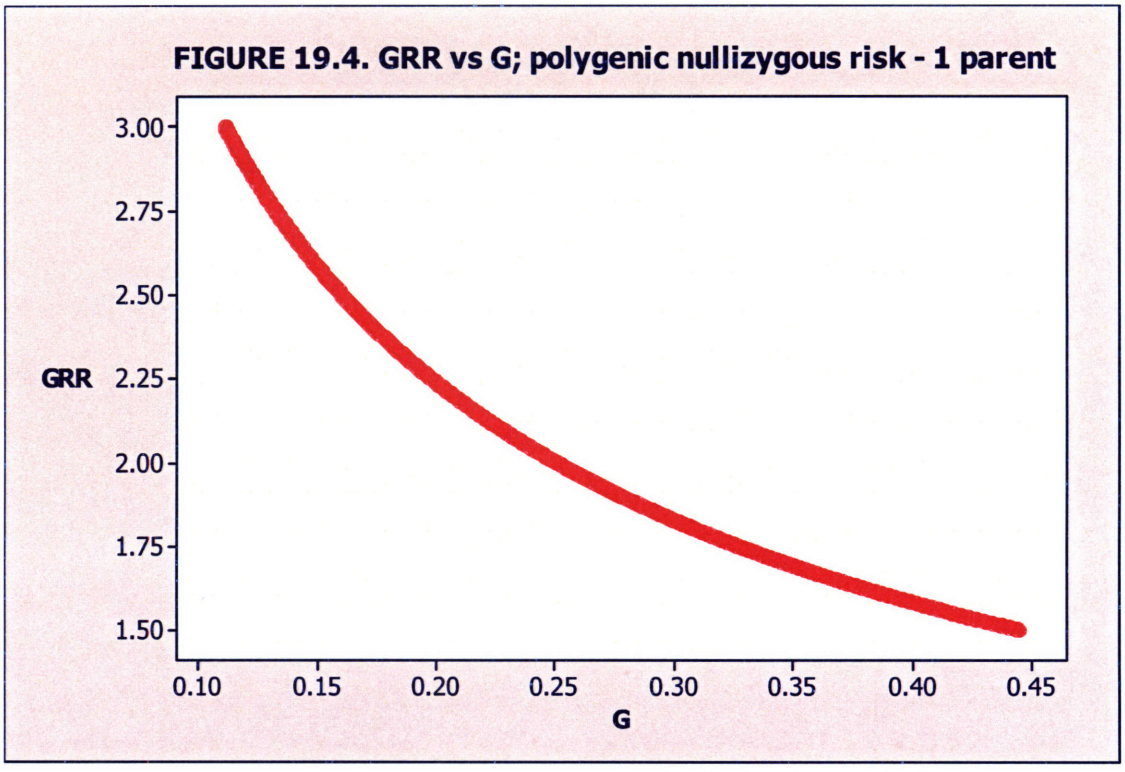


FIGURE 19.4. GRR vs G; polygenic nullizygous risk - 1 parent
GRR
3.00
2.75
2.50
2.25
2.00
1.75
1.50
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
G

TABLE 1: **Monogenic heterozygous risk** ($G_0 \leq 0.5$)

G_0	q	1-q	GRR
0.01	0.00503	0.99497	50
0.02	0.01010	0.98990	25
0.03	0.01523	0.98477	16.6667
0.04	0.02042	0.97958	12.5
0.05	0.02566	0.97434	10
0.06	0.03096	0.96904	8.3333
0.07	0.03632	0.96368	7.1429
0.08	0.04174	0.95826	6.25
0.09	0.04723	0.95277	5.5556
0.1	0.05279	0.94721	5
0.11	0.05841	0.94159	4.5455
0.12	0.06411	0.93589	4.1667
0.13	0.06988	0.93012	3.8462
0.14	0.07574	0.92426	3.5714
0.15	0.08167	0.91833	3.3333
0.16	0.08769	0.91231	3.125
0.1667	0.09175	0.90825	3
0.17	0.09380	0.90620	2.9412
0.18	0.1	0.9	2.7778
0.19	0.10630	0.89370	2.6316
0.2	0.11270	0.88730	2.5
0.21	0.11921	0.88079	2.3810
0.22	0.12583	0.87417	2.2727
0.23	0.13258	0.86742	2.1739
0.24	0.13944	0.86056	2.0833
0.25	0.14645	0.85355	2
0.26	0.15359	0.84641	1.9231
0.27	0.16088	0.83912	1.8519
0.28	0.16834	0.83166	1.7857
0.29	0.17596	0.82404	1.7241
0.3	0.18377	0.81623	1.6667
0.31	0.19178	0.80822	1.6129
0.32	0.2	0.8	1.5625

TABLE 1 cont'd: Monogenic heterozygous risk ($G_0 \leq 0.5$)

G_0	q	1-q	GRR
0.33	0.20845	0.79155	1.5152
0.3333	0.21132	0.78868	1.5
0.34	0.21716	0.78284	1.4706
0.35	0.22614	0.77386	1.4286
0.36	0.23542	0.76458	1.3889
0.37	0.24505	0.75495	1.3514
0.38	0.25505	0.74495	1.3158
0.39	0.26548	0.73452	1.2821
0.4	0.27639	0.72361	1.25
0.41	0.28787	0.71213	1.2195
0.42	0.3	0.7	1.1905
0.43	0.31292	0.68708	1.1628
0.44	0.32679	0.67321	1.1364
0.45	0.34189	0.65811	1.1111
0.46	0.35858	0.64142	1.0870
0.47	0.37753	0.62247	1.0638
0.48	0.4	0.6	1.0417
0.49	0.42929	0.57071	1.0204
0.5	0.5	0.5	1

TABLE 2: Monogenic nullizygous risk

G_0	q	GRR (1 parent)
0.01	0.1	10
0.02	0.14142	7.0711
0.03	0.17321	5.7735
0.04	0.2	5
0.05	0.22361	4.4721
0.06	0.24495	4.0825
0.07	0.26458	3.7796
0.08	0.28284	3.5355
0.09	0.3	3.3333
0.1	0.31623	3.1623
0.11	0.33166	3.0151
0.1111	0.33333	3
0.12	0.34641	2.8868
0.13	0.36056	2.7735
0.14	0.37417	2.6726
0.15	0.38730	2.5820
0.16	0.4	2.5
0.17	0.41231	2.4254
0.18	0.42426	2.3570
0.19	0.43589	2.2942
0.2	0.44721	2.2361
0.21	0.45826	2.1822
0.22	0.46904	2.1320
0.23	0.47958	2.0851
0.24	0.48990	2.0412
0.25	0.5	2
0.26	0.50990	1.9612
0.27	0.51962	1.9245
0.28	0.52915	1.8898
0.29	0.53852	1.8570
0.3	0.54772	1.8257
0.31	0.55678	1.7961
0.32	0.56569	1.7678

TABLE 2 cont'd: Monogenic nullizygous risk

G_0	q	GRR (1 parent)
0.33	0.57446	1.7408
0.34	0.58310	1.7150
0.35	0.59161	1.6903
0.36	0.60000	1.6667
0.37	0.60828	1.6440
0.38	0.61644	1.6222
0.39	0.62450	1.6013
0.4	0.63246	1.5811
0.41	0.64031	1.5617
0.42	0.64807	1.5430
0.43	0.65574	1.5250
0.44	0.66332	1.5076
0.4444	0.66667	1.5
0.45	0.67082	1.4907
0.46	0.67823	1.4744
0.47	0.68557	1.4586
0.48	0.69282	1.4434
0.49	0.70000	1.4286
0.5	0.70711	1.4142
0.51	0.71414	1.4003
0.52	0.72111	1.3868
0.53	0.72801	1.3736
0.54	0.73485	1.3608
0.55	0.74162	1.3484
0.56	0.74833	1.3363
0.57	0.75498	1.3245
0.58	0.76158	1.3131
0.59	0.76811	1.3019
0.6	0.77460	1.2910
0.61	0.78102	1.2804
0.62	0.78740	1.2700
0.63	0.79373	1.2599
0.64	0.8	1.25
0.65	0.80623	1.2403
0.66	0.81240	1.2309

TABLE 2 cont'd: Monogenic nullizygous risk

G_0	q	GRR (1 parent)
0.67	0.81854	1.2217
0.68	0.82462	1.2127
0.69	0.83066	1.2039
0.7	0.83666	1.1952
0.71	0.84261	1.1868
0.72	0.84853	1.1785
0.73	0.85440	1.1704
0.74	0.86023	1.1625
0.75	0.86603	1.1547
0.76	0.87178	1.1471
0.77	0.87750	1.1396
0.78	0.88318	1.1323
0.79	0.88882	1.1251
0.8	0.89443	1.1180
0.81	0.9	1.1111
0.82	0.90554	1.1043
0.83	0.91104	1.0976
0.84	0.91652	1.0911
0.85	0.92195	1.0847
0.86	0.92736	1.0783
0.87	0.93274	1.0721
0.88	0.93808	1.0660
0.89	0.94340	1.0600
0.9	0.94868	1.0541
0.91	0.95394	1.0483
0.92	0.95917	1.0426
0.93	0.96437	1.0370
0.94	0.96954	1.0314
0.95	0.97468	1.0260
0.96	0.97980	1.0206
0.97	0.98489	1.0153
0.98	0.98995	1.0102
0.99	0.99499	1.0050
1	1	1

TABLE 3: Multigenic heterozygous risk, m = 2 ($G_0 \leq 0.75$)

G_0	q	1-q	GRR
0.01	0.00251	0.99749	50.2506
0.02	0.00505	0.99495	25.2513
0.03	0.00762	0.99238	16.9186
0.04	0.01021	0.98979	12.7526
0.05	0.01282	0.98718	10.2532
0.06	0.01547	0.98453	8.5872
0.07	0.01815	0.98185	7.3974
0.08	0.02085	0.97915	6.5052
0.09	0.02359	0.97641	5.8114
0.1	0.02635	0.97365	5.2566
0.11	0.02915	0.97085	4.8027
0.12	0.03198	0.96802	4.4247
0.13	0.03485	0.96515	4.1049
0.14	0.03774	0.96226	3.8309
0.15	0.04068	0.95932	3.5935
0.16	0.04365	0.95635	3.3859
0.17	0.04666	0.95334	3.2028
0.18	0.04970	0.95030	3.0402
0.1827	0.05052	0.94948	3.0000
0.19	0.05279	0.94721	2.8947
0.2	0.05591	0.94409	2.7639
0.21	0.05908	0.94092	2.6457
0.22	0.06229	0.93771	2.5382
0.2238	0.06352	0.93648	2.5000
0.23	0.06555	0.93445	2.4402
0.24	0.06885	0.93115	2.3505
0.25	0.07220	0.92780	2.2679
0.26	0.07560	0.92440	2.1919
0.27	0.07905	0.92095	2.1215
0.28	0.08255	0.91745	2.0562
0.2892	0.08583	0.91417	2.0000
0.29	0.08611	0.91389	1.9955
0.3	0.08972	0.91028	1.9389
0.31	0.09339	0.90661	1.8860
0.32	0.09712	0.90288	1.8365

TABLE 3 cont'd: Multigenic heterozygous risk, m = 2 ($G_0 \leq 0.75$)

G_0	q	1-q	GRR
0.33	0.10092	0.89908	1.7901
0.34	0.10478	0.89522	1.7465
0.35	0.10870	0.89130	1.7054
0.36	0.11270	0.88730	1.6667
0.37	0.11677	0.88323	1.6301
0.38	0.12092	0.87908	1.5955
0.39	0.12515	0.87485	1.5628
0.4	0.12946	0.87054	1.5318
0.41	0.13386	0.86614	1.5023
0.4108	0.13422	0.86578	1.5000
0.42	0.13835	0.86165	1.4743
0.43	0.14294	0.85706	1.4477
0.44	0.14763	0.85237	1.4224
0.45	0.15242	0.84758	1.3982
0.46	0.15733	0.84267	1.3752
0.47	0.16235	0.83765	1.3532
0.48	0.16750	0.83250	1.3322
0.49	0.17278	0.82722	1.3121
0.5	0.17820	0.82180	1.2929
0.51	0.18377	0.81623	1.2745
0.52	0.18950	0.81050	1.2569
0.53	0.19540	0.80460	1.2400
0.54	0.20148	0.79852	1.2239
0.55	0.20775	0.79225	1.2083
0.56	0.21423	0.78577	1.1935
0.57	0.22094	0.77906	1.1792
0.58	0.22790	0.77210	1.1655
0.59	0.23513	0.76487	1.1523
0.6	0.24265	0.75735	1.1396
0.61	0.25050	0.74950	1.1275
0.62	0.25871	0.74129	1.1158
0.63	0.26732	0.73268	1.1045
0.64	0.27639	0.72361	1.0938
0.65	0.28598	0.71402	1.0834
0.66	0.29617	0.70383	1.0734

TABLE 3 cont'd: Multigenic heterozygous risk, m = 2 ($G_0 \leq 0.75$)

G_0	q	1-q	GRR
0.67	0.30705	0.69295	1.0638
0.68	0.31877	0.68123	1.0546
0.69	0.33151	0.66849	1.0458
0.7	0.34553	0.65447	1.0373
0.71	0.36123	0.63877	1.0292
0.72	0.37927	0.62073	1.0214
0.73	0.40097	0.59903	1.0140
0.74	0.42964	0.57036	1.0068
0.75	0.5	0.5	1

TABLE 4: **Multigenic heterozygous risk, m = 3** ($G_0 \leq 0.875$)

G_0	q	1-q	GRR
0.01	0.00168	0.99832	50.3339
0.02	0.00337	0.99663	25.3345
0.03	0.00508	0.99492	17.0017
0.04	0.00680	0.99320	12.8356
0.05	0.00855	0.99145	10.3362
0.06	0.01031	0.98969	8.6701
0.07	0.01210	0.98790	7.4802
0.08	0.01390	0.98610	6.5879
0.09	0.01572	0.98428	5.8941
0.1	0.01756	0.98244	5.3392
0.11	0.01943	0.98057	4.8852
0.12	0.02131	0.97869	4.5070
0.13	0.02322	0.97678	4.1872
0.14	0.02515	0.97485	3.9131
0.15	0.02710	0.97290	3.6756
0.16	0.02908	0.97092	3.4679
0.17	0.03108	0.96892	3.2847
0.18	0.03310	0.96690	3.1220
0.1883	0.03480	0.96520	3.0000
0.19	0.03515	0.96485	2.9765
0.2	0.03723	0.96277	2.8456
0.21	0.03933	0.96067	2.7272
0.22	0.04146	0.95854	2.6197
0.23	0.04362	0.95638	2.5215
0.2323	0.04412	0.95588	2.5000
0.24	0.04581	0.95419	2.4317
0.25	0.04803	0.95197	2.3490
0.26	0.05028	0.94972	2.2728
0.27	0.05256	0.94744	2.2023
0.28	0.05487	0.94513	2.1369
0.29	0.05722	0.94278	2.0761
0.3	0.05960	0.94040	2.0194
0.3036	0.06046	0.93954	2.0000
0.31	0.06202	0.93798	1.9664
0.32	0.06447	0.93553	1.9167

TABLE 4 cont'd: Multigenic heterozygous risk, m = 3 ($G_0 \leq 0.875$)

G_0	q	1-q	GRR
0.33	0.06697	0.93303	1.8702
0.34	0.06950	0.93050	1.8264
0.35	0.07208	0.92792	1.7852
0.36	0.07469	0.92531	1.7463
0.37	0.07735	0.92265	1.7096
0.38	0.08006	0.91994	1.6749
0.39	0.08281	0.91719	1.6420
0.4	0.08561	0.91439	1.6108
0.41	0.08847	0.91153	1.5812
0.42	0.09137	0.90863	1.5530
0.43	0.09433	0.90567	1.5262
0.44	0.09735	0.90265	1.5007
0.4403	0.09743	0.90257	1.5000
0.45	0.10042	0.89958	1.4763
0.46	0.10356	0.89644	1.4531
0.47	0.10677	0.89323	1.4309
0.48	0.11004	0.88996	1.4097
0.49	0.11338	0.88662	1.3895
0.5	0.11679	0.88321	1.3700
0.51	0.12028	0.87972	1.3514
0.52	0.12385	0.87615	1.3336
0.53	0.12751	0.87249	1.3165
0.54	0.13126	0.86874	1.3001
0.55	0.13510	0.86490	1.2843
0.56	0.13904	0.86096	1.2692
0.57	0.14308	0.85692	1.2546
0.58	0.14723	0.85277	1.2407
0.59	0.15151	0.84849	1.2272
0.6	0.15590	0.84410	1.2143
0.61	0.16043	0.83957	1.2018
0.62	0.16510	0.83490	1.1898
0.63	0.16992	0.83008	1.1783
0.64	0.17490	0.82510	1.1671
0.65	0.18005	0.81995	1.1564
0.66	0.18539	0.81461	1.1461

TABLE 4 cont'd: Multigenic heterozygous risk, m = 3 ($G_0 \leq 0.875$)

G_0	q	1-q	GRR
0.67	0.19094	0.80906	1.1362
0.68	0.19669	0.80331	1.1266
0.69	0.20269	0.79731	1.1174
0.7	0.20894	0.79106	1.1085
0.71	0.21547	0.78453	1.0999
0.72	0.22232	0.77768	1.0917
0.73	0.22951	0.77049	1.0837
0.74	0.23708	0.76292	1.0761
0.75	0.24509	0.75491	1.0688
0.76	0.25358	0.74642	1.0617
0.77	0.26263	0.73737	1.0549
0.78	0.27231	0.72769	1.0484
0.79	0.28275	0.71725	1.0422
0.8	0.29408	0.70592	1.0363
0.81	0.30649	0.69351	1.0306
0.82	0.32025	0.67975	1.0251
0.83	0.33574	0.66426	1.0200
0.84	0.35357	0.64643	1.0150
0.85	0.37484	0.62516	1.0104
0.86	0.40189	0.59811	1.0060
0.87	0.44264	0.55736	1.0019
0.875	0.5	0.5	1

TABLE 5: Multigenic heterozygous risk, m = 4 ($G_0 \leq 0.9375$)

G_0	q	1-q	GRR
0.01	0.00126	0.99874	50.3755
0.02	0.00253	0.99747	25.3759
0.03	0.00381	0.99619	17.0431
0.04	0.00510	0.99490	12.8769
0.05	0.00641	0.99359	10.3774
0.06	0.00773	0.99227	8.7112
0.07	0.00907	0.99093	7.5212
0.08	0.01042	0.98958	6.6289
0.09	0.01179	0.98821	5.9349
0.1	0.01317	0.98683	5.3799
0.11	0.01457	0.98543	4.9259
0.12	0.01598	0.98402	4.5476
0.13	0.01741	0.98259	4.2276
0.14	0.01886	0.98114	3.9534
0.15	0.02032	0.97968	3.7158
0.16	0.02180	0.97820	3.5081
0.17	0.02330	0.97670	3.3248
0.18	0.02482	0.97518	3.1619
0.19	0.02635	0.97365	3.0163
0.1912	0.02654	0.97346	3.0000
0.2	0.02791	0.97209	2.8853
0.21	0.02948	0.97052	2.7668
0.22	0.03108	0.96892	2.6591
0.23	0.03270	0.96730	2.5609
0.2367	0.03379	0.96621	2.5000
0.24	0.03433	0.96567	2.4709
0.25	0.03599	0.96401	2.3881
0.26	0.03768	0.96232	2.3118
0.27	0.03938	0.96062	2.2412
0.28	0.04111	0.95889	2.1757
0.29	0.04287	0.95713	2.1147
0.3	0.04465	0.95535	2.0579
0.31	0.04645	0.95355	2.0047
0.3109	0.04662	0.95338	2.0000
0.32	0.04829	0.95171	1.9550

TABLE 5 cont'd: Multigenic heterozygous risk, m = 4 ($G_0 \leq 0.9375$)

G_0	q	1-q	GRR
0.33	0.05015	0.94985	1.9083
0.34	0.05204	0.94796	1.8643
0.35	0.05396	0.94604	1.8230
0.36	0.05591	0.94409	1.7840
0.37	0.05790	0.94210	1.7471
0.38	0.05991	0.94009	1.7122
0.39	0.06196	0.93804	1.6792
0.4	0.06405	0.93595	1.6478
0.41	0.06617	0.93383	1.6181
0.42	0.06833	0.93167	1.5897
0.43	0.07052	0.92948	1.5628
0.44	0.07276	0.92724	1.5371
0.45	0.07504	0.92496	1.5126
0.4553	0.07628	0.92372	1.5000
0.46	0.07737	0.92263	1.4892
0.47	0.07974	0.92026	1.4668
0.48	0.08216	0.91784	1.4455
0.49	0.08463	0.91537	1.4250
0.5	0.08715	0.91285	1.4054
0.51	0.08972	0.91028	1.3866
0.52	0.09235	0.90765	1.3686
0.53	0.09504	0.90496	1.3513
0.54	0.09779	0.90221	1.3347
0.55	0.10060	0.89940	1.3187
0.56	0.10349	0.89651	1.3034
0.57	0.10644	0.89356	1.2886
0.58	0.10947	0.89053	1.2744
0.59	0.11258	0.88742	1.2607
0.6	0.11577	0.88423	1.2475
0.61	0.11905	0.88095	1.2348
0.62	0.12242	0.87758	1.2226
0.63	0.12589	0.87411	1.2108
0.64	0.12946	0.87054	1.1994
0.65	0.13315	0.86685	1.1884
0.66	0.13695	0.86305	1.1778

TABLE 5 cont'd: Multigenic heterozygous risk, m = 4 ($G_0 \leq 0.9375$)

G_0	q	1-q	GRR
0.67	0.14088	0.85912	1.1676
0.68	0.14495	0.85505	1.1577
0.69	0.14916	0.85084	1.1482
0.7	0.15353	0.84647	1.1390
0.71	0.15807	0.84193	1.1302
0.72	0.16279	0.83721	1.1216
0.73	0.16770	0.83230	1.1133
0.74	0.17283	0.82717	1.1053
0.75	0.17820	0.82180	1.0976
0.76	0.18383	0.81617	1.0902
0.77	0.18974	0.81026	1.0830
0.78	0.19597	0.80403	1.0761
0.79	0.20255	0.79745	1.0695
0.8	0.20953	0.79047	1.0631
0.81	0.21696	0.78304	1.0569
0.82	0.22490	0.77510	1.0510
0.83	0.23343	0.76657	1.0453
0.84	0.24265	0.75735	1.0399
0.85	0.25268	0.74732	1.0347
0.86	0.26368	0.73632	1.0297
0.87	0.27588	0.72412	1.0250
0.88	0.28956	0.71044	1.0205
0.89	0.30519	0.69481	1.0163
0.9	0.32345	0.67655	1.0123
0.91	0.34553	0.65447	1.0086
0.92	0.37385	0.62615	1.0052
0.93	0.41524	0.58476	1.0021
0.9375	0.5	0.5	1

TABLE 6: **Multigenic heterozygous risk, m = 5** ($G_0 \leq 0.96875$)

G_0	q	1-q	GRR
0.01	0.00101	0.99899	50.4004
0.02	0.00202	0.99798	25.4008
0.03	0.00305	0.99695	17.0679
0.04	0.00408	0.99592	12.9016
0.05	0.00513	0.99487	10.4020
0.06	0.00619	0.99381	8.7358
0.07	0.00726	0.99274	7.5457
0.08	0.00834	0.99166	6.6533
0.09	0.00943	0.99057	5.9593
0.1	0.01054	0.98946	5.4042
0.11	0.01165	0.98835	4.9501
0.12	0.01278	0.98722	4.5717
0.13	0.01393	0.98607	4.2516
0.14	0.01508	0.98492	3.9774
0.15	0.01625	0.98375	3.7397
0.16	0.01744	0.98256	3.5318
0.17	0.01864	0.98136	3.3485
0.18	0.01985	0.98015	3.1856
0.19	0.02108	0.97892	3.0398
0.1929	0.02144	0.97856	3.0000
0.2	0.02232	0.97768	2.9087
0.21	0.02358	0.97642	2.7902
0.22	0.02486	0.97514	2.6824
0.23	0.02615	0.97385	2.5841
0.2393	0.02737	0.97263	2.5000
0.24	0.02746	0.97254	2.4940
0.25	0.02878	0.97122	2.4112
0.26	0.03013	0.96987	2.3347
0.27	0.03149	0.96851	2.2640
0.28	0.03288	0.96712	2.1984
0.29	0.03428	0.96572	2.1373
0.3	0.03570	0.96430	2.0804
0.31	0.03714	0.96286	2.0272
0.3154	0.03793	0.96207	2.0000
0.32	0.03861	0.96139	1.9773

TABLE 6 cont'd: Multigenic heterozygous risk, m = 5 ($G_0 \leq 0.96875$)

G_0	q	1-q	GRR
0.33	0.04009	0.95991	1.9305
0.34	0.04160	0.95840	1.8865
0.35	0.04314	0.95686	1.8450
0.36	0.04469	0.95531	1.8059
0.37	0.04627	0.95373	1.7689
0.38	0.04788	0.95212	1.7339
0.39	0.04952	0.95048	1.7008
0.4	0.05118	0.94882	1.6693
0.41	0.05287	0.94713	1.6394
0.42	0.05459	0.94541	1.6110
0.43	0.05634	0.94366	1.5839
0.44	0.05812	0.94188	1.5581
0.45	0.05994	0.94006	1.5335
0.46	0.06179	0.93821	1.5100
0.4644	0.06261	0.93739	1.5000
0.47	0.06368	0.93632	1.4875
0.48	0.06560	0.93440	1.4660
0.49	0.06756	0.93244	1.4454
0.5	0.06956	0.93044	1.4257
0.51	0.07161	0.92839	1.4067
0.52	0.07369	0.92631	1.3886
0.53	0.07583	0.92417	1.3711
0.54	0.07801	0.92199	1.3544
0.55	0.08024	0.91976	1.3383
0.56	0.08252	0.91748	1.3228
0.57	0.08486	0.91514	1.3078
0.58	0.08725	0.91275	1.2935
0.59	0.08971	0.91029	1.2796
0.6	0.09223	0.90777	1.2663
0.61	0.09482	0.90518	1.2534
0.62	0.09747	0.90253	1.2410
0.63	0.10020	0.89980	1.2290
0.64	0.10302	0.89698	1.2175
0.65	0.10591	0.89409	1.2063
0.66	0.10889	0.89111	1.1955

TABLE 6 cont'd: Multigenic heterozygous risk, m = 5 ($G_0 \leq 0.96875$)

G_0	q	1-q	GRR
0.67	0.11197	0.88803	1.1851
0.68	0.11515	0.88485	1.1751
0.69	0.11844	0.88156	1.1653
0.7	0.12184	0.87816	1.1559
0.71	0.12537	0.87463	1.1469
0.72	0.12903	0.87097	1.1381
0.73	0.13284	0.86716	1.1296
0.74	0.13680	0.86320	1.1214
0.75	0.14093	0.85907	1.1134
0.76	0.14525	0.85475	1.1057
0.77	0.14977	0.85023	1.0983
0.78	0.15451	0.84549	1.0911
0.79	0.15950	0.84050	1.0842
0.8	0.16475	0.83525	1.0775
0.81	0.17032	0.82968	1.0711
0.82	0.17622	0.82378	1.0649
0.83	0.18251	0.81749	1.0589
0.84	0.18924	0.81076	1.0531
0.85	0.19647	0.80353	1.0475
0.86	0.20430	0.79570	1.0422
0.87	0.21282	0.78718	1.0371
0.88	0.22216	0.77784	1.0322
0.89	0.23251	0.76749	1.0275
0.9	0.24411	0.75589	1.0231
0.91	0.25731	0.74269	1.0189
0.92	0.27260	0.72740	1.0149
0.93	0.29082	0.70918	1.0112
0.94	0.31335	0.68665	1.0078
0.95	0.34303	0.65697	1.0047
0.96	0.38752	0.61248	1.0020
0.96875	0.5	0.5	1

TABLE 7: Multigenic nullizygous risk, m = 2

G_0	q	GRR
0.01	0.07080	7.5457
0.02	0.10025	5.4648
0.03	0.12294	4.5399
0.04	0.14214	3.9868
0.05	0.15912	3.6083
0.06	0.17454	3.3281
0.07	0.18877	3.1097
0.0760	0.19683	3.0000
0.08	0.20207	2.9332
0.09	0.21462	2.7866
0.1	0.22653	2.6622
0.11	0.23791	2.5550
0.1157	0.24420	2.5000
0.12	0.24883	2.4612
0.13	0.25935	2.3782
0.14	0.26951	2.3041
0.15	0.27937	2.2374
0.16	0.28894	2.1769
0.17	0.29826	2.1217
0.18	0.30735	2.0710
0.19	0.31623	2.0242
0.1955	0.32104	2.0000
0.2	0.32492	1.9809
0.21	0.33344	1.9407
0.22	0.34180	1.9031
0.23	0.35001	1.8680
0.24	0.35808	1.8349
0.25	0.36603	1.8038
0.26	0.37385	1.7745
0.27	0.38158	1.7467
0.28	0.38919	1.7204
0.29	0.39672	1.6954
0.3	0.40415	1.6716
0.31	0.41151	1.6489
0.32	0.41878	1.6272

TABLE 7 cont'd: Multigenic nullizygous risk, m = 2

G_0	q	GRR
0.33	0.42599	1.6065
0.34	0.43312	1.5867
0.35	0.44020	1.5676
0.36	0.44721	1.5494
0.37	0.45417	1.5318
0.38	0.46108	1.5149
0.3891	0.46735	1.5000
0.39	0.46795	1.4986
0.4	0.47477	1.4829
0.41	0.48154	1.4677
0.42	0.48829	1.4531
0.43	0.49499	1.4389
0.44	0.50167	1.4252
0.45	0.50831	1.4119
0.46	0.51493	1.3990
0.47	0.52153	1.3865
0.48	0.52810	1.3744
0.49	0.53466	1.3626
0.5	0.54120	1.3512
0.51	0.54772	1.3400
0.52	0.55424	1.3292
0.53	0.56074	1.3186
0.54	0.56725	1.3083
0.55	0.57374	1.2983
0.56	0.58024	1.2885
0.57	0.58673	1.2790
0.58	0.59323	1.2696
0.59	0.59974	1.2605
0.6	0.60625	1.2516
0.61	0.61278	1.2429
0.62	0.61932	1.2344
0.63	0.62588	1.2261
0.64	0.63246	1.2179
0.65	0.63906	1.2099
0.66	0.64568	1.2021

TABLE 7 cont'd: Multigenic nullizygous risk, m = 2

G_0	q	GRR
0.67	0.65234	1.1945
0.68	0.65903	1.1869
0.69	0.66575	1.1796
0.7	0.67252	1.1723
0.71	0.67933	1.1652
0.72	0.68618	1.1583
0.73	0.69310	1.1514
0.74	0.70007	1.1447
0.75	0.70711	1.1381
0.76	0.71421	1.1316
0.77	0.72140	1.1252
0.78	0.72867	1.1189
0.79	0.73603	1.1127
0.8	0.74350	1.1066
0.81	0.75107	1.1006
0.82	0.75877	1.0947
0.83	0.76661	1.0889
0.84	0.77460	1.0831
0.85	0.78275	1.0775
0.86	0.79110	1.0719
0.87	0.79965	1.0664
0.88	0.80845	1.0610
0.89	0.81752	1.0556
0.9	0.82691	1.0503
0.91	0.83666	1.0451
0.92	0.84685	1.0399
0.93	0.85757	1.0347
0.94	0.86894	1.0297
0.95	0.88113	1.0247
0.96	0.89443	1.0197
0.97	0.90928	1.0147
0.98	0.92660	1.0098
0.99	0.94868	1.0049
1	1	1

TABLE 8: Multigenic nullizygous risk, m = 3

G_0	q	GRR
0.01	0.05783	6.4123
0.02	0.08192	4.7103
0.03	0.10051	3.9529
0.04	0.11625	3.4995
0.05	0.13020	3.1888
0.0580	0.14042	3.0000
0.06	0.14288	2.9586
0.07	0.15460	2.7789
0.08	0.16556	2.6335
0.09	0.17592	2.5126
0.0912	0.17708	2.5000
0.1	0.18577	2.4100
0.11	0.19519	2.3214
0.12	0.20425	2.2438
0.13	0.21298	2.1751
0.14	0.22143	2.1136
0.15	0.22963	2.0582
0.16	0.23762	2.0080
0.1617	0.23893	2.0000
0.17	0.24540	1.9620
0.18	0.25300	1.9198
0.19	0.26044	1.8809
0.2	0.26774	1.8448
0.21	0.27489	1.8111
0.22	0.28193	1.7797
0.23	0.28885	1.7503
0.24	0.29567	1.7226
0.25	0.30239	1.6965
0.26	0.30902	1.6719
0.27	0.31558	1.6486
0.28	0.32205	1.6264
0.29	0.32846	1.6053
0.3	0.33481	1.5853
0.31	0.34109	1.5661
0.32	0.34732	1.5478

TABLE 8 cont'd: Multigenic nullizygous risk, m = 3

G_0	q	GRR
0.33	0.35351	1.5303
0.34	0.35964	1.5135
0.3483	0.36471	1.5000
0.35	0.36573	1.4973
0.36	0.37179	1.4818
0.37	0.37781	1.4669
0.38	0.38379	1.4525
0.39	0.38975	1.4386
0.4	0.39569	1.4253
0.41	0.40160	1.4123
0.42	0.40749	1.3998
0.43	0.41336	1.3877
0.44	0.41922	1.3759
0.45	0.42506	1.3646
0.46	0.43090	1.3535
0.47	0.43673	1.3428
0.48	0.44255	1.3324
0.49	0.44838	1.3222
0.5	0.45420	1.3123
0.51	0.46003	1.3027
0.52	0.46586	1.2934
0.53	0.47170	1.2842
0.54	0.47755	1.2753
0.55	0.48342	1.2666
0.56	0.48929	1.2581
0.57	0.49519	1.2498
0.58	0.50111	1.2417
0.59	0.50705	1.2338
0.6	0.51302	1.2260
0.61	0.51902	1.2185
0.62	0.52506	1.2110
0.63	0.53113	1.2037
0.64	0.53723	1.1966
0.65	0.54339	1.1896
0.66	0.54959	1.1827

TABLE 8 cont'd: Multigenic nullizygous risk, m = 3

G_0	q	GRR
0.67	0.55584	1.1760
0.68	0.56215	1.1693
0.69	0.56852	1.1628
0.7	0.57495	1.1565
0.71	0.58145	1.1502
0.72	0.58804	1.1440
0.73	0.59470	1.1379
0.74	0.60146	1.1320
0.75	0.60831	1.1261
0.76	0.61527	1.1203
0.77	0.62234	1.1146
0.78	0.62954	1.1090
0.79	0.63687	1.1034
0.8	0.64436	1.0980
0.81	0.65200	1.0926
0.82	0.65983	1.0873
0.83	0.66786	1.0820
0.84	0.67610	1.0768
0.85	0.68460	1.0717
0.86	0.69336	1.0667
0.87	0.70244	1.0617
0.88	0.71187	1.0567
0.89	0.72170	1.0518
0.9	0.73201	1.0470
0.91	0.74287	1.0422
0.92	0.75440	1.0374
0.93	0.76673	1.0327
0.94	0.78007	1.0280
0.95	0.79473	1.0233
0.96	0.81117	1.0187
0.97	0.83023	1.0140
0.98	0.85356	1.0094
0.99	0.88575	1.0047
1	1	1

TABLE 9: Multigenic nullizygous risk, m = 4

G_0	q	GRR
0.01	0.05009	5.7227
0.02	0.07098	4.2474
0.03	0.08710	3.5905
0.04	0.10077	3.1970
0.0470	0.10933	3.0000
0.05	0.11288	2.9272
0.06	0.12389	2.7270
0.07	0.13409	2.5708
0.0754	0.13927	2.5000
0.08	0.14363	2.4443
0.09	0.15265	2.3391
0.1	0.16123	2.2496
0.11	0.16945	2.1723
0.12	0.17735	2.1047
0.13	0.18498	2.0447
0.1383	0.19109	2.0000
0.14	0.19236	1.9910
0.15	0.19954	1.9426
0.16	0.20652	1.8987
0.17	0.21334	1.8585
0.18	0.22000	1.8215
0.19	0.22653	1.7874
0.2	0.23293	1.7557
0.21	0.23922	1.7262
0.22	0.24541	1.6986
0.23	0.25150	1.6728
0.24	0.25750	1.6485
0.25	0.26343	1.6255
0.26	0.26928	1.6038
0.27	0.27507	1.5833
0.28	0.28079	1.5637
0.29	0.28646	1.5452
0.3	0.29208	1.5275
0.31	0.29765	1.5105
0.3165	0.30123	1.5000
0.32	0.30317	1.4944

TABLE 9 cont'd: Multigenic nullizygous risk, m = 4

G_0	q	GRR
0.33	0.30866	1.4789
0.34	0.31411	1.4640
0.35	0.31953	1.4497
0.36	0.32492	1.4360
0.37	0.33028	1.4227
0.38	0.33562	1.4100
0.39	0.34095	1.3977
0.4	0.34625	1.3858
0.41	0.35154	1.3743
0.42	0.35681	1.3632
0.43	0.36208	1.3524
0.44	0.36734	1.3419
0.45	0.37259	1.3318
0.46	0.37785	1.3219
0.47	0.38310	1.3123
0.48	0.38835	1.3030
0.49	0.39361	1.2940
0.5	0.39888	1.2851
0.51	0.40415	1.2765
0.52	0.40944	1.2682
0.53	0.41474	1.2600
0.54	0.42006	1.2520
0.55	0.42540	1.2442
0.56	0.43076	1.2366
0.57	0.43614	1.2291
0.58	0.44155	1.2218
0.59	0.44700	1.2147
0.6	0.45247	1.2077
0.61	0.45798	1.2008
0.62	0.46353	1.1941
0.63	0.46913	1.1875
0.64	0.47477	1.1811
0.65	0.48046	1.1747
0.66	0.48620	1.1685

TABLE 9 cont'd: Multigenic nullizygous risk, m = 4

G_0	q	GRR
0.67	0.49201	1.1624
0.68	0.49787	1.1564
0.69	0.50381	1.1505
0.7	0.50982	1.1447
0.71	0.51591	1.1390
0.72	0.52209	1.1334
0.73	0.52835	1.1279
0.74	0.53472	1.1224
0.75	0.54120	1.1171
0.76	0.54779	1.1118
0.77	0.55451	1.1065
0.78	0.56137	1.1014
0.79	0.56838	1.0963
0.8	0.57555	1.0913
0.81	0.58291	1.0864
0.82	0.59046	1.0815
0.83	0.59824	1.0767
0.84	0.60625	1.0719
0.85	0.61455	1.0672
0.86	0.62314	1.0625
0.87	0.63209	1.0579
0.88	0.64143	1.0533
0.89	0.65123	1.0487
0.9	0.66156	1.0442
0.91	0.67252	1.0398
0.92	0.68423	1.0353
0.93	0.69687	1.0309
0.94	0.71069	1.0265
0.95	0.72604	1.0221
0.96	0.74350	1.0178
0.97	0.76408	1.0134
0.98	0.78990	1.0090
0.99	0.82691	1.0046
1	1	1

TABLE 10: Multigenic nullizygous risk, m = 5

G_0	q	GRR
0.01	0.04481	5.2460
0.02	0.06350	3.9258
0.03	0.07793	3.3376
0.0395	0.08958	3.0000
0.04	0.09017	2.9851
0.05	0.10103	2.7434
0.06	0.11090	2.5640
0.0643	0.11490	2.5000
0.07	0.12004	2.4239
0.08	0.12860	2.3104
0.09	0.13669	2.2159
0.1	0.14440	2.1356
0.11	0.15178	2.0662
0.12	0.15888	2.0054
0.1209	0.15954	2.0000
0.13	0.16574	1.9515
0.14	0.17238	1.9032
0.15	0.17883	1.8597
0.16	0.18512	1.8201
0.17	0.19126	1.7839
0.18	0.19726	1.7506
0.19	0.20315	1.7198
0.2	0.20892	1.6913
0.21	0.21459	1.6647
0.22	0.22018	1.6398
0.23	0.22568	1.6164
0.24	0.23110	1.5944
0.25	0.23646	1.5737
0.26	0.24175	1.5541
0.27	0.24699	1.5355
0.28	0.25217	1.5178
0.29	0.25730	1.5010
0.2906	0.25762	1.5000
0.3	0.26239	1.4850
0.31	0.26744	1.4697
0.32	0.27246	1.4550

TABLE 10 cont'd: Multigenic nullizygous risk, m = 5

G_0	q	GRR
0.33	0.27744	1.4409
0.34	0.28239	1.4275
0.35	0.28731	1.4145
0.36	0.29222	1.4020
0.37	0.29710	1.3900
0.38	0.30196	1.3784
0.39	0.30681	1.3672
0.4	0.31164	1.3564
0.41	0.31646	1.3459
0.42	0.32128	1.3358
0.43	0.32609	1.3260
0.44	0.33090	1.3164
0.45	0.33570	1.3072
0.46	0.34051	1.2982
0.47	0.34532	1.2894
0.48	0.35013	1.2809
0.49	0.35496	1.2727
0.5	0.35979	1.2646
0.51	0.36464	1.2567
0.52	0.36950	1.2491
0.53	0.37437	1.2416
0.54	0.37927	1.2342
0.55	0.38419	1.2271
0.56	0.38913	1.2201
0.57	0.39410	1.2133
0.58	0.39910	1.2066
0.59	0.40413	1.2000
0.6	0.40920	1.1936
0.61	0.41431	1.1873
0.62	0.41946	1.1811
0.63	0.42465	1.1751
0.64	0.42989	1.1691
0.65	0.43519	1.1633
0.66	0.44054	1.1575

TABLE 10 cont'd: Multigenic nullizygous risk, m = 5

G_0	q	GRR
0.67	0.44595	1.1519
0.68	0.45143	1.1464
0.69	0.45697	1.1409
0.7	0.46260	1.1356
0.71	0.46830	1.1303
0.72	0.47410	1.1251
0.73	0.47999	1.1200
0.74	0.48598	1.1149
0.75	0.49208	1.1099
0.76	0.49830	1.1050
0.77	0.50465	1.1002
0.78	0.51115	1.0954
0.79	0.51780	1.0907
0.8	0.52461	1.0860
0.81	0.53162	1.0814
0.82	0.53883	1.0769
0.83	0.54626	1.0724
0.84	0.55395	1.0679
0.85	0.56191	1.0635
0.86	0.57019	1.0591
0.87	0.57884	1.0548
0.88	0.58789	1.0505
0.89	0.59741	1.0462
0.9	0.60749	1.0420
0.91	0.61822	1.0378
0.92	0.62975	1.0336
0.93	0.64225	1.0294
0.94	0.65599	1.0253
0.95	0.67136	1.0211
0.96	0.68898	1.0170
0.97	0.70998	1.0128
0.98	0.73668	1.0087
0.99	0.77582	1.0044
1	1	1

TABLE 11: Polygenic heterozygous risk, n = 2 ($G_0 \leq 0.25$)

G_0	q	1-q	GRR
0.01	0.05279	0.94721	25
0.02	0.07657	0.92343	12.5
0.03	0.09578	0.90422	8.3333
0.04	0.11270	0.88730	6.25
0.05	0.12825	0.87175	5
0.06	0.14289	0.85711	4.1667
0.07	0.15691	0.84309	3.5714
0.08	0.17049	0.82951	3.125
0.0833	0.17494	0.82506	3
0.09	0.18377	0.81623	2.7778
0.1	0.19687	0.80313	2.5
0.11	0.20988	0.79012	2.2727
0.12	0.22288	0.77712	2.0833
0.125	0.22940	0.77060	2
0.13	0.23595	0.76405	1.9231
0.14	0.24917	0.75083	1.7857
0.15	0.26262	0.73738	1.6667
0.16	0.27639	0.72361	1.5625
0.1667	0.28582	0.71418	1.5
0.17	0.29061	0.70939	1.4706
0.18	0.30540	0.69460	1.3889
0.19	0.32096	0.67904	1.3158
0.2	0.33754	0.66246	1.25
0.21	0.35553	0.64447	1.1905
0.22	0.37558	0.62442	1.1364
0.23	0.39896	0.60104	1.0870
0.24	0.42893	0.57107	1.0417
0.25	0.5	0.5	1

TABLE 12: Polygenic nullizygous risk, n = 2-5

G_0	p (n = 2)	p (n = 3)	p (n = 4)	p (n = 5)	GRR (1 parent)
0.01	0.68377	0.53584	0.43766	0.36904	10
0.02	0.62394	0.47900	0.38676	0.32376	7.0711
0.03	0.58382	0.44257	0.35488	0.29577	5.7735
0.04	0.55279	0.41520	0.33126	0.27522	5
0.05	0.52713	0.39304	0.31234	0.25887	4.4721
0.06	0.50508	0.37431	0.29649	0.24523	4.0825
0.07	0.48563	0.35803	0.28280	0.23350	3.7796
0.08	0.46817	0.34358	0.27073	0.22320	3.5355
0.09	0.45228	0.33057	0.25992	0.21400	3.3333
0.1	0.43766	0.31871	0.25011	0.20567	3.1623
0.11	0.42410	0.30780	0.24112	0.19806	3.0151
0.1111	0.42265	0.30664	0.24016	0.19726	3
0.12	0.41143	0.29769	0.23282	0.19106	2.8868
0.13	0.39954	0.28826	0.22510	0.18456	2.7735
0.14	0.38831	0.27941	0.21789	0.17849	2.6726
0.15	0.37767	0.27108	0.21112	0.17280	2.5820
0.16	0.36754	0.26319	0.20473	0.16745	2.5
0.17	0.35789	0.25571	0.19868	0.16238	2.4254
0.18	0.34864	0.24859	0.19293	0.15758	2.3570
0.19	0.33978	0.24179	0.18746	0.15302	2.2942
0.2	0.33126	0.23528	0.18223	0.14866	2.2361
0.21	0.32305	0.22903	0.17723	0.14450	2.1822
0.22	0.31513	0.22303	0.17243	0.14051	2.1320
0.23	0.30748	0.21725	0.16782	0.13668	2.0851
0.24	0.30007	0.21168	0.16338	0.13300	2.0412
0.25	0.29289	0.20630	0.15910	0.12945	2
0.26	0.28593	0.20109	0.15497	0.12603	1.9612
0.27	0.27916	0.19605	0.15098	0.12272	1.9245
0.28	0.27257	0.19117	0.14711	0.11953	1.8898
0.29	0.26616	0.18642	0.14336	0.11643	1.8570
0.3	0.25992	0.18181	0.13972	0.11343	1.8257
0.31	0.25383	0.17733	0.13619	0.11052	1.7961
0.32	0.24788	0.17296	0.13275	0.10769	1.7678

TABLE 12 cont'd: Polygenic nullizygous risk, n = 2-5

G_0	p (n = 2)	p (n = 3)	p (n = 4)	p (n = 5)	GRR (1 parent)
0.33	0.24207	0.16871	0.12941	0.10494	1.7408
0.34	0.23639	0.16456	0.12615	0.10227	1.7150
0.35	0.23084	0.16052	0.12298	0.09966	1.6903
0.36	0.22540	0.15657	0.11989	0.09712	1.6667
0.37	0.22008	0.15271	0.11687	0.09464	1.6440
0.38	0.21486	0.14893	0.11392	0.09222	1.6222
0.39	0.20975	0.14524	0.11104	0.08986	1.6013
0.4	0.20473	0.14163	0.10822	0.08756	1.5811
0.41	0.19980	0.13809	0.10546	0.08530	1.5617
0.42	0.19497	0.13462	0.10277	0.08309	1.5430
0.43	0.19022	0.13122	0.10012	0.08093	1.5250
0.44	0.18555	0.12788	0.09753	0.07882	1.5076
0.4444	0.18350	0.12642	0.09640	0.07789	1.5
0.45	0.18096	0.12461	0.09499	0.07675	1.4907
0.46	0.17645	0.12140	0.09250	0.07471	1.4744
0.47	0.17201	0.11824	0.09006	0.07272	1.4586
0.48	0.16764	0.11514	0.08766	0.07077	1.4434
0.49	0.16334	0.11210	0.08531	0.06885	1.4286
0.5	0.15910	0.10910	0.08300	0.06697	1.4142
0.51	0.15493	0.10616	0.08072	0.06512	1.4003
0.52	0.15082	0.10326	0.07849	0.06330	1.3868
0.53	0.14676	0.10041	0.07629	0.06151	1.3736
0.54	0.14277	0.09760	0.07413	0.05976	1.3608
0.55	0.13883	0.09484	0.07201	0.05803	1.3484
0.56	0.13494	0.09211	0.06991	0.05633	1.3363
0.57	0.13110	0.08943	0.06785	0.05466	1.3245
0.58	0.12732	0.08679	0.06582	0.05302	1.3131
0.59	0.12358	0.08418	0.06383	0.05140	1.3019
0.6	0.11989	0.08161	0.06186	0.04980	1.2910
0.61	0.11624	0.07908	0.05992	0.04823	1.2804
0.62	0.11264	0.07658	0.05800	0.04668	1.2700
0.63	0.10909	0.07412	0.05612	0.04515	1.2599
0.64	0.10557	0.07168	0.05426	0.04365	1.2500
0.65	0.10210	0.06928	0.05242	0.04216	1.2403
0.66	0.09867	0.06691	0.05061	0.04070	1.2309

TABLE 12 cont'd: Polygenic nullizygous risk, n = 2-5

G_0	p (n = 2)	p (n = 3)	p (n = 4)	p (n = 5)	GRR (1 parent)
0.67	0.09527	0.06457	0.04883	0.03926	1.2217
0.68	0.09191	0.06225	0.04706	0.03783	1.2127
0.69	0.08859	0.05997	0.04532	0.03643	1.2039
0.7	0.08531	0.05771	0.04361	0.03504	1.1952
0.71	0.08206	0.05548	0.04191	0.03367	1.1868
0.72	0.07884	0.05328	0.04023	0.03232	1.1785
0.73	0.07566	0.05110	0.03858	0.03098	1.1704
0.74	0.07251	0.04895	0.03694	0.02966	1.1625
0.75	0.06940	0.04682	0.03532	0.02836	1.1547
0.76	0.06631	0.04471	0.03372	0.02707	1.1471
0.77	0.06325	0.04263	0.03214	0.02580	1.1396
0.78	0.06023	0.04056	0.03058	0.02454	1.1323
0.79	0.05723	0.03853	0.02904	0.02330	1.1251
0.8	0.05426	0.03651	0.02751	0.02207	1.1180
0.81	0.05132	0.03451	0.02600	0.02085	1.1111
0.82	0.04840	0.03253	0.02450	0.01965	1.1043
0.83	0.04551	0.03058	0.02302	0.01846	1.0976
0.84	0.04265	0.02864	0.02156	0.01728	1.0911
0.85	0.03982	0.02672	0.02011	0.01612	1.0847
0.86	0.03700	0.02482	0.01868	0.01497	1.0783
0.87	0.03422	0.02294	0.01726	0.01383	1.0721
0.88	0.03145	0.02108	0.01585	0.01270	1.0660
0.89	0.02871	0.01923	0.01446	0.01159	1.0600
0.9	0.02600	0.01741	0.01308	0.01048	1.0541
0.91	0.02330	0.01560	0.01172	0.00939	1.0483
0.92	0.02063	0.01380	0.01037	0.00830	1.0426
0.93	0.01798	0.01202	0.00903	0.00723	1.0370
0.94	0.01535	0.01026	0.00770	0.00617	1.0314
0.95	0.01274	0.00851	0.00639	0.00512	1.0260
0.96	0.01015	0.00678	0.00509	0.00407	1.0206
0.97	0.00759	0.00506	0.00380	0.00304	1.0153
0.98	0.00504	0.00336	0.00252	0.00202	1.0102
0.99	0.00251	0.00167	0.00126	0.00100	1.0050
1	0	0	0	0	1

TABLE 13: Polygenic nullizygous risk, n = 10-25

G_0	p (n = 10)	p (n = 15)	p (n = 20)	p (n = 25)	GRR (1 parent)
0.01	0.20567	0.14230	0.10875	0.08799	10
0.02	0.17766	0.12226	0.09317	0.07526	7.0711
0.03	0.16082	0.11031	0.08393	0.06773	5.7735
0.04	0.14866	0.10174	0.07732	0.06235	5
0.05	0.13911	0.09503	0.07216	0.05816	4.4721
0.06	0.13122	0.08952	0.06792	0.05471	4.0825
0.07	0.12450	0.08483	0.06432	0.05180	3.7796
0.08	0.11864	0.08074	0.06119	0.04926	3.5355
0.09	0.11343	0.07713	0.05842	0.04702	3.3333
0.1	0.10875	0.07388	0.05594	0.04501	3.1623
0.11	0.10449	0.07093	0.05369	0.04319	3.0151
0.1111	0.10404	0.07062	0.05345	0.04299	3
0.12	0.10059	0.06824	0.05163	0.04152	2.8868
0.13	0.09698	0.06575	0.04973	0.03998	2.7735
0.14	0.09363	0.06344	0.04796	0.03856	2.6726
0.15	0.09050	0.06128	0.04632	0.03723	2.5820
0.16	0.08756	0.05926	0.04478	0.03599	2.5
0.17	0.08479	0.05735	0.04333	0.03482	2.4254
0.18	0.08217	0.05556	0.04196	0.03371	2.3570
0.19	0.07968	0.05385	0.04067	0.03267	2.2942
0.2	0.07732	0.05223	0.03944	0.03168	2.2361
0.21	0.07507	0.05069	0.03826	0.03073	2.1822
0.22	0.07291	0.04922	0.03715	0.02983	2.1320
0.23	0.07085	0.04781	0.03608	0.02897	2.0851
0.24	0.06887	0.04646	0.03505	0.02814	2.0412
0.25	0.06697	0.04516	0.03406	0.02735	2
0.26	0.06514	0.04391	0.03312	0.02658	1.9612
0.27	0.06337	0.04271	0.03220	0.02585	1.9245
0.28	0.06167	0.04154	0.03132	0.02514	1.8898
0.29	0.06002	0.04042	0.03047	0.02445	1.8570
0.3	0.05842	0.03934	0.02965	0.02379	1.8257
0.31	0.05688	0.03829	0.02886	0.02315	1.7961
0.32	0.05538	0.03727	0.02808	0.02253	1.7678

TABLE 13 cont'd: Polygenic nullizygous risk, n = 10-25

G_0	p (n = 10)	p (n = 15)	p (n = 20)	p (n = 25)	GRR (1 parent)
0.33	0.05392	0.03628	0.02734	0.02193	1.7408
0.34	0.05251	0.03532	0.02661	0.02135	1.7150
0.35	0.05114	0.03439	0.02590	0.02078	1.6903
0.36	0.04980	0.03348	0.02522	0.02023	1.6667
0.37	0.04850	0.03260	0.02455	0.01969	1.6440
0.38	0.04723	0.03174	0.02390	0.01917	1.6222
0.39	0.04599	0.03090	0.02327	0.01866	1.6013
0.4	0.04478	0.03008	0.02265	0.01816	1.5811
0.41	0.04360	0.02928	0.02204	0.01767	1.5617
0.42	0.04245	0.02850	0.02145	0.01720	1.5430
0.43	0.04132	0.02774	0.02088	0.01674	1.5250
0.44	0.04022	0.02699	0.02032	0.01629	1.5076
0.4444	0.03974	0.02667	0.02007	0.01609	1.5
0.45	0.03914	0.02627	0.01976	0.01584	1.4907
0.46	0.03808	0.02555	0.01923	0.01541	1.4744
0.47	0.03705	0.02485	0.01870	0.01499	1.4586
0.48	0.03603	0.02417	0.01818	0.01457	1.4434
0.49	0.03504	0.02350	0.01768	0.01417	1.4286
0.5	0.03406	0.02284	0.01718	0.01377	1.4142
0.51	0.03311	0.02219	0.01669	0.01338	1.4003
0.52	0.03217	0.02156	0.01622	0.01299	1.3868
0.53	0.03125	0.02094	0.01575	0.01262	1.3736
0.54	0.03034	0.02033	0.01529	0.01225	1.3608
0.55	0.02945	0.01973	0.01483	0.01189	1.3484
0.56	0.02857	0.01914	0.01439	0.01153	1.3363
0.57	0.02771	0.01856	0.01395	0.01118	1.3245
0.58	0.02687	0.01799	0.01353	0.01084	1.3131
0.59	0.02604	0.01743	0.01310	0.01050	1.3019
0.6	0.02522	0.01688	0.01269	0.01016	1.2910
0.61	0.02441	0.01634	0.01228	0.00984	1.2804
0.62	0.02362	0.01581	0.01188	0.00952	1.2700
0.63	0.02284	0.01528	0.01148	0.00920	1.2599
0.64	0.02207	0.01477	0.01110	0.00889	1.2500
0.65	0.02131	0.01426	0.01071	0.00858	1.2403
0.66	0.02056	0.01376	0.01033	0.00828	1.2309

TABLE 13 cont'd: Polygenic nullizygous risk, n = 10-25

G_0	p (n = 10)	p (n = 15)	p (n = 20)	p (n = 25)	GRR (1 parent)
0.67	0.01982	0.01326	0.00996	0.00798	1.2217
0.68	0.01910	0.01277	0.00960	0.00768	1.2127
0.69	0.01838	0.01229	0.00923	0.00739	1.2039
0.7	0.01768	0.01182	0.00888	0.00711	1.1952
0.71	0.01698	0.01135	0.00853	0.00683	1.1868
0.72	0.01629	0.01089	0.00818	0.00655	1.1785
0.73	0.01561	0.01044	0.00784	0.00627	1.1704
0.74	0.01494	0.00999	0.00750	0.00600	1.1625
0.75	0.01428	0.00954	0.00717	0.00574	1.1547
0.76	0.01363	0.00911	0.00684	0.00547	1.1471
0.77	0.01298	0.00867	0.00651	0.00521	1.1396
0.78	0.01235	0.00825	0.00619	0.00496	1.1323
0.79	0.01172	0.00783	0.00588	0.00470	1.1251
0.8	0.01110	0.00741	0.00556	0.00445	1.1180
0.81	0.01048	0.00700	0.00525	0.00421	1.1111
0.82	0.00987	0.00659	0.00495	0.00396	1.1043
0.83	0.00927	0.00619	0.00465	0.00372	1.0976
0.84	0.00868	0.00579	0.00435	0.00348	1.0911
0.85	0.00809	0.00540	0.00405	0.00325	1.0847
0.86	0.00751	0.00501	0.00376	0.00301	1.0783
0.87	0.00694	0.00463	0.00348	0.00278	1.0721
0.88	0.00637	0.00425	0.00319	0.00255	1.0660
0.89	0.00581	0.00388	0.00291	0.00233	1.0600
0.9	0.00525	0.00351	0.00263	0.00210	1.0541
0.91	0.00470	0.00314	0.00235	0.00188	1.0483
0.92	0.00416	0.00278	0.00208	0.00167	1.0426
0.93	0.00362	0.00242	0.00181	0.00145	1.0370
0.94	0.00309	0.00206	0.00155	0.00124	1.0314
0.95	0.00256	0.00171	0.00128	0.00103	1.0260
0.96	0.00204	0.00136	0.00102	0.00082	1.0206
0.97	0.00152	0.00101	0.00076	0.00061	1.0153
0.98	0.00101	0.00067	0.00050	0.00040	1.0102
0.99	0.00050	0.00033	0.00025	0.00020	1.0050
1	0	0	0	0	1

DISCUSSION, CONCLUSIONS, AND SUGGESTIONS FOR FURTHER RESEARCH

Stipulations

Range of values considered for all common late-onset cancers, independent of modes of inheritance:

$3.0 > GRR > 1.5$ (for most late-onset cancers, Table 14)(Hemminki, Rawal et al. 2004)

$G_0 > 0.01$ (definition of "common" disease)

$q > 0$ (mean value 0.03, 99% u.c.l. ~ 0.22, from Morgenthaler & Thilly, 2006)

Range of values considered for CRC, using estimates undergoing refinement by additional analyses, based on the Swedish Family Cancer Database:

$3.0 > GRR > 1.5$ (1.77 for rectal cancer, 1.86 for CRC, 2.02 for colon cancer, 2.04 for colorectal adenocarcinoma, 2.58 for age-matched parents-offspring with colon cancer) (Hemminki, Rawal et al. 2004; Hemminki, Granstrom et al. 2005)

$G_0 > 0.26$ (from Figure 6)

Monogenic heterozygous risk

[Figures 8.1-8.4 and Table 1]

All common late-onset cancers

It is clear from inspection of Table 1 that for risk of a common late onset cancer to be inherited as a monogenic heterozygous condition, values of *GRR* must be less than 3.0 if the fraction of the population at risk is greater than 0.16. Only a few forms of cancer - skin, prostate, colorectal, breast, and lung (among cigarette smokers) - meet this criterion for minimum lifetime risk, *mLR*, which we use an estimate of the minimum value of G_0. Pancreatic cancer, the fifth most deadly form of cancer in the U.S., whose mortality estimates are essentially identical with incidence estimates, has an *mLR* for males of only 0.08-0.09. Assuming *pro tempore* that environmental risks now reach nearly all persons,

values of *GRR* for pancreatic cancer would have to lie between 5 and 6, were risk to be encoded as heterozygosity in a single gene. As the Hemminki group has reported few estimates of *GRR* for late-onset cancers higher than 3.0, but has reported a general absence of spousal risk, it appears that risk of pancreatic and the less common forms of cancer is excluded from consideration of genetic risk from monogenic heterozygosity.

Colorectal cancer

Using the assumption $E\sim1$, the observation $G_0 > 0.26$, and a strictly preliminary range of estimates of *GRR* for male CRC risk of $1.7 < GRR < 2.5$ reported by the Hemminki group it appears possible that CRC risk can be conferred by monogenic heterozygosity. Values of $G_0 > 0.26$ and $1.7 < GRR < 1.9$ are found in Table 1. These values correspond to values of $0.15 < q < 0.17$ which is less than the 99% u.c.l. calculated for the values of q for all non-deleterious gene inactivating mutant fractions in human genes. If, for instance, were recalculations for CRC to find $GRR > 2.0$, then this mode of simple Mendelian inheritance would appear to be excluded.

Monogenic nullizygous risk

[Figures 9.1-9.4 and Table 2]

All common late-onset cancers

It is clear from inspection of Table 2 that for risk of a common late onset cancer to be inherited as a monogenic nullizygous condition, values of *GRR* must be less than 2.5 if the fraction of the population at risk is greater than 0.16. Only a few forms of cancer - skin, prostate, colorectal, breast, and lung (among cigarette smokers) - meet this criterion for minimum lifetime risk, *mLR*, which we use an estimate of the minimum value of G_0. Pancreatic cancer has an *mLR* for males of only 0.08-0.09. Assuming *pro tempore* that environmental risks now reach nearly all persons, *i.e.* $E\sim1$, values of *GRR* for pancreatic cancer would have to lie between 3.3 and 3.5, were risk to be encoded as nullizygosity in a single gene. As the Hemminki group has reported few estimates of *GRR* for late-onset cancers higher than 3.0, but has reported a general absence of spousal risk, it appears that

the less common forms of cancer are excluded from consideration of genetic risk from monogenic nullizygosity.

Colorectal cancer

Using the assumption E~1, the observation $G_0 > 0.26$, and a strictly preliminary range of estimates of *GRR* for male CRC risk of $1.7 < GRR < 2.5$ reported by the Hemminki group it appears possible that CRC risk can be conferred by monogenic nullizygosity. Values of $G_0 > 0.26$ and $1.7 < GRR < 1.9$ are found in Table 2. These values correspond, however, to values of $0.5 < q < 0.6$ which is considerably than the 99% u.c.l. calculated for the values of q for all non-deleterious gene inactivating mutant fractions in human genes. If, for instance, were recalculations for CRC to find $GRR > 2.0$ then this mode of simple Mendelian inheritance would appear to be excluded.

Multigenic heterozygous risk (m=2, each gene contributing equal risk)

[Figures 10.1-10.4 and Table 3]

All common late-onset cancers

It is clear from inspection of Table 3 that for risk of a common late onset cancer to be inherited as a bigenic heterozygous condition, values of *GRR* must be less than 3.0 if the fraction of the population at risk is greater than 0.19. Cancers of the skin, prostate, colorectum, breast, and lung (among cigarette smokers), meet this criterion for minimum lifetime risk, *mLR*, which we use an estimate of the minimum value of G_0. Pancreatic cancer has an *mLR* for males of only 0.08-0.09. Assuming *pro tempore* that environmental risks now reach nearly all persons, values of *GRR* for pancreatic cancer would have to lie between 5.8 and 6.5 were risk to be encoded in equal amount as heterozygosity in each of two separate genes. As the Hemminki group has reported few estimates of *GRR* for late-onset cancers higher than 3.0, but has reported a general absence of spousal risk, it appears that risk the less common forms of cancer are excluded from consideration of genetic risk from bigenic heterozygosity.

Colorectal cancer

Using the assumption E~1, the observation $G_0 > 0.26$, and a strictly preliminary range of estimates of *GRR* for male CRC risk of $1.7 < GRR < 2.5$ reported by the Hemminki group it appears possible that CRC risk can be conferred by bigenic heterozygosity. Values of $G_0 > 0.26$ and $1.7 < GRR < 2.2$ are found in Table 3. These values correspond to values of $0.07 < q < 0.11$ which is less than the 99% u.c.l. calculated for the values of q for all non-deleterious gene inactivating mutant fractions in human genes. If, for instance, recalculations for CRC to find $GRR > 2.2$, then this mode of simple Mendelian inheritance would appear to be excluded.

Multigenic heterozygous risk (m=3, each gene contributing equal risk)

[Figures 11.1-11.4 and Table 4]

All common late-onset cancers

It is clear from inspection of Table 4 that for risk of a common late onset cancer to be inherited as a trigenic heterozygous condition, values of *GRR* must be less than 3.0 if the fraction of the population at risk is greater than 0.19. Cancers of the skin, prostate, colorectum, breast, and lung (among cigarette smokers), meet this criterion for minimum lifetime risk, *mLR*, which we use an estimate of the minimum value of G_0. Pancreatic cancer has an *mLR* for males of only 0.08-0.09. Assuming *pro tempore* that environmental risks now reach nearly all persons, values of *GRR* for pancreatic cancer would have to lie between 5.9 and 6.6 were risk to be encoded in equal amount as heterozygosity in each of three separate genes. As the Hemminki group has reported few estimates of *GRR* for late-onset cancers higher than 3.0, but has reported a general absence of spousal risk, it appears that risk the less common forms of cancer are excluded from consideration of genetic risk from trigenic heterozygosity.

Colorectal cancer

Using the assumption E~1, the observation $G_0 > 0.26$, and a strictly preliminary range of estimates of *GRR* for male CRC risk of $1.7 < GRR < 2.5$ reported by the Hemminki group

it appears possible that CRC risk can be conferred by trigenic heterozygosity. Values of $G_0 > 0.26$ and $1.7 < GRR < 2.3$ are found in Table 4. These values correspond to values of $0.05 < q < 0.08$ which approach the mean values of q of ~0.03 for all non-deleterious gene-inactivating mutant fractions in human genes. If, for instance, recalculations for CRC to find $GRR > 2.3$, then this mode of simple Mendelian inheritance would appear to be excluded.

Multigenic heterozygous risk (m=4, each gene contributing equal risk)

[Figures 12.1-12.4 and Table 5]

All common late-onset cancers

It is clear from inspection of Table 5 that for risk of a common late onset cancer to be inherited as a tetragenic heterozygous condition, values of *GRR* must be less than 3.0 if the fraction of the population at risk is greater than 0.19. Cancers of the skin, prostate, colorectum, breast, and lung (among cigarette smokers), meet this criterion for minimum lifetime risk, *mLR*, which we use an estimate of the minimum value of G_0. Pancreatic cancer has an *mLR* for males of only 0.08-0.09. Assuming *pro tempore* that environmental risks now reach nearly all persons, values of *GRR* for pancreatic cancer would have to lie between 5.9 and 6.6 were risk to be encoded in equal amount as heterozygosity in each of four separate genes. As the Hemminki group has reported few estimates of *GRR* for late-onset cancers higher than 3.0, but has reported a general absence of spousal risk, it appears that risk the less common forms of cancer are excluded from consideration of genetic risk from tetragenic heterozygosity.

Colorectal cancer

Using the assumption E~1, the observation $G_0 > 0.26$, and a strictly preliminary range of estimates of *GRR* for male CRC risk of $1.7 < GRR < 2.5$ reported by the Hemminki group it appears possible that CRC risk can be conferred by tetragenic heterozygosity. Values of $G_0 > 0.26$ and $1.7 < GRR < 2.3$ are found in Table 5. These values correspond to values of $0.04<q<0.06$ which approach the mean values of q of ~0.03 for all non-deleterious

gene inactivating mutant fractions in human genes. If, for instance, were recalculations for CRC to find $GRR > 2.3$ then this mode of simple Mendelian inheritance would appear to be excluded.

Multigenic heterozygous risk (m = 5, each gene contributing equal risk)

[Figures 13.1-13.4 and Table 6]

All common late-onset cancers

It is clear from inspection of Table 6 that for risk of a common late onset cancer to be inherited as a pentagenic heterozygous condition, values of *GRR* must be less than 3.0 if the fraction of the population at risk is greater than 0.19. Cancers of the skin, prostate, colorectum, breast, and lung (among cigarette smokers), meet this criterion for minimum lifetime risk, *mLR*, which we use an estimate of the minimum value of G_0. Pancreatic cancer has an *mLR* for males of only 0.08-0.09. Assuming *pro tempore* that environmental risks now reach nearly all persons, values of *GRR* for pancreatic cancer would have to lie between 5.9 and 6.7 were risk to be encoded in equal amount as heterozygosity in each of 5 separate genes. As the Hemminki group has reported few estimates of *GRR* for late-onset cancers higher than 3.0, but has reported a general absence of spousal risk, it appears that risk the less common forms of cancer are excluded from consideration of genetic risk from pentagenic heterozygosity.

Colorectal cancer

Using the assumption $E \sim 1$, the observation $G_0 > 0.26$, and a strictly preliminary range of estimates of *GRR* for male CRC risk of $1.7 < GRR < 2.5$ reported by the Hemminki group it appears possible that CRC risk can be conferred by pentagenic heterozygosity. Values of $G_0 > 0.26$ and $1.70 < GRR < 2.33$ are found in Table 6. These values correspond to values of $0.03 < q < 0.05$ which includes the mean values of q of ~0.03 for all non-deleterious gene inactivating mutant fractions in human genes. If, for instance, were recalculations for CRC to find $GRR > 2.3$, then this mode of simple Mendelian inheritance would appear to be excluded.

Hypothesis about multigenic heterozygous risk derived from this exercise

Once, however, one finds that risk as multigenic heterozygosity (m= 5, 6, 7) is in accord with values of q estimated for all human genes carrying nondeleterious gene-inactivating, one, in the words given to Hercule Poirot, "is given furiously to think". These findings offer a fairly simple hypothesis bringing together some disparate points of CRC epidemiology, physiology and familial risk. Could risk for late-onset CRC be inherited as heterozygosity for any of 5-7 proto-oncogenes? No oncogenes for which activation is required for neoplastic transformation have been found in all tumor cells in human tumors. This conclusion stands despite the presence of sometimes large tumor sectors "activated" for *RAS* and other murine and avian oncogenes. Could there be a set of conditions of heterozygosity such $A^{+/-}$, $B^{+/-}$, $C^{+/-}$, $D^{+/-}$, $E^{+/-}$, $F^{+/-}$ in which heterozygosity for any of these genes would confer risk of tumor promotion by alteration of the active allele in a preneoplastic stem cell? Homozygous and nullizygous persons for genes *A-F* would not be at risk as two rare events would be required for activation of one gene copy and alteration of the other (homozygotes) while re-activation and then specific oncogenic alteration of an inactivated gene copy (nullizygotes) would be at rates below any reasonable expectation Such a scenario is in accord with estimates of rates of promotion by single mutations in colorectal cancer in U.S. European American males (Herrero-Jimenez, Tomita-Mitchell et al. 2000) and is worth considering further.

Multigenic nullizygous risk (m=2, each gene contributing equal risk)

[Figures 14.1-14.4 and Table 7]

All common late-onset cancers

It is clear from inspection of Table 7 that for risk of a common late onset cancer to be inherited as a bigenic nullizygous condition, values of *GRR* must be less than 3.0 if the fraction of the population at risk is greater than 0.076. Pancreatic cancer, the fifth most deadly cancer form in the U.S. and for which mortality estimates are essentially identical with incidence has an *mLR* for males of 0.08-0.09. But *mLR* for leukemia and other

cancers appear to be lower than 0.04. Assuming *pro tempore* that environmental risks now reach nearly all persons, values of *GRR* for pancreatic cancer would have to lie between 2.8 and 2.9 were risk to be encoded in equal amount as nullizygosity in each of two separate genes. As the Hemminki group has reported few estimates of *GRR* for late-onset cancers higher than 3.0, but has reported a general absence of spousal risk, it appears that risk the less common forms of cancer are excluded from consideration of genetic risk from bigenic nullizygosity.

Colorectal cancer

Using the assumption E~1, the observation $G_0 > 0.26$, and a strictly preliminary range of estimates of *GRR* for male CRC risk of $1.7 < GRR < 2.5$ reported by the Hemminki group it appears possible that CRC risk can be conferred by bigenic nullizygosity. Values of $G_0 > 0.26$ and $1.72 < GRR < 1.77$ are found in Table 7. These values correspond to values of $0.37 < q < 0.39$ which is well above the 99% u.c.l. (0.22) calculated for the values of q for all non-deleterious gene inactivating mutant fractions in human genes. If, for instance, recalculations for CRC to find $GRR > 1.8$, then this bigenic mode of simple Mendelian inheritance would appear to be excluded.

Multigenic nullizygous risk (m=3, each gene contributing equal risk)

[Figures 15.1-15.4 and Table 8]

All common late-onset cancers

It is clear from inspection of Table 8 that for risk of a common late onset cancer to be inherited as a trigenic nullizygous condition, values of *GRR* must be less than 3.0 if the fraction of the population at risk is greater than 0. 06. Pancreatic cancer, the fifth most deadly cancer form in the U.S. and for which mortality estimates are essentially identical with incidence has an *mLR* for males of only 0.08-0.09. Assuming *pro tempore* that environmental risks now reach nearly all persons, values of *GRR* for pancreatic cancer would have to lie between 2.5 and 2.6 were risk to be encoded in equal amount as nullizygosity in each of three separate genes. As the Hemminki group has reported few

estimates of *GRR* for late-onset cancers higher than 3.0, but has reported a general absence of spousal risk, it appears the less common forms of cancer are excluded from consideration of genetic risk from trigenic nullizygosity.

Colorectal cancer

Using the assumption $E \sim 1$, the observation $G_0 > 0.26$, and a strictly preliminary range of estimates of *GRR* for male CRC risk of $1.7 < GRR < 2.5$ reported by the Hemminki group it appears possible that CRC risk *cannot* be conferred by trigenic nullizygosity . Values of $G_0 > 0.26$ and $GRR > 1.7$ are not found in Table 8. If, for instance, were recalculations for CRC to find $GRR > 2.3$ then this mode of simple Mendelian inheritance would appear to be excluded with greater statistical assurance.

Multigenic nullizygous risk (m=4, each gene contributing equal risk)

[Figures 16.1-16.4 and Table 9]

All common late-onset cancers

It is clear from inspection of Table 9 that for risk of a common late onset cancer to be inherited as a tetragenic nullizygous condition, values of *GRR* must be less than 3.0 if the fraction of the population at risk is greater than 0.047. Pancreatic cancer, the fifth most deadly cancer form in the U.S. and for which mortality estimates are essentially identical with incidence has an *mLR* for males of only 0.08-0.09. Assuming *pro tempore* that environmental risks now reach nearly all persons, values of *GRR* for pancreatic cancer would have to lie between 2.34 and 2.44 were risk to be encoded in equal amount as heterozygosity in each of four separate genes. As the Hemminki group has reported few estimates of *GRR* for late-onset cancers higher than 3.0, but has reported a general absence of spousal risk, it appears that risk the less common forms of cancer should be included when considering genetic risk from tetragenic nullizygosity.

Colorectal cancer

Using the assumption E~1, the observation $G_0 > 0.26$, and a strictly preliminary range of estimates of *GRR* for male CRC risk of $1.7 < GRR < 2.5$ reported by the Hemminki group it appears possible that CRC risk can be conferred by tetragenic heterozygosity. Values of $G_0 > 0.26$ and $GRR > 1.7$ are not found in Table 9. This mode, tetragenic nullizygous risk, of simple Mendelian inheritance would appear to be excluded.

Multigenic nullizygous risk (m = 5, each gene contributing equal risk)

[Figures 17.1-17.4 and Table 10]

All common late-onset cancers

It is clear from inspection of Table 10 that for risk of a common late-onset cancer to be inherited as a pentagenic nullizygous condition, values of *GRR* must be less than 3.0, if the fraction of the population at genetic risk is greater than 0.04. Several forms of cancer, such as leukemia, CNS cancers, and lymphoma, may meet this criterion for minimum lifetime risk, *mLR*, which we use an estimate of the minimum value of G_0. Pancreatic cancer, the fifth most deadly cancer form in the U.S. and for which mortality estimates are essentially identical with incidence has an *mLR* for males of only 0.08-0.09. Assuming *pro tempore* that environmental risks now reach nearly all persons, values of *GRR* for pancreatic cancer would have to lie between 2.2 and 2.3 were risk to be encoded in equal amount as nullizygosity for each of 5 separate genes. As the Hemminki group has reported few estimates of *GRR* for late-onset cancers higher than 3.0, but has reported a general absence of spousal risk, it appears that risk from pentagenic nullizygosity should be included when considering genetic risk for the less common forms of cancer.

Colorectal cancer

Using the assumption E~1, the observation $G_0 > 0.26$ and a strictly preliminary range of estimates of *GRR* for male CRC risk of $1.7 < GRR < 2.5$ reported by the Hemminki group it does not appear to be possible that CRC risk can be conferred by pentagenic nullizygosity. No values of $G_0 > 0.26$ and $GRR > 1.70$ are found in Table 10. This mode of simple Mendelian inheritance would appear to be excluded.

Polygenic heterozygous risk (n = 2, each gene contributing equal risk)

[Figures 18.1-18.4, Table 11]

All common late-onset cancers

It is clear from inspection of Table 11 that for risk of a common late onset cancer to be inherited as a polygenic (n=2) heterozygous condition, values of *GRR* must be less than 3.0 if the fraction of the population at risk is greater than 0.083. Cancers of the skin, prostate, colorectum, breast, and lung (among cigarette smokers), meet this criterion for minimum lifetime risk, *mLR*, which we use an estimate of the minimum value of G_0. Pancreatic cancer has an *mLR* for males of only 0.08-0.09. Assuming *pro tempore* that environmental risks now reach nearly all persons, values of *GRR* for pancreatic cancer would have to lie between 2.8 and 3.0 were risk to be encoded in equal amount as heterozygosity in each of two separate genes. As the Hemminki group has reported few estimates of *GRR* for late-onset cancers higher than 3.0, but has reported a general absence of spousal risk, it appears that risk the less common forms of cancer are excluded from consideration of genetic risk from polygenic (n=2) heterozygosity.

Colorectal cancer

Using the assumption E~1, the observation $G_0 > 0.26$, and a strictly preliminary range of estimates of *GRR* for male CRC risk of $1.7 < GRR < 2.5$ reported by the Hemminki group, it does not appear that CRC risk can be conferred by polygenic (n=2) heterozygosity. No values of $G_0 > 0.26$ and $GRR > 1.7$ are found in Table 11. This mode of simple Mendelian inheritance would appear to be excluded as the basis of genetic risk for CRC. It can be seen by inspection that all values of polygenic (n>2) nullizygosity are similarly excluded for CRC and the several other most common forms of cancer.

Polygenic homozygous risk (n = 2-25, each gene contributing equal risk)

[Figures 19.1-19.4, Table 12] (N.B. values are for (1-q) = p in this table)

All common late-onset cancers

It is clear from inspection of Table 12 that for risk of a common late onset cancer to be inherited as a polygenic (n>2) homozygous condition, values of *GRR* must be less than 3.0 if the fraction of the population at risk is greater than 0.11. Cancers of the skin, prostate, colorectum, breast, and lung (among cigarette smokers), meet this criterion for minimum lifetime risk, *mLR*, which we use an estimate of the minimum value of G_0. Pancreatic cancer has an *mLR* for males of only 0.08-0.09. Assuming *pro tempore* that environmental risks now reach nearly all persons, values of $GRR < 3.0$ do not exist for polygenic (n>2) homozygosity for each of two to five separate genes for pancreatic cancer. It appears that risk the less common forms of cancer, including pancreatic cancer, are excluded from consideration of genetic risk from polygenic (n>2) homozygosity.

Colorectal cancer

Using the assumption $E \sim 1$, the observation $G_0 > 0.26$, and a strictly preliminary range of estimates of *GRR* for male CRC risk of $1.7 < GRR < 2.5$ reported by the Hemminki group it appears possible that CRC risk can be conferred by polygenic (n>2) homozygosity. Values of $G_0 > 0.26$ and $1.7 < GRR < 1.96$ are found in Table 12. This mode of simple Mendelian inheritance would appear to be worthy of further consideration as the basis of genetic risk for CRC.

These observations are summarized in Table 15 as calculated values of G_0 and *GRR*.

Monogenic risk of CRC

Using CRC as an example with $G_0 > 0.26$ and $1.5 < GRR < 3.0$, a key conclusion is reached: both monogenic heterozygous and nullizygous conditions of risk are included in the set of possible modes of CRC risk inheritance. Were we, however, to use the estimate, $G_0 \sim 0.35$, derived from the relatively small set of proctoscopic studies reported by Atkin (1993), monogenic nullizygous risk but not monogenic heterozygous risk is still included for values of *GRR* between 1.5 and 1.7. Monogenic risks require values of q

considerably above the 99% u.c.l. of 0.22 estimated for genes carrying gene-inactivating non-deleterious mutations. Calculating the value of *FRR(t)* for father to son and mother to daughter transmission of CRC risk for the age-intervals 50-54,...,70-74, should provide ten separate estimates of *GRR* for CRC and permit some refinement of the rang we are obliged to use herein. For example, if *GRR* for CRC is found to exceed 2.0, all forms of monogenic risk will be seen to be excluded.

Multigenic risk of CRC

Using CRC as an example with $G_0 > 0.26$ and $1.5 < GRR < 3.0$, it is clear that multigenic heterozygosity is not excluded for a range $2.5 > GRR > 1.5$ while multigenic nullizygous risks are not excluded within a more narrow range $\sim 1.7 > GRR > 1.5$. Were the estimate of CRC $G_0 = 0.35$ applied only more narrow ranges of *GRR* near 1.5 would continue to include these possibilities. Recalculation of *GRR* as outlined above is clearly the next crucial step on this scientific path. Note that if CRC *GRR* were found to be >2.0 the only mode of Mendelian risk inheritance not excluded would be any of several levels of multigenic heterozygosity.

Polygenic risk of CRC

Using CRC as an example with $G_0 > 0.26$ and $GRR < 3.0$, it appears that these modes are generally excludable on the basis of the estimated minimum of G_0 and *GRR* alone. This exclusion is, however, limited to the unlikely combination of genes carrying high values of q. A useful follow-up in this regard would be to apply the general equations for polygenicity in which values of q range broadly.

The calculations summarized in Table 16 offer food for additional thought. Using Morgenthaler and Thilly's estimation of a mean mutant fraction of 0.03 for gene inaxctivating mutations in genes that carry such non-deleterious alleles and their estimate of a 99% u.c.l of 0.22 we may apply that range to the values of Table 14. Note that high values of q cannot be excluded, because a high value may occur once by chance and a few such values might conceivably account for the values of G_0 for the most common late-onset forms of cancer: skin, prostate, colorectal and breast. Multigenic or polygenic

modes with multiple mutations each with high inherited mutant fractions may reasonably be considered unlikely. But a model of multigenicity with genes carrying different mutant fractions with the most frequent accounting for 1/3 to 2/3 of the total risk can be created using the general equations derived herein and this possibility is worth pursuing. When the multigenic or polygenic models show average values of q approaching the expected mean of $q = 0.03$, special attention is due, as the expected mean of several genes drawn by chance is 0.03.

The most important conclusion of this thesis is that it is possible to build a logical quantitative structure to evaluate the probability that risk for a particular form of cancer is transmitted by any of a series of simple Mendelian modes.

This process began with the use of lifetime incidence and mortality data to estimate the minimum value of G_0, as in Figures 5 and 6 for colorectal cancers. In this we applied the reasoning developed in Herrero-Jimenez et al. (2000).

The second step, prescribed by our arguments, was taken to estimate the genetic relative risk, *GRR*, by determining specific values of familial risk *FRR* in the age intervals 50-54 up to 70-74 for father-to-son and mother-to-daughter risk transmission. This work provides a valuable redefinition of *GRR*, not heretofore developed, that should chart a path for these calculations for all common cancers recorded in the Swedish Family Cancer Database, organized by Prof. K. Hemminki.

The third step was to use the graphs and tables provided to discover if the values of G_0 and *GRR* are excluded for each form of simple Mendelian inheritance posited. In these, we extend the work of Li and Sacks (1954) by creating formulæ for *GRR* in the cases of multigenic and polygenic risks.

The application of this method to colorectal cancer in parent-to-child transmission in Sweden indicates the important conclusion that monogenic forms of transmission are not excluded by existing data but that recalculation of *GRR* (in progress) may in fact do so.

Application of this general method should be of immediate use to those who are planning pan-genomic searches for genes that confer risk to common diseases such as cancers.

A major caveat must be stressed. If cardiovascular disease(s) share common determinants of genetic risk with any other common disease such as CRC, estimates of G_0 derived from lifetime incidence or mortality data will be gross underestimates. Such estimates would need to be increased via a model in which persons at risk of the disease observed competed with cardiovascular disease in each age interval. Creation of such a quantitative model would prove useful in setting bounds on estimates of gene frequencies. Herein, we assumed no interaction of cardiovascular risk, but future research may uncover links, heretofore not discovered.

In this effort, we have tried to contribute to the development of a novel logical means to analyze the genetics of common diseases such as cancers in the human population. In concluding our part in this effort, we are confident that we have made original and apparently important progress. We hope that will serve to guide others and stimulate them to press on to more precise estimates of key parameters, and investigate more carefully the possibility of shared genetic risks among cancers and other common diseases.

TABLE 14: SIR (Standardized Incidence Ratios) for offspring with parental history Adaptation with permission (Hemminki, Rawal et al. 2004)

Site	SIR
Thyroid	7.13
Testicular	4.58
Esophageal	3.82
Ovarian	2.91
Endometrial	2.74
Prostate	2.42
Skin (SCC)	2.39
Multiple Myeloma	2.33
Endocrine	2.23
CRC	2.04
Cervix	1.95
Lung	1.90
Breast	1.80
Non-Hodgkin's	1.76
Bladder	1.73
Kidney	1.67
Nervous System	1.63
Leukemia	1.53
Pancreas	1.53

TABLE 15: Values of G_0 corresponding to values of *GRR* of 3.0, 2.5, 2.0, and 1.5

	***GRR* = 3.0**	***GRR* = 2.5**	***GRR* = 2.0**	***GRR* = 1.5**
mono.het	0.1667	0.2	0.25	0.3333
mono.nul	0.1111	0.16	0.25	0.4444
multi.het.2	0.1827	0.2238	0.2892	0.4108
multi.het.3	0.1883	0.2323	0.3036	0.4403
multi.het.4	0.1912	0.2367	0.3109	0.4553
multi.het.5	0.1929	0.2393	0.3154	0.4644
multi.nul.2	0.0760	0.1157	0.1955	0.3891
multi.nul.3	0.0580	0.0912	0.1617	0.3483
multi.nul.4	0.0470	0.0754	0.1383	0.3165
multi.nul.5	0.0395	0.0643	0.1209	0.2906
poly.het.2	0.0833	0.1	0.125	0.1667
poly.hom	0.1111	0.16	0.25	0.4444

TABLE 16: Values of q corresponding to values of *GRR* of 3.0, 2.5, 2.0, and 1.5

	***GRR* = 3.0**	***GRR* = 2.5**	***GRR* = 2.0**	***GRR* = 1.5**
mono.het	0.09175	0.11270	0.14645	0.21132
mono.nul	0.33333	0.4	0.5	0.66667
multi.het.2	0.05052	0.06352	0.08583	0.13422
multi.het.3	0.03480	0.04412	0.06046	0.09743
multi.het.4	0.02654	0.03379	0.04662	0.07628
multi.het.5	0.02144	0.02737	0.03793	0.06261
multi.nul.2	0.19683	0.24420	0.32104	0.46735
multi.nul.3	0.14042	0.17708	0.23893	0.36471
multi.nul.4	0.10933	0.13927	0.19109	0.30123
multi.nul.5	0.08958	0.11490	0.15954	0.25762
poly.het.2	0.17494	0.19687	0.22940	0.28582
poly.hom.2	0.42265	0.36754	0.29289	0.18350
poly.hom.3	0.30664	0.26319	0.20630	0.12642
poly.hom.4	0.24016	0.20473	0.15910	0.09640
poly.hom.5	0.19726	0.16745	0.12945	0.07789
poly.hom.10	0.10404	0.08756	0.06697	0.03974
poly.hom.15	0.07062	0.05926	0.04516	0.02667
poly.hom.20	0.05345	0.04478	0.03406	0.02007
poly.hom.25	0.04299	0.03599	0.02735	0.01609

REFERENCES

Armitage, P. and R. Doll (1957). "A two-stage theory of carcinogenesis in relation to the age distribution of human cancer." Br J Cancer **11**(2): 161-9.

Atkin, W., P. Rogers, et al. (2004). "Wide variation in adenoma detection rates at screening flexible sigmoidoscopy." Gastroenterology **126**(5): 1247-56.

Beckman, R. A. and L. A. Loeb (2006). "Efficiency of carcinogenesis with and without a mutator mutation." Proc Natl Acad Sci U S A **103**(38): 14140-5.

Cascorbi, I. (2003). "Pharmacogenetics of cytochrome p4502D6: genetic background and clinical implication." Eur J Clin Invest **33 Suppl 2**: 17-22.

Czene, K., P. Lichtenstein, et al. (2002). "Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database." Int J Cancer **99**(2): 260-6.

Frank, S. A. (2004). "Age-specific acceleration of cancer." Curr Biol **14**(3): 242-6.

Goldgar, D. E., D. F. Easton, et al. (1994). "Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands." J Natl Cancer Inst **86**(21): 1600-8.

Gostjeva, E. V. and W. G. Thilly (2005). "Stem cell stages and the origins of colon cancer: a multidisciplinary perspective." Stem Cell Rev **1**(3): 243-51.

Gruber, S. B., N. A. Ellis, et al. (2002). "BLM heterozygosity and the risk of colorectal cancer." Science **297**(5589): 2013.

Hageman, G. S., D. H. Anderson, et al. (2005). "A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration." Proc Natl Acad Sci U S A **102**(20): 7227-32.

Hemminki, K. and B. Chen (2004). "Familial association of colorectal adenocarcinoma with cancers at other sites." Eur J Cancer **40**(16): 2480-7.

Hemminki, K. and B. Chen (2004). "Familial association of leukemia with colorectal cancer." Leuk Res **28**(10): 1113-5.

Hemminki, K. and B. Chen (2004). "Familial risk for colon and rectal cancers." Int J Cancer **111**(5): 809-10.

Hemminki, K. and B. Chen (2004). "Familial risk for colorectal cancers are mainly due to heritable causes." Cancer Epidemiol Biomarkers Prev **13**(7): 1253-6.

Hemminki, K., C. Dong, et al. (2001). "Cancer risks to spouses and offspring in the Family-Cancer Database." Genet Epidemiol **20**(2): 247-57.

Hemminki, K., C. Granstrom, et al. (2005). "The Swedish Family-Cancer Database: Update, Application to Colorectal Cancer and Clinical Relevance." Hereditary Cancer in Clinical Practice **3**(1): 7-18.

Hemminki, K. and Y. Jiang (2002). "Cancer risks among long-standing spouses." Br J Cancer **86**(11): 1737-40.

Hemminki, K. and Y. Jiang (2002). "Familial and second gastric carcinomas: a nationwide epidemiologic study from Sweden." Cancer **94**(4): 1157-65.

Hemminki, K., X. Li, et al. (2001). "The nation-wide Swedish family-cancer database--updated structure and familial rates." Acta Oncol **40**(6): 772-7.

Hemminki, K., R. Rawal, et al. (2004). "Genetic epidemiology of cancer: from families to heritable genes." Int J Cancer **111**(6): 944-50.

Hemminki, K., P. Vaittinen, et al. (1999). "Endometrial cancer in the family-cancer database." Cancer Epidemiol Biomarkers Prev **8**(11): 1005-10.

Herrero-Jimenez, P., G. Thilly, et al. (1998). "Mutation, cell kinetics, and subpopulations at risk for colon cancer in the United States." Mutat Res **400**(1-2): 553-78.

Herrero-Jimenez, P., A. Tomita-Mitchell, et al. (2000). "Population risk and physiological rate parameters for colon cancer. The union of an explicit model for carcinogenesis with the public health records of the United States." Mutat Res **447**(1): 73-116.

Kimura, M. and J. F. Crow (1964). "The Number of Alleles That Can Be Maintained in a Finite Population." Genetics **49**: 725-38.

Lander, E. S. and D. Botstein (1989). "Mapping mendelian factors underlying quantitative traits using RFLP linkage maps." Genetics **121**(1): 185-99.

Lander, E. S. and N. J. Schork (1994). "Genetic dissection of complex traits." Science **265**(5181): 2037-48.

Li, C. C. and L. Sacks (1954). "The Derivation of Joint Distribution and Correlation between Relatives by the Use of Stochastic Matrices." Biometrics **10**(3): 347-360.

Lichtenstein, P., N. V. Holm, et al. (2000). "Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland." N Engl J Med **343**(2): 78-85.

Luebeck, E. G. and S. H. Moolgavkar (2002). "Multistage carcinogenesis and the incidence of colorectal cancer." Proc Natl Acad Sci U S A **99**(23): 15095-100.

McDougall, I., F. H. Brown, et al. (2005). "Stratigraphic placement and age of modern humans from Kibish, Ethiopia." Nature **433**(7027): 733-6.

Meijers-Heijboer, H., J. Wijnen, et al. (2003). "The CHEK2 1100delC mutation identifies families with a hereditary breast and colorectal cancer phenotype." Am J Hum Genet **72**(5): 1308-14.

Michor, F., Y. Iwasa, et al. (2004). "Dynamics of cancer progression." Nat Rev Cancer **4**(3): 197-205.

Morgenthaler, S. and W. G. Thilly (2006). "A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: A cohort allelic sums test (CAST)." Mutat Res.

Muniappan, B. P. and W. G. Thilly (2002). "The DNA polymerase beta replication error spectrum in the adenomatous polyposis coli gene contains human colon tumor mutational hotspots." Cancer Res **62**(11): 3271-5.

Rees, J. L. (2000). "The melanocortin 1 receptor (MC1R): more than just red hair." Pigment Cell Res **13**(3): 135-40.

Rees, J. L. (2004). "The genetics of sun sensitivity in humans." Am J Hum Genet **75**(5): 739-51.

Sankaranarayanan, K. and R. Chakraborty (2000). "Ionizing radiation and genetic risks. XI. The doubling dose estimates from the mid-1950s to the present and the conceptual change to the use of human data on spontaneous mutation rates and mouse data on induced mutation rates for doubling dose calculations." Mutat Res **453**(2): 107-27.

Stein, W. D. (1991). "Analysis of cancer incidence data on the basis of multistage and clonal growth models." Adv Cancer Res **56**: 161-213.

Stein, W. D. and A. D. Stein (1990). "Testing and characterizing the two-stage model of carcinogenesis for a wide range of human cancers." J Theor Biol **145**(1): 95-122.

Strange, R. C. and A. A. Fryer (1999). "The glutathione S-transferases: influence of polymorphism on cancer susceptibility." IARC Sci Publ(148): 231-49.

Thilly, W. G. (2003). "Have environmental mutagens caused oncomutations in people?" Nat Genet **34**(3): 255-9.

Tomita-Mitchell, A., B. P. Muniappan, et al. (1998). "Single nucleotide polymorphism spectra in newborns and centenarians: identification of genes coding for risk of mortal disease." Gene **223**(1-2): 381-91.

Zheng, W., K. Khrapko, et al. (2006). "Origins of human mitochondrial point mutations as DNA polymerase gamma-mediated errors." Mutat Res **599**(1-2): 11-20.