Models of Interaction between Metabolic Genes and Environmental Exposure in Cancer Susceptibility

Emanuela Taïoli,1,2 Carlo Zacchetti,3 and Seymour Garte4

1Nelson Institute of Environmental Medicine and Kaplan Cancer Center, New York University Medical Center, New York, NY 10010 USA; 2Ospedale Maggiore Policlinico IRCCS, University of Milano, Milano Italy; 3Clinica Del Lavoro, Milano, Italy

Polymorphic metabolic genes that confer enhanced genetic susceptibility to the carcinogenic effects of certain environmental carcinogens act according to a type 2 interaction between genetic and environmental risk factors. This type of interaction, for which the gene has no effect on disease outcome by itself but only modifies the risk associated with exposure, must be treated differently from other types of gene–environment interaction. We present a method to analyze different dose effects often seen in studies involving these genes. We define a low exposure–gene effect, when a greater degree of gene environment interaction appears at lower doses of exposure (the interaction follows an inverse dose function), and a converse high exposure–gene effect, when the interaction increases as a function of dose. Using a standard logistic regression model, we define a new term, $\alpha$, that can be determined as a function of exposure dose in order to analyze epidemiological studies for the type of gene–exposure effect. These models are illustrated by the use of hypothetical case–control data as well as examples from the literature. Key words: biomarkers, case-control, epidemiology, logistic regression, polymorphisms. Environ Health Perspect 106:67–70 (1998). [Online 21 January 1998] http://ehpnet1.niehs.nih.gov/docs/1998/106p67-70taiolil/abstract.html

Both environmental and genetic factors have been identified as playing roles in the development of many chronic diseases including cancer (1–6). Despite the importance of environmental exposure in human cancer, the evidence for some form of genetic influence on almost all cancer etiology is very strong (7,8). Human subjects show not only a wide variation of genetic polymorphisms but also an extremely broad range of phenotypic responses to environmental stimuli. In this scenario, it is important to understand the type of interactions that occur between genetic and environmental factors. With the development of new biomarkers of genetic cancer susceptibility, new paradigms for the classical terms of interaction and confounding that are appropriate for the several different mechanisms of gene–environment interaction are needed (9).

A number of genes involved in metabolism of carcinogens have been shown to play a role in cancer risk. In most cases, the putative biochemical mechanism by which such genetic factors exert their effects is fairly straightforward and is related to the actual dose of the active carcinogenic metabolite that reaches the genome in the target cell. By definition, these genes play a role in cancer risk only in the context of interaction with the environment because the substrates of their gene products are xenobiotic chemicals or their metabolites. This follows the type 2 form of gene–environment interaction (GEI) as previously described by Khoury (10,11) and Ottman (12). In this model of interaction, the cancer is caused by exposure to an environmental agent. If there is no exposure, the presence or absence of the genetic risk factor is irrelevant for disease causation. This model of GEI is the most suitable to explain human carcinogenesis related to metabolic susceptibility genes such as cytochrome P450 1A1 (CYPIA1), N-acetyltransferase (NAT), glutathione-Stransferase (GST), etc.

When the dose of environmental exposure (such as smoking) is analysed with respect to genotype of a metabolic susceptibility gene, two apparently divergent patterns are seen. The first instance could be termed the low exposure–gene (LEG) effect. This is seen when a decreasing degree of interaction occurs as a function of exposure. For example, in a case–control study, the proportion of cases with the genetic risk factor (GRF) or polymorphism would have lower exposure than the proportion of cases without the GRF (13,14). If the endpoint is not cancer but, for example, some marker of exposure such as adducts, subjects with polymorphisms should tend to have higher relative levels of adducts at lower doses of exposure; at high exposures, no difference in endpoint would be observed between those with and without the GRF (15). The LEG effect is very often observed with cases of type 2 GEI such as cancer susceptibility metabolic genes and has been discussed in the context of these genes (16–18).

A high exposure–gene (HEG) effect is observed when there is an increased degree of interaction as a function of exposure dose. In a case–control study, cases with the GRF would have higher environmental exposure doses than cases without the GRF; in other words, the higher the dose, the greater the effect of having the GRF on any other endpoint (such as disease, adducts, etc.). This high exposure–gene effect has been seen with GSTM1 and lung cancer (19,20). However, a study of GSTM1 and asbestososis (21) showed a LEG effect, suggesting that this phenomenon is not simply gene specific, but must be related to the mechanism of action of the gene product leading to the endpoint being measured. It is important to understand that, for both types of dose-related gene–exposure interaction defined here, having the GRF is not protective, and persons with the GRF are at either equal or higher risk than those without the GRF. The difference between the HEG and LEG effects is whether the effect of increasing exposure dose magnifies or diminishes the relative risk associated with the GRF.

We present a method to identify these two patterns of dose effects and illustrate this method with data from a hypothetical case–control study of a metabolic gene polymorphism showing type 2 GEI, as well as with examples from the literature. The method is also able to distinguish between these effects and the presence of a true protective effect of the genetic variant.

Results and Discussion

A common way to describe interactions between the effects of an environmental agent and a genetic risk factor is to use both terms in a multiple regression model and to include a term that multiplies the GRF by the environmental agent. The coefficient of this interactive term then determines whether interaction is present:

$$G(Y) = a + b_g E + b_{gE} E G,$$  \hspace{1cm} (1)

where $Y$ is the odds of disease, $a$ is a constant, $E$ is the environmental exposure, $G$ is the GRF, and $E G$ is the interaction term. The coefficients $b_g$, $b_{gE}$, and $b_{gE}$ are determined by regression analyses of the data.

Address correspondence to E. Taïoli, Epidemiology Unit, Ospedale Maggiore Policlinico, Via F. Sforza 28, 20122 Milano, Italy.

This work was supported by NIH grants ES00260, ES08363, and ES66/CA103919 and an IRCCS Institutional Grant. We thank Pier Alberto Berzatti for his support and encouragement. Received 30 May 1997; accepted 6 October 1997.
regression analysis using an appropriate computer program. If we assume that in the absence of environmental exposure the presence of the GRF by itself has no effect on disease outcome (which is the definition of type 2 GEI), then $b_e$ in the regression model of Equation 1 is defined as 0, which leads to Equation 2:

$$G(Y) = a + b_e E + b_{eg} EG,$$

(2)

which can also be written

$$G(Y) = a + (b_e + b_{eg} G)E = a + b^* E.$$  (3)

This expression corresponds to the assumption that the risk of disease is due to the action only of the environmental exposure and the only effect of the GRF is to modify the coefficient of the exposure term. Now we can say that

$$b^* = b_e + b_{eg} G = b_e (1 + \alpha G)$$  (4)

where $\alpha = b_{eg}/b_e$.

In a case–control study, $\alpha$ is the ratio of two log odds ratios (ORs).

For example, let us assume that the effect of GRF such as the isoleucine (Ile) to valine (Val) polymorphism in the CYP1A1 gene is to increase the enzymatic activity of the gene product, as has been shown (22–25). The result of having this GRF is an increased level of metabolism, presumably leading to an increased concentration of the ultimate carcinogen, given a particular exposure dose. While this scenario may represent an oversimplification, it can be seen that the effect of the GRF is to quantitatively modify the effect of the exposure term. This would be reflected in a value greater than 0 for the term $\alpha$. If $\alpha$ is negative, the genetic factor should be protective. If the GRF has no effect on the exposure (e.g., if the exposure is to an agent that is not a substrate for the gene product), then $b_{eg}$ and $\alpha = 0$. If the GRF is absent, $G = 0$. In either case, the risk is a function of exposure only, with no contribution from the gene.

We can rewrite the regression model of Equation 2 if there are data for the effects of multiple ($n$) levels of exposure (doses):

$$G(Y) = a + b_{e1} E_1 + b_{e2} E_2 + \ldots + b_{en} E_n + b_{eg1} GE_1 + b_{eg2} GE_2 + \ldots + b_{egan} GE_n,$$  (5)

where $b_{e1} = b_e$ for exposure level $E_1$, $b_{eg1} = b_{eg}$ for exposure level $E_1$, and $G$ is the GRF. Using this notation, $E_i$ is the dose and $b_i$ is defined as 0 and does not appear. For each dose level (apart from the reference) from Equation 4:

$$\alpha_i = b_{eg}/b_{ei}.$$  (6)

If values of $\alpha$ are plotted against dose, several outcomes are possible. If the slope of this plot is positive, the gene–environment interaction follows a HEG effect; if the slope is negative, a LEG effect is operative. If $\alpha < 0$ and $b_i > 0$ at any particular dose level, the genetic factor is protective at that level. Such a scenario, whereby a particular genetic polymorphism may be a risk factor at one level of exposure but protective at a different level, is possible given the highly complex web of interconnecting metabolic pathways that usually operate in carcinogenic mechanisms.

The risk of disease or any other endpoint is always a function of dose. Figure 1 illustrates a hypothetical dose–response relationship for a type 2 GEI. At low–exposure levels (area A in the figure), the curves with and without the GRF diverge as a function of increasing dose, showing an increased interaction (HEG effect). On the other hand, at high doses, near a putative saturation value (Area C), the curves converge toward saturation, showing a decreased interaction (LEG effect). The area in between, Area B, exhibits a mixed pattern. This model may or may not explain the observed LEG and HEG effects in type 2 GEI.

Since $\alpha$ is dependent on two odds ratios (or rate ratios) and each ratio is a function of the risk of disease at a particular dose, we can assume that $\alpha$ is some function of dose. The particular function is likely to vary for different specific chemical exposures. If the function of $\alpha$ with respect to dose is known, we can determine the form of the exposure–gene effect by the sign of the first derivative of $\alpha$ with respect to exposure, $d\alpha/dE$. Although several feasible dose–response models could be used to determine the function of $\alpha$ with respect to exposure level, there is no clear reason to choose any particular such model, given the current understanding. Therefore, rather than attempt to more precisely define $d\alpha/dE$, we have used hypothetical case–control studies representative of the various scenarios presented above.

**Hypothetical data.** We have created a dataset using the EGRET package (Serc and Cytel, Seattle, WA), including 5,000 cases and 5,000 controls, in which the frequency of the genetic polymorphism among the controls was arbitrarily defined as 20% and the frequency of the levels of exposure (no exposure, low, medium, high) was arbitrarily set to 90%, 5%, 3%, and 2%, respectively.

### Table 1. Hypothetical high exposure–gene effect

| Gene | Environmental exposure | OR | $b_e$ | $b_{eg}$ | $\alpha$ |
|------|------------------------|----|-------|---------|---------|
| -    | None                   | 1.0| 0     | 0       |         |
| -    | + (low)                | 2.471 | 0.995 | 1.222 |       |
| -    | + (medium)             | 3.394 | 1.050 | 1.050 |       |
| -    | + (high)               | 4.505 | 1.050 | 1.050 |       |
| +    | None                   | 0.894 | 0     | 0.112 |       |
| +    | + (low)                | 2.709 | 0.905 | 0.997 | 1.102 |
| +    | + (medium)             | 6.675 | 1.222 | 1.898 | 1.553 |
| +    | + (high)               | 13.450 | 1.505 | 2.599 | 1.727 |

**Abbreviations:** OR: odds ratio. \(\cdot\): absent; \(+\): present. $\alpha = b_{eg}/b_e$.

### Table 2. Hypothetical low exposure–gene effect

| Gene | Environmental exposure | OR | $b_e$ | $b_{eg}$ | $\alpha$ |
|------|------------------------|----|-------|---------|---------|
| -    | None                   | 1.0| 0     | 0       |         |
| -    | + (low)                | 2.029 | 0.708 | 1.147 |       |
| -    | + (medium)             | 3.148 | 1.408 | 1.408 |       |
| -    | + (high)               | 4.086 | 1.408 | 1.408 |       |
| +    | None                   | 1.043 | 0     | 0.043 |       |
| +    | + (low)                | 4.402 | 0.708 | 1.482 | 2.094 |
| +    | + (medium)             | 5.289 | 1.147 | 1.866 | 1.453 |
| +    | + (high)               | 5.936 | 1.147 | 1.866 | 1.453 |

**Abbreviations:** OR: odds ratio. \(\cdot\): absent; \(+\): present. $\alpha = b_{eg}/b_e$. 

**Figure 1.** Relationship between levels of exposure and endpoint, showing three regions with different exposure–gene effects. A, high exposure–gene effect; B, mixed effect; C, low exposure–gene effect; dotted line, genetic risk factor present; solid line, genetic risk factor absent.
To show the HEG effect, we set the ORs of the three levels of exposure as 2.0, 3.0, and 4.0, the OR for the presence of the genetic polymorphism as 1.0 (as from the definition of GEI type 2), and the OR for the presence of the gene and the three levels of environmental exposure to 3.0, 6.0, and 12.0. To show the LEG response, we kept the same ORs of the three levels of exposure and for the presence of the genetic polymorphism, while the OR for the gene and the three levels of environmental exposure were set to 4.0, 5.0, and 6.0.

The results are presented in Tables 1 and 2. In Table 1 the value of \( \alpha \) rose from 1.1 to 1.7, confirming a HEG effect; in Table 2, \( \alpha \) decreased from 2.1 to 1.3, as hypothesized in the LEG effect.

**Examples of type 2 GEI analyses.**

Both types of exposure–gene effects have been observed in several studies of genetic susceptibility genes that include information on exposure dose. Kihara et al. (20) illustrate a high exposure–gene effect. Table 3 shows the ORs for each category of smoking exposure and genotype. The first observation is that when the GRF is present but there is no exposure, the OR is equal to the reference (absence of both gene and exposure); this defines a type 2 interaction. As the exposure level increases, the risk of disease increases; the increase is higher when the GRF is present for each category of exposure level. Table 4 shows the coefficients \( (b_1, b_2) \) obtained from the multiple logistic regression model using the SAS statistical package Genmod (SAS Institute, Cary, NC). Also shown in Table 4 are the values for \( \alpha \), the interaction term, which increases directly as a function of dose. Thus, in this case there is a HEG effect. Similar results can be obtained using the data from other sources for this gene and smoking-related lung cancer (19).

An example of a LEG effect is seen for the CYP1A1 Ile to Val polymorphism in exon 7 of the gene as a GRF for smoking-induced lung cancer. The data from Nakachi et al. (26), shown in Tables 3 and 4, show that although the OR for cancer increases for both genotypes as a function of dose, the ratio of the risks decreases at higher doses.

The decrease of \( \alpha \) with increasing exposures illustrates this. We have also observed a low exposure–gene effect for the association of the African-American-specific polymorphism in CYP1A1 with lung adenocarcinoma in smokers (14), and other groups have reported similar findings using adducts as an endpoint for NAT2 and for CYP1A2 (Landi et al., personal communication; (15)).

Metabolic genes that modify cancer susceptibility play no role in carcinogenesis in the absence of a relevant carcinogenic exposure, assuming that endogenous substrates are not involved. If they are, then the term exposure must be modified to include such agents. In the analysis of this GEI, defined as type 2, it is important to consider the exposure dose and two different forms of exposure–gene interaction. We have shown that the two forms of dose effect, the low exposure–gene effect and the high exposure–gene effect, may be analyzed and distinguished from each other and from other types of effects (such as protection). In our example, we have not addressed the question of whether multiplicative versus additive models of interaction should be used (27). For areas A and C of Figure 1, a multiplicative model was the hypothesis underly- ing our analysis, and we have observed that use of an additive model has an effect only on the magnitude of the coefficients and \( \alpha \) values but not on the direction of the dose-dependent effect of the gene (not shown). For Area B, this is not the case.

While at this point the biological mechanisms responsible for the two types of exposure effects are not known, it is possible to speculate that these effects may simply be a reflection of the shape of the dose–response curves for individuals with and without the GRF. If we assume that the effect of a genetic susceptibility factor is to increase the carcinogenic response at any particular dose (e.g., by causing increased enzymatic activity or by altering the metabolic profile of an agent), the dose–response curve will be shifted to the left. An important assumption is that the GRF has no effect on the maximal response. At low dose levels, the presence of the GRF will lead to an increase in the slope of the dose–response curve, and subjects with the GRF should respond more to higher doses of environmental agents than subjects without the GRF, leading to the observation of a HEG effect. It may be safely assumed that for any toxicological endpoint (including carcinogenesis) a saturating dose must always exist at which no further effect can be seen with increasing dose. At doses close to this maximum saturating level, the dose–response curve for subjects with the GRF will exhibit a decreased slope so that, although the overall response in GRF positive individuals is higher than in GRF negatives, the increase of the effect with increasing dose is lower in the positives than in the negatives. Therefore, subjects with the GRF may exhibit disease (or high adduct levels, etc.) at lower doses of environmental agent than those without the GRF (the LEG effect). This assumes that both exposure and the GRF have a positive influence on the endpoint. For situations in which either the exposure (such as a therapeutic or preventive agent) or the gene have a negative effect on the endpoint, the reverse would be observed. According to this model, genes such as CYP1A1 show low exposure–gene effect with respect to lung cancer because the doses of carcinogen present in cigarette smoke are so high.

A low exposure–gene effect implies that for GRF positive individuals, even a low carcinogen dose is highly risky; people carrying the polymorphism are at higher risk of cancer in comparison to members of the general population who are exposed to a carcinogen such as tobacco smoke. Only complete smoking cessation, as well as the avoidance of other relevant exposures, can lead to cancer prevention in the susceptible group.

This analytical approach may be used to determine whether a HEG or LEG effect is operative. Type 2 GEI is the correct model for a particular study if the odds ratio for the unexposed group with the GRF is close to 1. However, when the data are of insufficient statistical power to allow for an accurate determination of this odds ratio, it may be necessary to use the type 2 model of LEG effect to estimate the maximum effect that could be seen.
ratio, an alternative is to apply knowledge regarding the mechanism of action of the gene related to the exposure to decide whether a type 2 interaction is logical within a mechanistic context. Comparison of α values between studies and meta-analysis of α for specific gene–exposure combinations may prove valuable in the future.

The detection of cancer genetic susceptibility has profound positive public health implications for cancer prevention. Detailed study of the interactions of these genes with environmental carcinogens holds the promise of allowing the definition of subsets of individuals of varying sensitivity and responsiveness. While the entire issue of genetic susceptibility differences among people has important ethical, legal, and political issues, increased knowledge in this area (such as the specific form of exposure–gene effect discussed here) will provide benefits in helping to resolve these issues.

REFERENCES

1. Hirayama T. Family history in childhood malignancies with special reference to genetic–environment interaction. In: Genetic Epidemiology of Cancer (Lynch HT, Hirayama T, eds). Boca Raton FL: CRC Press, 1989:111–117.
2. Hegele RA. Gene–environment interactions in atherosclerosis. Mol Cell Biochem 133(2):177–186 (1992).
3. Hwang SJ, Beatty TH, Panny SR, Street NA, Joseph JM, Gordon S, McIntosh I, Franconomo CA. Association study of transforming growth factor alpha (TGF alpha) TaqI polymorphism and oral clefts: indication of gene–environment interaction in a population-based sample of infants with birth defects. Am J Epidemiol 141(7):629–636 (1995).
4. Brennan P, Silman AJ. An investigation of gene–environment interaction in the etiology of rheumatoid arthritis. Am J Epidemiol 140(5):453–460 (1994).
5. Hayden MR, Liu MS, Ma Y. Gene environment interaction and plasma triglyceride levels: the crucial role of apolipoprotein lipases. Clin Genet 40(6 Spec No):15–18 (1991).
6. Tietz L, Abel L, Rakotovao R. Effect of ignoring genotype–environment interaction on segregation analysis of quantitative traits. Genet Epidemiol 10(6):581–586 (1993).
7. Hirvonen A. Genetic factors in individual responses to environmental exposures. J Occup Environ Med 37(1):37–43 (1995).
8. Harris CC. Interindividual variation among humans in carcinogen metabolism, DNA adduct formation and DNA repair. Carcinogenesis 10(3):1563–1566 (1989).
9. Hulka BS. Epidemiological studies using biological markers: issues for epidemiologists. Cancer Epidemiol Biomarkers Prev 1:13–19 (1991).
10. Khoury MJ, James LM. Population and familial relative risks of disease associated with environmental factors in the presence of gene–environment interaction. Am J Epidemiol 131(11):1241–1250 (1990).
11. Khoury MJ, Wagener DK. Epidemiological evaluation of the use of genetics to improve the predictive value of disease risk factors. Am J Hum Genet 56(4):835–844 (1995).
12. Ottman R. An epidemiologic approach to gene–environment interaction. Genet Epidemiol 7:177–185 (1990).
13. Nakachi K, Imai K, Hayashi S, Kawajiri K. Polymorphisms of the CYPIA1 and glutathione S-transferase genes associated with susceptibility to lung cancer in relation to cigarette dose in a Japanese population. Cancer Res 53:2394–2399 (1993).
14. Taioli E, Crofts F, Demopoulos R, Trachman J, Toniolo P, Garte SJ. An African-American specific CYPIA1 polymorphism is associated with adenocarcinoma of the lung. Cancer Res 55:472–473 (1995).
15. Vines P, Bartsch H, Caporaso N, Harrington A, Kadiu F, Landi MT, Malaveille C, Shields PG, Skipper P, Talaska G. Genetically based N-acetylation metabolism polymorphism and low level environmental exposure to carcinogens. Nature 369:154–156 (1994).
16. Taioli E, Garte SJ. Lung cancer risk in African-Americans in relation to a race-specific CYPIA1 polymorphism [letter]. Cancer Res 56:4275–4277 (1996).
17. Vines P. Molecular epidemiology: low dose carcinogens and genetic susceptibility. Int J Cancer 71:1–3 (1997).
18. Garte SJ, Zocchetti C, Taioli E. Gene–environment interactions in the application of biomarkers of cancer susceptibility in epidemiology. In: Methodological Issues in the Use of Biomarkers in Cancer Epidemiology (Toniolo P, Boffetta P, Shuker D, Hulka B, Pearce N, Rothman N, eds). Lyon: International Agency for Research on Cancer, 1987:251–264.
19. Hirvonen A, Husgafvel-Pursiainen K, Anttila S, Vainio H. The GSTM1 null genotype as a potential risk modifier for squamous cell carcinoma of the lung. Carcinogenesis 14:1479–1481 (1993).
20. Khara M, Kihara M, Noda K. Lung cancer risk of GSTM1 null genotype is dependent on the extent of tobacco smoke exposure. Carcinogenesis 15:415–418 (1994).
21. Smith CM, Kelsey KT, Wiencke JK, Leyden K, Levin S, Christiani DC. Inherited glutathione-S-transferase deficieny is a risk factor for pulmonary asbestosis. Cancer Epidemiol Biomarkers Prev 3:471–477 (1994).
22. Crofts F, Taioli E, Trachman J, Cosma GN, Currie D, Toniolo P, Garte SJ. Functional significance of different human CYPIA1 genotypes. Carcinogenesis 15:2961–2963 (1994).
23. Landi MT, Bertazzi PA, Shields PG, Clark G, Lucier GW, Garte SJ, Cosma G, Caporaso NE. Restriction fragment length polymorphism, mRNA expression and enzymatic product of CYPIA1 gene in humans. Pharmacogenetics 4:242–246 (1994).
24. Cosma G, Crofts F, Currie D, Wing J, Toniolo P, Garte SG. Racial differences in restriction fragment length polymorphisms and messenger RNA inducibility of the human CYPIA1 gene. Cancer Epidemiol Biomarkers Prev 2:53–57 (1993).
25. Taioli E, Crofts F, Trachman J, Bayo S, Brador L, Toniolo P, Garte SJ. Racial differences in CYPIA1 genotype and function. Toxicol Lett 77:357–362 (1995).
26. Nakachi K, Imai K, Hayashi S, Watanabe J, Kawajiri K. Genetic susceptibility to squamous cell carcinoma of the lung in relation to cigarette smoking dose. Cancer Res 51:5177–5181 (1991).
27. Brown CC, Chu KC. Additive and multiplicative models and multistage carcinogenesis theory. Risk Anal 9:99–105 (1989).