Published in final edited form as:
Eur J Cardiovasc Nurs. 2012 December ; 11(4): 472–478. doi:10.1016/j.ejcnurse.2011.06.001.

Gene-Environment Interactions in Cardiovascular Disease

Elena Flowers, RN, MS, MAS [doctoral candidate],
School of Nursing at the University of California, San Francisco

Erika Sivarajan Froelicher, RN, PhD, FAAN [Professor], and
School of Nursing, Department of Physiologic Nursing, and School of Medicine, Department of Epidemiology and Biostatistics at the University of California, San Francisco

Bradley E. Aouizerat, PhD, MAS [Associate Professor]
School of Nursing, Department of Physiologic Nursing, and Institute for Human Genetics at the University of California, San Francisco

Abstract

Background—Historically, models to describe disease were exclusively nature-based or nurture-based. Current theoretical models for complex conditions such as cardiovascular disease acknowledge the importance of both biologic and non-biologic contributors to disease. A critical feature is the occurrence of interactions between numerous risk factors for disease. The interaction between genetic (i.e. biologic, nature) and environmental (i.e. non-biologic, nurture) causes of disease is an important mechanism for understanding both the etiology and public health impact of cardiovascular disease.

Objectives—The purpose of this paper is to describe theoretical underpinnings of gene-environment interactions, models of interaction, methods for studying gene-environment interactions, and the related concept of interactions between epigenetic mechanisms and the environment.

Discussion—Advances in methods for measurement of genetic predictors of disease have enabled an increasingly comprehensive understanding of the causes of disease. In order to fully describe the effects of genetic predictors of disease, it is necessary to place genetic predictors within the context of known environmental risk factors. The additive or multiplicative effect of the interaction between genetic and environmental risk factors is often greater than the contribution of either risk factor alone.

Keywords

Confounding; Environment and Public Health; Gene Expression; Genetic Variation; Methods; Phenotype

Corresponding Author: Elena Flowers, RN, MS, University of California, San Francisco, School of Nursing, Department of Physiologic Nursing, 2 Koret Way, #N611Y, San Francisco, CA 94143, Telephone: 001-415-606-9285, eflowers@ucsf.edu.

Introduction

Health promotion and disease prevention is a primary component of nursing practice. The cardiovascular disease pandemic necessitates that nursing practice focus on prevention and treatment of cardiovascular disease through risk assessment and stratification, and risk reduction through lifestyle and behavioral change. Although progress has been made in treating cardiovascular disease and its risk factors, the prevalence of this largely preventable condition worldwide remains unacceptably high.

Over the past three decades, the nature versus nurture debate over disease etiology has given way to more complex models that accommodate a larger number of causal pathways and allow for interactions between risk factors on many levels. These pathways include both individual and environmental level factors, and incorporate the influences of both nature and nurture into a single comprehensive model. The examination of known environmental causes of disease, coupled with the recent scientific advances allowing for rapid and affordable detection of genetic variation, are advancing our understanding of the individual and combined roles of genetic predisposition and environmental influences on cardiovascular disease. Interactions, including those occurring between an individual's psychosocial and physical environment and their unique genetic "makeup," are now accepted to have an important role in nearly all cardiovascular disease conditions and underlying biological processes.

Unaccounted gene-environment interactions are hypothesized to be a common alternative explanation for paradoxical findings about the etiology and prevention of cardiovascular disease (1). Environmental exposures can increase or decrease the effect of genetic predisposition, and genetic predisposition can modify the effects of the environment (2, 4). Lifestyle interventions to prevent and treat cardiovascular disease and its risk factors are in part unsuccessful because of incomplete understanding of the underlying biology of the disease and how manifestation of disease is moderated by the environment. Increasing the effectiveness of interventions may be possible through a more comprehensive understanding of the biology of cardiovascular disease, including interactions between genetic predisposition and environmental components such as lifestyle and behavioral variables. The aim of this paper is to describe theoretical underpinnings supporting the interaction between genetics and the environment and the onset and progression of cardiovascular disease.

Gene-Environment Interactions

Evidence supports the existence of gene-environment interactions for nearly every disease condition, including mental health disorders (5, 6), cardiovascular and metabolic disease (7-12), infectious disease (13, 14), and trauma and injury (15). While the field of quantitative genetics aims to identify specific gene loci responsible for disease, genetic epidemiology places what is known about genetic predictors of disease in the context of a population, searching for mechanisms of disease that include both genetic predisposition and environmental factors. The purposes of studying gene-environment interactions are to understand the complete etiology of a disease inclusive of multiple discrete and interacting pathways, and to determine the public health impact of individual factors within a specific population so that interventions can be designed to maximize health and minimize disease.

When an interaction between the genotype and environmental factors is present, this interaction is said to exert a main effect on the likelihood of developing disease. Additional marginal effects result from the independent contributions of the genotype and the environmental factors. Studies of genetic determinants of disease or environmental risk factors for disease are often designed to assess marginal effects only; however in many cases, the main effect of the interaction is hypothesized to be a far greater contributor to

disease than either marginal effect alone (1). Inconsistent and inconclusive findings from studies of the marginal effects of genetic determinants of disease are common. Failure to identify the presence of a gene-environment interaction with significant main effects is a likely alternative explanation for incongruous findings (1).

Models of Interaction

There are two commonly discussed types of interaction: statistical and biologic (3). Statistical interaction is strictly a mathematical phenomenon, in which the measured effects of one variable depend on the level of a second variable. By contrast, biologic interaction refers to the intersection of what are considered to be discrete pathways relevant to the maintenance of homeostasis or even the onset and progression of a physiologic condition. Both of these concepts are central to consideration of gene-environment interactions. Because the aim of studying gene-environment interactions is to discover new mechanisms of disease or describe the causes of deviation from expected expression of disease, biologic interaction is the very definition of gene-environment interaction. In order to quantify the presence of biologic interaction and make meaningful inferences about observations of interactions, incorporation of statistical interaction should be included during data analysis in order to accurately model the true underlying condition.

Gene-environment interactions can be either complementary or antagonistic (16). In the case of complementary interactions, both factors (i.e. environmental exposure and genetic predisposition) work in the same direction on disease risk. For example, an allele for the familial hypercholesterolemia (FH) gene might increase susceptibility to atherosclerosis by up-regulating the production of low-density lipoprotein cholesterol, and high intake of saturated fat may increase the likelihood of atherosclerosis by increasing low-density lipoprotein cholesterol levels. These two factors act in complement to increase an individual's risk of developing atherosclerosis. In contrast, antagonist interaction occurs when the direction of effect of two variables oppose each other. In the case of hypertension, carrying one of the known risk alleles for essential hypertension will increase an individual's likelihood of developing high blood pressure, whereas engaging in moderate physical activity most days of the week exerts the opposing effect of decreasing lifetime risk of hypertension.

Gene – Gene Interactions—Implicit in the definition of multi-factorial traits is that risk factors, be they environmental or genetic, can interact. The interaction of two or more genes is termed epistasis. Procedurally, studying gene-gene interactions is similar to studying gene-environment interactions. Although epistasis is related to gene-environment interactions and shares underlying principles, it is beyond the scope of this paper, and will not be discussed in further detail. An excellent review of epistasis can be found elsewhere (17).

Gene – Environment Correlations—A similar but distinct phenomenon from gene-environment interactions is gene-environment correlations. Correlations occur when a genetic marker is highly associated with, and possibly causal of, a behavioral characteristic or exposure that predisposes the outcome of interest (18). A thorough description of work describing gene-environment correlations, including clinical examples, can be found elsewhere (19).

Models of Disease Risk

The risk of disease in the presence of environmental and genetic risk factors can be depicted using a classic 2×2 table (Table 1). For simplicity, we consider both environmental and

genetic predictors and the disease outcome to be dichotomous. When genetic and environmental risk factors interact, four scenarios are possible (Table 2).

There are two primary mathematical models of risk that describe the relationship between multiple predictors of human disease: additive and multiplicative. Additive models assert that the effects of each predictor are summed in order to determine an individual's likelihood of disease. Predictors can still be complementary or antagonistic, however the net effect of all predictors is the sum of each predictor's effect on the disease outcome. In the case of gene-environment interactions, the effect of the environment depends on the genotype of an individual (2), and a statistical interaction is considered to be present when there is departure from the simple additive model (4). We can return to the atherosclerosis example from above and assume a simplified scenario in which there are only two predictors of disease: saturated fat intake and the FH risk allele (this example is chosen for simplicity only, and the following discussion does not encompass the entirety of biological mechanisms describing atherosclerosis risk). If no interaction is present, then an individual's relative risk of atherosclerosis (RR_a) is equal to the risk associated with their FH genotype (RR_g) plus the risk associated with their saturated fat intake (RR_e), which can be expressed mathematically as $RR_a = RR_g + RR_e$. Note that the mathematical term in this equation can be addition or subtraction, depending on whether the two effects are complementary or antagonistic. If an interaction is present, then a person who has a high saturated fat intake with no copies of the risk allele will have a relative risk for atherosclerosis equal to the harmful effect of saturated fat ($RR_a = RR_e$). However, for an individual who does carry the FH risk allele, the risk of saturated fat may no longer be significant, and atherosclerosis relative risk will be equal to the risk of the FH risk allele ($RR_a = RR_g$). In the presence of an interaction, the risk for atherosclerosis associated with saturated fat depends on FH genotype.

In contrast, multiplicative models assert that the effects of the genotype depend on the environment (2). In the presence of a biologic interaction, departure from the multiplicative model will be observed (4). If we apply a multiplicative model to the dichotomous atherosclerosis example for gene-environment interaction and again simplify to assume only two predictors, we can say that the effect of the FH risk allele depends on saturated fat intake. If no interaction is present, then an individual's relative risk of developing atherosclerosis is equal to the sum of the risk associated with their FH genotype and the risk associated with their saturated fat intake ($RR_a = RR_g + RR_e$). In the case of an interaction, the effect of the FH risk allele will depend on saturated fat intake. For an individual who consumes a large amount of saturated fat, carrying the FH risk allele would have no effect on the risk of atherosclerosis ($RR_a = RR_g \times RR_e$).

An alternative example for a continuous trait is blood pressure in individuals who experience a high level of autonomic stimulation during the workday (e.g. firefighters, air traffic controllers). An individual with no genetic predisposition to hypertension may have a normal blood pressure (e.g. 110/70 mmgHg) even in the presence of a high level of autonomic stress. In contrast, the blood pressure of an individual who is genetically predisposed to hypertension will depend on the level of autonomic work-related stress. In the absence of autonomic stress, the relative risk of hypertension (RR_h) is equal to the genetic risk for hypertension ($RR_h = RR_g$), which may result in a moderately elevated blood pressure (e.g. 145/90 mmHg). For the individual who is genetically predisposed to hypertension and exposed to a high level of work-related autonomic stress, the relative risk of hypertension could be multiplicative ($RR_h = RR_e \times RR_g$), and blood pressure may be significantly elevated beyond the expected effects of either risk factor alone (e.g. 180/100 mmHg).

The choice of which model to consider depends on two primary considerations. Although a comprehensive discussion of model selection is available elsewhere (20), each will be summarized briefly here. The first is the biologic relationship between the predictors. Commonly, when two predictors act on the same pathway, a simple additive model is assumed (20). Conversely, when two factors are thought to act on discrete physiologic pathways, the effect is often more than additive, and a multiplicative model is assumed (3, 20). In addition, the research question and aim of a study determine how a model is selected (2, 20). In the development of predictive models of disease for public health purposes and clinical decision-making, the underlying mechanisms of disease are not as important as the predictive capability of observable risk factors. In this scenario, an additive model that includes surrogate non-causal markers[1] will often suffice (3). For studies of disease etiology, the aim is to understand the mechanisms by which disease is ccurring, and multiplicative models may be necessary in order to correctly specify the relationship between predictors (3, 20). Often statistical modeling of both additive and multiplicative relationships is performed in order to determine the possible implications of model misspecification. In the case where additive and multiplicative models do not differ, the additive model is usually selected for simplicity and ease of interpretability.

Timing and Spectrum of Exposure—The occurrence of a gene-environment interaction is dynamic. As a result of changes in both gene expression and environmental exposure, interactions can occur at one time-point during the lifespan, periodically throughout life, or for longer durations. With regard to genetic exposure, a copy of the full genome is present in every cell of an organism; however not all genes are expressed at all times. Numerous genes are involved only during development and maturation, and once the adult stage is reached, expression of these genes is silenced. In contrast, some genes are only expressed in response to an environmental exposure or trigger. For example, a traumatic event such as a fracture will prompt localized expression of genes involved in inflammation and bone growth and remodeling that are not normally expressed in healthy osteocytes. Thus, the timing of an exposure can determine whether a gene-environment interaction occurs.

An organism's environment is typically in a state of constant flux. For humans, environmental factors can change throughout the day as well as over the course of a lifetime. Returning to the example of blood pressure and autonomic stimulation, work-related stress exerts immediate effects on hemodynamics and inflammatory processes, and individuals who are genetically predisposed to hypertension may be more likely to experience elevations in blood pressure during stressful work hours. In the case that this stress resolves at the completion of the workday, some physiologic phenomena, (e.g. hemodynamics) may to return to a non-stressed baseline state, whereas others, (e.g. inflammation-induced damage to the arterial wall) are permanent. Repeated exposure to the environmental stimuli can result in cumulative permanent physiologic changes.

Alternatively, some individuals are exposed to second-hand cigarette smoke during childhood, and subsequently this exposure is removed from the environment during adulthood. Others will grow up in a smoke free environment but then partner with a smoker of cigarettes during adulthood. The effects of secondhand smoke exposure may differ for the developing pediatric vascular endothelium compared to adult vascular endothelium, and irreversible damage could occur during a critical developmental period that will not affect the adult exposed to second-hand smoke in the same way.

[1]A known genetic locus or environmental measure that is not directly causal of disease but highly correlated with the causal region or exposure. Surrogate markers may be selected for reasons of cost or ease of measurement.

The spectrum of both genetic and environmental exposures can also determine whether an interaction occurs. Genetic dose is variable between individuals, and the individual's genotype will affect the level of exposure: some genes have dominant and recessive patterns of inheritance while others are co-dominant (21). For the dominant inheritance pattern, an individual will be affected if they carry just one copy of the risk allele, whereas the recessive pattern of inheritance requires two copies of a risk allele for an individual to be affected. In the case of co-dominance, differing alleles for a given gene are equally expressed. Similarly, differing doses of environmental exposures exert differing effects. In some cases, there is a threshold effect in which no adverse consequences are observed until a threshold level of exposure is reached. In other cases, environmental exposure is continuous, and increasing doses will exert a linear increase in harmful or protective effects. Thus, quantifying the dose of both genetic and environmental exposure is important in order to detect the presence of gene-environment nteractions.

Methodological Issues

Study Design—There are three primary epidemiologic study designs that are appropriate for studies of this nature: cohort, cross-sectional, and case-only (Table 3). The selection of study design depends on the aim of the study (i.e. investigations of disease etiology compared to assessing the impact of environmental exposure in the context of genetic predisposition), what is known about genetic determinants of the outcome of interest, the prevalence and/or incidence of the disease, and the resources required to perform the study. Prospective population-based cohort studies are the gold standard for gene-environment interaction studies, offering substantially decreased risk of measurement error and subsequent bias, however they are generally extremely resource-intensive (22). The classic case-control design can also be used for studies of gene-environment interactions, but are more susceptible to confounding compared to other study designs. For genetic association studies and gene-environment interaction studies, controls can be selected from among family members, which typically decreases the potential for confounding due to population stratification and can increase the power to detect an effect, but can result in the detection relationships that are less relevant at the population level (22, 23). A modification of the classic case-control design that is well suited to studies of gene-environment interactions is the case-only design. In case-only studies, inclusion criteria limit sample selection to individuals with the outcome of interest (24, 25). A limitation of the case-only design is that a priori knowledge of causal regions of the genome is required. A complete discussion of study designs appropriate for nvestigation of gene-environment interactions is available elsewhere (22-25).

Measurement—For studies of gene-environment interactions, the most common sources of measurement bias arise from misclassification of both environmental and genetic exposure. As discussed above, environmental exposures can vary over the course of a lifetime, which poses a challenge to accurate measurement (i.e., recall bias, biomarker stability). In some instances, individuals with a disease may be more likely to recall and/or report exposure because of their disease, resulting in differential misclassification and thus biased estimates of association and interaction (26, 27). Genotype is also subject to misclassification. When stringent quality control standards are implemented during laboratory analysis, the likelihood of misclassification is diminished for studies of relatively rare disorders (e.g., phenylketonuria) for which functional polymorphisms in single genes can be quantified (2, 4). However for disease conditions that are multifactorial, the principle of linkage disequilibrium is often exploited in order to identify regions of the genome that are associated with the outcome of interest. Genotype measurement by linkage disequilibrium is efficient and decreases genotyping costs. The principle of linkage disequilibrium is dependent on population substructure, or sub-populations determined by

geographic ancestry that have shared common allele frequencies. The correlation between a region of the genome and an outcome is a group-averaged statistic for a given sub-population, which can result in misclassification of the genotype for an individual if they differ from the group norm (2, 4). When cases and controls are sampled from the same study base (sub-population), this type of misclassification is likely to occur with equal frequency, which will not result in biased estimates of association, but may decrease the likelihood of detection of a true interaction (26, 27). The least accurate method of quantifying genetic exposure is family history, which is also subject to misclassification as the actual genotype of the individual in unknown, and recall of family history can be incorrect (2, 4). Similar to recall of environmental exposure, accuracy of recall of family history may be differential between cases and controls, resulting in biased measures of association and interaction (26, 27).

Epigenetics and Gene-Environment Interactions

Epigenetics is the study of mechanisms that result in changes to the phenotype or appearance of an organism that do not result from underlying changes to the genetic sequence. Commonly studied epigenetic mechanisms can affect the DNA, as in the case of DNA methylation[2] or histone acetylation[3], or post-transcriptional modifications such as microRNA[4] regulation of gene expression. Traditional gene-environment interactions occur when the protein encoded by a particular gene interacts with an environmental exposure. Similarly, expression of a gene can be affected by exposure to an environmental factor, resulting in silencing or increased or decreased expression of a gene that may persist due to lasting but potentially reversible changes. For example, exposure to tobacco smoke may result in up-regulation of genes associated with platelet activation, and increased expression of these genes will persist while tobacco smoke is present at regular intervals. However, when this stimulus is removed for an extended period of time, epigenetic up-regulation of inflammatory genes will cease, and platelet activity will return to a normal, healthy physiologic state. In contrast, concomitant changes in genes mediating inflammation may not return to baseline due to changes in DNA methylation or histone acetylation. With regard to platelet activation, the deleterious effects of exposure to tobacco smoke are reversible; however for inflammation, there can be long-lasting or permanent alterations in genes expression, resulting in cumulative physiologic damage over the lifespan. The same principles that apply to gene-environment interactions, including timing and spectrum of environmental exposure, apply to interactions between epigenetic mechanisms and environmental exposures.

Conclusion

Humans exist within a dynamic environment and are subject to factors exerting effects on health outcomes on a number of levels. The current paradigm for understanding causes of cardiovascular disease within a complex system suggests that these conditions are rarely, if ever, the result of a single causal factor. The conceptual frameworks underlying cardiovascular disease postulate that these conditions occur in the presence of numerous genetic and environmental risk factors, that interactions between these factors occur on several levels, and that these interactions account for significant primary effects on the likelihood of disease occurrence (1, 28, 29). In some cases, an interaction between individual gene loci and environmental exposure is believed to have a greater effect than the individual marginal effects of either factor alone (1). Failing to account for the presence of a

[2]addition of a methyl group to DNA that silences expression of some genes
[3]changes to the protein-DNA folding structure that prevent transcription of select regions of the genome
[4]small interfering RNA molecule that represses transcription of a messenger RNA sequence preventing polypeptide formation

gene-environment interaction can result in incorrect conclusions about the etiology of cardiovascular disease, and is often attributed as a cause of incongruous study findings.

Gene-environment interactions have important implications for both nurse-clinicians and nurse-researchers. Worldwide, a current emphasis of nursing practice is to identify and treat individuals suffering from cardiovascular risk factors in order to prevent the onset and sequelae of cardiovascular disease. Throughout the twentieth century, nursing practice has focused largely on behavioral interventions and modification of the environment. The genomic era of healthcare both facilitates and necessitates that nurses also understand genetic predisposition for cardiovascular disease, and most importantly, place genetic predisposition within the context of an individual's environment.

Acknowledgments

Funding for Elena Flowers' doctoral studies was provided by NIH/NIGMS R25 GM56847 NIH/NCRR/OD UCSF-CTSI Grant TL1 RR024129

Funding for Bradley E. Aouizerat was provided by NIH/NCRR/OD UCSF-CTSI Grant Number KL2 RR024130

References

1. Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet. 2005 Apr; 6(4):287–98. [PubMed: 15803198]
2. Ottman R. Gene-environment interaction: definitions and study designs. Prev Med. 1996 Nov-Dec; 25(6):764–70. [PubMed: 8936580]
3. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. Am J Epidemiol. 1980 Oct; 112(4):467–70. [PubMed: 7424895]
4. Yang Q, Khoury MJ. Evolving methods in genetic epidemiology. III. Gene-environment interaction in epidemiologic research. Epidemiol Rev. 1997; 19(1):33–43. [PubMed: 9360900]
5. van der Zwaluw CS, Engels RC. Gene-environment interactions and alcohol use and dependence: current status and future challenges. Addiction. 2009 Jun; 104(6):907–14. [PubMed: 19466917]
6. van Os J, Rutten BP, Poulton R. Gene-environment interactions in schizophrenia: review of epidemiological findings and future directions. Schizophr Bull. 2008 Nov; 34(6):1066–82. [PubMed: 18791076]
7. Andreasen CH, Andersen G. Gene-environment interactions and obesity--further aspects of genomewide association studies. Nutrition. 2009 Oct; 25(10):998–1003. [PubMed: 19596186]
8. Hirvonen A. Gene-environment interactions in chronic pulmonary diseases. Mutat Res. 2009 Jul 10; 667(1-2):132–41. [PubMed: 19563928]
9. Andreassi MG. Metabolic syndrome, diabetes and atherosclerosis: influence of gene-environment interaction. Mutat Res. 2009 Jul 10; 667(1-2):35–43. [PubMed: 19028510]
10. Ordovas JM, Shen J. Gene-environment interactions and susceptibility to metabolic syndrome and other chronic diseases. J Periodontol. 2008 Aug; 79(8 Suppl):1508–13. [PubMed: 18673004]
11. Qi L, Hu FB, Hu G. Genes, environment, and interactions in prevention of type 2 diabetes: a focus on physical activity and lifestyle changes. Curr Mol Med. 2008 Sep; 8(6):519–32. [PubMed: 18781959]
12. Visvikis-Siest S, Siest G. The STANISLAS Cohort: a 10-year follow-up of supposed healthy families. Gene-environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. Clin Chem Lab Med. 2008; 46(6):733–47. [PubMed: 18601594]
13. Alcais A, Abel L, Casanova JL. Human genetics of infectious diseases: between proof of principle and paradigm. J Clin Invest. 2009 Sep; 119(9):2506–14. [PubMed: 19729848]
14. Hill AV. The genomics and genetics of human infectious disease susceptibility. Annu Rev Genomics Hum Genet. 2001; 2:373–400. [PubMed: 11701655]

15. Reddy AJ, Kleeberger SR. Genetic polymorphisms associated with acute lung injury. Pharmacogenomics. 2009 Sep; 10(9):1527–39. [PubMed: 19761373]
16. Rothman KJ. The estimation of synergy or antagonism. Am J Epidemiol. 1976 May; 103(5):506–11. [PubMed: 1274952]
17. Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum Mol Genet. 2002 Oct 1; 11(20):2463–8. [PubMed: 12351582]
18. Rutter M. Gene-environment interdependence. Dev Sci. 2007 Jan; 10(1):12–8. [PubMed: 17181693]
19. Rutter M. Commentary: causal processes leading to antisocial behavior. Dev Psychol. 2003 Mar; 39(2):372–8. [PubMed: 12661891]
20. Greenberg RS, Kleinbaum DG. Mathematical modeling strategies for the analysis of epidemiologic research. Annu Rev Public Health. 1985; 6:223–45. [PubMed: 3994813]
21. Jorde, LB. Medical genetics. In: St Louis, Mo, editor. New. London: Mosby; 2003.
22. Dempfle A, Scherag A, Hein R, Beckmann L, Chang-Claude J, Schafer H. Gene-environment interactions for complex traits: definitions, methodological requirements and challenges. Eur J Hum Genet. 2008 Oct; 16(10):1164–72. [PubMed: 18523454]
23. Witte JS, Gauderman WJ, Thomas DC. Asymptotic bias and efficiency in case-control studies of candidate genes and gene-environment interactions: basic family designs. Am J Epidemiol. 1999 Apr 15; 149(8):693–705. [PubMed: 10206618]
24. Andrieu N, Goldstein AM. Epidemiologic and genetic approaches in the study of gene-environment interaction: an overview of available methods. Epidemiol Rev. 1998; 20(2):137–47. [PubMed: 9919434]
25. Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: case-control studies with no controls! Am J Epidemiol. 1996 Aug 1; 144(3):207–13. [PubMed: 8686689]
26. Hernan MA, Cole SR. Invited Commentary: Causal diagrams and measurement bias. Am J Epidemiol. 2009 Oct 15; 170(8):959–62. discussion 63-4. [PubMed: 19755635]
27. Szklo, M.; Nieto, FJ. Epidemiology : beyond the basics. 2. Sudbury, MA: Jones and Bartlett Publishers; 2007.
28. Bronfenbrenner, U. The ecology of human development : experiments by nature and design. Cambridge, Mass: Harvard University Press; 1979.
29. Hancock T. The mandala of health: a model of the human ecosystem. Fam Community Health. 1985 Nov; 8(3):1–10. [PubMed: 10274086]

Table 1

Effects of Genetic and Environmental Risk Factors on Relative Risk of Disease (RR) [a, b]

	High risk genotype present	High risk genotype absent
Environmental risk factor present	$RR_{ge} \gg 1$[c]	$RR_e > 1$[c]
Environmental risk factor absent	$RR_g > 1$[c]	$RR = 1$[c]

[a] Assumes a simple scenario in which the exposure and the genotype are dichotomous and result in a synergistic effect of both factors to increase the risk of disease

[b] Adapted from Ottman, R. (1996). Gene-environment interaction: definitions and study designs. *Prev Med, 25*(6), 764-770.

A relative risk of one is the level of risk for the general population.

Table 2

Possible Interaction Effects Between Multiple Risk Factors on Relative Risk of Disease (RR)[a]

	Model of Disease Risk	
	Additive	**Multiplicative**
No interaction	$RR_{ge} = RR_g + RR_e - 1$	$RR_{ge} = RR_g \times RR_e - 1$
Complementary interaction	$RR_{ge} > RR_g + RR_e - 1$	$RR_{ge} > RR_g \times RR_e - 1$
Antagonistic interaction	$RR_{ge} < RR_g + RR_e - 1$	$RR_{ge} < RR_g \times RR_e - 1$

RR_g = relative risk of disease when genetic risk factor is present

RR_e = relative risk of disease when environmental risk factor is present

RR_{ge} = relative risk of disease when both genetic and environmental risk factors are present

[a]Adapted from Ottman, R. (1996). Gene-environment interaction: definitions and study designs. *Prev Med, 25*(6), 764-770.

Table 3

Gene-environment interaction study designs

Study Design	Features	Pros	Cons
Cohort	• Selection of sample occurs before onset of disease (case status unknown at outset of study) • Longitudinal follow up of entire sample	• A priori knowledge of causal regions of the genome not required • Can infer causal relationships between exposure and outcome • Accurate measurement timing of environmental exposure • Decreases likelihood of survivor and recall biases	• Time-consuming (often requires years of follow-up time) • Expensive • Require extremely large sample sizes to study rare or heterogeneous conditions
Case-Control	• Purposeful sampling of individuals with outcome of interest and controls typically matched on pre-specified criteria • Can be cross-sectional or longitudinal	• Increased power to study rare conditions • A priori knowledge of causal regions of the genome not required • Less expensive and time-consuming than cohort design	• Cannot make causal inferences with cross-sectional time-frame • Difficult to determine selection criteria for appropriate controls • Possible confounding due to population stratification
Case-Only	• Sample consists of only individuals known to have outcome of interest • Case status determined by genotype (presence of known genetic determinants of disease)	• Highly useful for studies of gene-environment interaction • Smaller sample size often possible • Eliminates problem of appropriate control selection in case-control design • Less expensive and time consuming than cohort design	• Requires knowledge of causal regions of the genome • Does not allow for estimation of the main effects of environmental and genetic exposure • Cannot make causal inferences with cross-sectional timeframe