

## REVIEW

# Gene–environment interactions for complex traits: definitions, methodological requirements and challenges

Astrid Dempfle<sup>\*</sup>,<sup>1</sup>, André Scherag<sup>1,3</sup>, Rebecca Hein<sup>2</sup>, Lars Beckmann<sup>2</sup>, Jenny Chang-Claude<sup>2</sup> and Helmut Schäfer<sup>1</sup>

<sup>1</sup>Institute of Medical Biometry and Epidemiology, Philipps University Marburg, Marburg, Germany; <sup>2</sup>Unit of Genetic Epidemiology, Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany

Genetic and environmental risk factors and their interactions contribute to the development of complex diseases. In this review, we discuss methodological issues involved in investigating gene–environment ( $G \times E$ ) interactions in genetic–epidemiological studies of complex diseases and their potential relevance for clinical application. Although there are some important examples of interactions and applications, the widespread use of the knowledge about  $G \times E$  interaction for targeted intervention or personalized treatment (pharmacogenetics) is still beyond current means. This is due to the fact that convincing evidence and high predictive or discriminative power are necessary conditions for usefulness in clinical practice. We attempt to clarify conceptual differences of the term ‘interaction’ in the statistical and biological sciences, since precise definitions are important for the interpretation of results. We argue that the investigation of  $G \times E$  interactions is more rewarding for the detailed characterization of identified disease genes (ie at advanced stages of genetic research) and the stratified analysis of environmental effects by genotype or *vice versa*. Advantages and disadvantages of different epidemiological study designs are given and sample size requirements are exemplified. These issues as well as a critical appraisal of common methodological concerns are finally discussed.

*European Journal of Human Genetics* (2008) 16, 1164–1172; doi:10.1038/ejhg.2008.106; published online 4 June 2008

**Keywords:** effect modification; genetic association; pharmacogenetics; interaction

## Introduction

It is generally accepted that both genetic and environmental factors contribute to the development of complex diseases. Thus, gene–environment ( $G \times E$ ) interaction is a hot topic in human genetics and there are great expecta-

tions for potential applications. Personalized medicine or individualized lifestyle recommendations based on the genetic profile are being promoted as the future of public health. Substantial funds devoted to study the genetics of human diseases are justified by these expectations. However, up to now, there are only a few replicated, biologically plausible and methodologically sound examples of  $G \times E$  interactions with a proven clinical relevance<sup>1,2</sup> and even less are used in daily clinical routines.<sup>3</sup> The extent to which  $G \times E$  interactions are of general importance for the development of common, complex diseases is currently unknown, even though important examples exist. Formal genetic evidence for  $G \times E$  interaction can consist of the observation that a certain exposure has different effects in

<sup>\*</sup>Correspondence: Dr A Dempfle, Institute of Medical Biometry and Epidemiology, Philipps University Marburg, Bunsenstr. 3, Marburg 35037, Germany.

Tel: +49 6421 28 66504; Fax: +49 6421 28 68921;

E-mail: dempfle@med.uni-marburg.de

<sup>3</sup>Current address: Institute of Medical Informatics, Biometry and Epidemiology, University of Duisburg-Essen, Essen, Germany.

Received 6 February 2007; revised 20 February 2008; accepted 6 May 2008; published online 4 June 2008

different populations or ethnic groups or in people with different genetically determined phenotypes. One example is exposure to sunlight that raises the risk of melanoma much more in fair-skinned than in dark-skinned people, that is there is an interaction between ultraviolet light and skin pigmentation.<sup>4</sup>

Constant advances in genotyping technology now enable genome-wide association studies and researchers are tempted to investigate their data as comprehensively as possible, including  $G \times E$  interactions. In this review, we present the perspectives for clinical applications, clarify definitions, discuss the range of application, and the design and required sample size of epidemiological  $G \times E$  studies. We conclude with some cautionary remarks on methodological challenges of such studies.

### Potential applications of $G \times E$ interactions in public health and clinical care

The most important area of application for  $G \times E$  interactions is personalized medicine, both in prevention and treatment (pharmacogenetics). Regarding the first, personalized prevention recommendations could be developed if the effects of an environmental risk factor strongly depend on an identified genetic polymorphism. In this sense, the assessment of the effects of genotypes in different exposure strata or *vice versa* of environmental exposures on disease risk in different genotype groups might be useful, even without *a priori* knowledge of the precise biological mechanisms underlying the statistical interaction. However, even the existence of a strong interaction does not imply that high-risk individuals can be easily identified for a targeted intervention, as usually many other factors will be important in disease development. This is for example the case for most so-called ‘sporadic cancers’ where presumably a strong stochastic element is involved in carcinogenesis, making accurate prediction of individual disease risk almost impossible.<sup>2</sup> Moreover, most study designs will not yield unbiased estimates of effects – the influence of the investigated risk factors is often overestimated.<sup>5,6</sup> From a public health perspective, the idea of personalized recommendations and targeted intervention has been questioned, as the overall benefit of small changes at a population level may be larger than that of large changes in high-risk individuals.<sup>7</sup> Whenever the interaction results only in a stronger or smaller detrimental effect of an exposure in the different genotype groups, all individuals may benefit from avoiding the exposure if the exposure is causally related to the disease. It is this very situation in which general recommendations are advisable, for example like those regarding exercise, smoking and diet.<sup>8</sup> Personalized recommendations, however, may be considered reasonable for cases when an exposure has a null or negative effect in one

genotype group and a protective effect in another genotype group.

Also the second area of application, pharmacogenetics,<sup>9</sup> relies on the existence of such strong  $G \times E$  interactions. It is implicit that individuals with different genotypes will benefit from different medication in a predictable manner.<sup>3</sup> Although it is plausible that the different reactions of patients to drugs may depend on their individual genetic ‘make-up’, the systematic study of such interactions is still in its beginnings. A prerequisite for widespread use in clinical practice is that the genetic variant is a sufficiently strong predictor of harm or benefit.<sup>5</sup> One example is anticoagulant treatment, where it is known that warfarin clearance depends on the genotype of the metabolizing enzyme cytochrome P-450 2C9 (CYP2C9). About one-third of Caucasian patients possess one of the polymorphisms that require a reduced maintenance dose of warfarin to avoid adverse side effects. Prior to integration of genetic information in clinical practice randomized, controlled clinical trials will be required to demonstrate the benefits of including CYP2C9 genotype in warfarin dosing (together with other covariates) compared to traditional dose-finding methods.<sup>10,11</sup> For a more detailed view of the potential impact of pharmacogenetics on public health we refer to a review by Goldstein *et al.*<sup>12</sup>

### Definition and meaning of interaction

While reviewing the data, one will often notice that both different connotations and different concepts of the term interaction are used by statisticians, clinicians, biologists and geneticists.<sup>13,14</sup> Frequently, a precise definition is completely omitted, which may lead to some confusion and controversy between scientists of different disciplines. Quite commonly in general contexts, ‘ $G \times E$  interaction’ is used in a very loose sense, meaning some sort of interplay between genetic and environmental factors. However, a specific mode of joint action or a certain relationship between statistical risks is not implied in many cases. Sometimes it is even used to express that several factors contribute to disease risk, without excluding the possibility of complete independence. In these cases using for example the term ‘joint action’ would be preferable. If ‘interaction’ is used in a narrower sense, it can refer to a biological (causal) or statistical level and we will define it here, introducing commonly used statistical terminology and finally distinguishing it from confounding.

Biological interaction is defined as the joint effect of two factors that act together in a direct physical or chemical reaction and the coparticipation of two or more factors in the same causal mechanism of disease development.<sup>15</sup> Further notations are causal or mechanical interaction. Examples of biological interaction are the direct reaction of a certain exposure with, for example, an enzyme whose

detoxification ability depends on the genotype of a certain gene. A good overview of possible causal relationships and interaction mechanisms is given by Ottman.<sup>16</sup> Such etiological mechanisms have to be explored by functional studies.

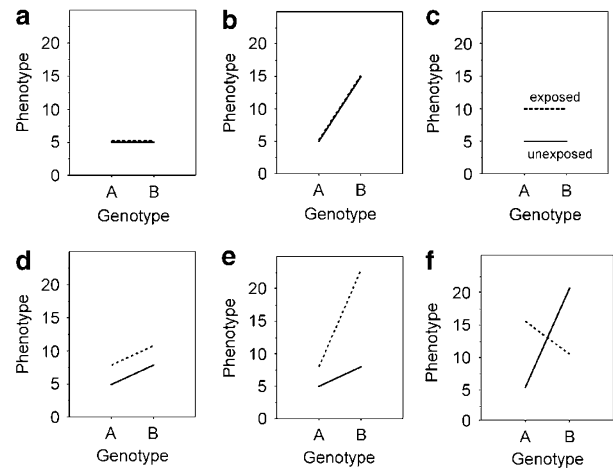
On the other hand, there is the definition of statistical interaction, which does not imply any inference about particular biological modes of action. Statistical interaction (or heterogeneity of effects) is usually defined as ‘departure from additivity of effects on a specific outcome scale’.<sup>14</sup> If only one factor is present, its effect on the risk of disease is called main effect. In the case where two or more risk factors are present, the marginal effect of a risk factor is its average effect across all levels of the other risk factors. The risk factors are said to interact, if the effect of one risk factor depends on the level of the other risk factor (Table 1). Several equivalent terms denoting statistical interaction exist, such as non-additivity, effect measure modification or heterogeneity of effects. The joint effect of two risk factors refers to both their marginal effects and their interaction effect. The joint effect can vary from less than additive (subadditive) to more than multiplicative (supramultiplicative) of the individual marginal effects. Theoretical models for such interaction relationships have been explored especially for cancer development, where carcinogens act at different stages.<sup>17</sup>

Interactions are sometimes divided into removable and nonremovable:<sup>18</sup> if a monotone transformation (eg taking logarithms or square roots of quantitative phenotypes) exists that removes the interaction<sup>19</sup> (Figure 1), it is called removable. This implies that there is an additive relationship between the variables, just on a different scale. Therefore, nonremovable interactions are usually of greater interest. To complete the terminology, nonremovable interaction effects are also called crossover effects<sup>20</sup> or qualitative interactions (as opposed to quantitative, ie removable interactions).

Furthermore, it is necessary to distinguish between interaction and confounding of environmental and genetic factors. Confounding refers to a mixing of extraneous effects with the effect of interest,<sup>14</sup> for example a (true but unmeasured) risk factor of disease that is correlated with the investigated risk factor and results in a noncausative association. In the context of interactions, this could primarily be a correlation between the genetic and environmental risk factors, which could be misinterpreted as an interaction if the statistical model used does not account for the correlation but treats them as independent. Such a gene–environment correlation can occur in samples with latent population substructure (eg unintentionally including groups of different ethnicity) where both risk allele frequencies and exposure frequencies vary between subpopulations. It can also result from the influence of genes on behavior like alcohol consumption or food and satiety responsiveness that in turn are related

**Table 1** Example of additive and multiplicative models of relative risks for an environmental and a genetic risk factor

Environmental risk factor	Genetic risk factor			
	Additive model		Multiplicative model	
	Absent	Present	Absent	Present
Absent	1	2	1	2
Present	1.5	2.5	1.5	3



**Figure 1** Examples of main and interaction effects. Phenotypic values depending on genotype *G* (two groups, eg under a dominant genetic model) and exposure *E* (also two groups, exposed (dotted line) and unexposed (solid line)). (a) Neither *G* nor *E* have a main effect and there is no interaction; (b) *G* has a main effect, *E* has no main effect, no interaction; (c) *E* has a main effect, *G* has no main effect, no interaction; (d) both *G* and *E* have main effects, no interaction; (e) *G* and *E* have main effects and there is an interaction (which could be removed by changing the phenotype scale, eg to a logarithmic scale); (f) *G* and *E* have main effects and there is an interaction (which cannot be removed by any monotone transformation).

to diseases such as coronary heart disease or obesity. In many other contexts confounding would not be a serious concern, as genotype and environmental risk factors will usually be independent – genotypes are fixed throughout life and are thus not influenced by or associated with environmental exposures (cf. concept of ‘Mendelian randomization’<sup>21,22</sup>). At the data level, confounding and interaction may lead to similar patterns, especially in partial collection designs such as the case-only design. An identified interaction should therefore be carefully interpreted to consider whether confounding could explain part of the observed effect.

#### When should $G \times E$ interactions be investigated?

The analysis of  $G \times E$  interactions in genetic epidemiology can be done at both different time points during the

research process and with varying scopes. The relevant research questions that could be addressed by a  $G \times E$  interaction study include the identification of new disease genes, the characterization of gene effects, the clinical relevance of a  $G \times E$  interaction and the public health impact of it.

In the phase of identification of genetic risk factors, accounting for a  $G \times E$  interaction might increase the power to detect genes with small marginal effects,<sup>23–25</sup> especially if the effect of a gene is only relevant in an etiological subgroup of patients, defined by a certain exposure. Here, the interaction is not of specific interest *per se*. Especially for high-throughput genotyping of polymorphisms in hundreds of candidate genes or genome-wide association studies with several hundred thousands of polymorphisms, the inclusion and testing of interactions greatly increase the number of statistical tests and thus the need to correct for multiple testing. Joint tests of marginal and interaction effects<sup>25</sup> may provide power over a wide range of unknown true situations. However, in the absence of very strong interaction, tests for marginal gene effects are still the most powerful to identify a disease-related gene.

Alternatively, a  $G \times E$  study can be part of the detailed characterization of gene effects for genes that have already been shown to be involved in disease etiology but whose effect may vary across different environmental strata. In this case, the interaction itself is of interest and the aim of an initial study may be primarily hypothesis generating (exploratory), possibly investigating several environmental factors or different polymorphisms within one gene to provide effect size estimates. The next step would be to establish clinical relevance of a detected  $G \times E$  interaction, which involves confirmatory testing of one specific *a priori* hypothesis within the clinical population and under the circumstances proposed for later application. It also includes the estimation of the strength of the interaction (effect size, eg odds ratio). Ideally, such investigations will be part of a randomized controlled (phase III) trial. Finally, assessments of the public health impact of an established  $G \times E$  interaction depend on the strength of the interaction, exposure frequency and allele frequencies. More importantly, however, the ascertainment strategy and the study design will require careful considerations to enable generalizations of the study results.

### Study designs for $G \times E$

Common family- and population-based designs for association studies can be extended for  $G \times E$  interaction. Table 2 lists different designs with their respective advantages and disadvantages and research situations in which such a design would be suitable. Family-based designs protect against bias due to population stratification

with both differential exposure and genotype distribution in subgroups. In population-based designs, data on a quantitative trait or a disease phenotype are collected from unrelated individuals, either prospectively (cohort) or retrospectively (case–control). If a large prospective cohort exists, a nested case–control study can reduce selection and possibly stratification biases and be a good compromise regarding cost and efficiency.<sup>29</sup> For the relative merits of cohort and case–control designs see also the discussion started by Clayton and McKeigue,<sup>21</sup> who argue that case–control studies are more feasible and cost efficient than cohort studies for modest disease risks and that exposure misclassification bias is not a serious threat in the case of  $G \times E$  interactions. Others however stress this possible bias and emphasize the merit of cohorts in studying multiple end points and especially different diseases in one sample.<sup>30–33</sup>

If the interest is limited to  $G \times E$  interaction, the special ‘case-only’ design exists that has the practical advantage that no controls need to be collected.<sup>34</sup> This design is based on the assumption that genotype and environmental exposure are independent in the population that the case sample is drawn from, so that exposure should not differ among subgroups defined by genotype. Since, in the presence of a  $G \times E$  interaction, specific combinations of genotypes and exposure lead to increased risk of disease and thus are more prevalent among cases, differences in exposure will be observable between genotype groups in cases. Because of the independence assumption, the case-only design is more efficient than the traditional case–control design, but this assumption is not assessable in the case sample alone. Therefore, the design is prone to bias and confounding, especially if there is exposure misclassification (keeping in mind that especially lifetime environmental exposures are not as accurately measurable as genotypes).<sup>35–38</sup> Another drawback is that although estimation of the  $G \times E$  interaction is possible, the estimation of the joint effect of exposure and genotype is impossible<sup>39</sup> even though the latter usually is of greater importance for the public health aspect of a  $G \times E$  investigation. As a consequence, the practical applicability of this design is limited and it is rarely applied. The case–control design is better suitable to address the relevant research questions,<sup>40</sup> and if one is willing to make the assumption of gene–environment independence, analysis methods exist that also leverage this.<sup>39,41</sup>

Two special, nonstandard applications of  $G \times E$  interactions occur in infectious disease and pharmacogenetic studies. In infectious diseases, only individuals exposed to the infectious agent can contract the disease, thus the environmental factor is a necessary causal factor. Genes may modify the risk of infection (or disease severity) for those exposed.<sup>42–44</sup> Examples are the *CCR5* gene for HIV infection,<sup>45</sup> malaria and heterozygosity for sickle cell anemia<sup>46</sup> or variant Creutzfeld–Jakob disease and a

**Table 2** Study designs for genetic association studies that can include  $G \times E$  interactions with their main advantages and disadvantages and the situations in which these designs are most suitable

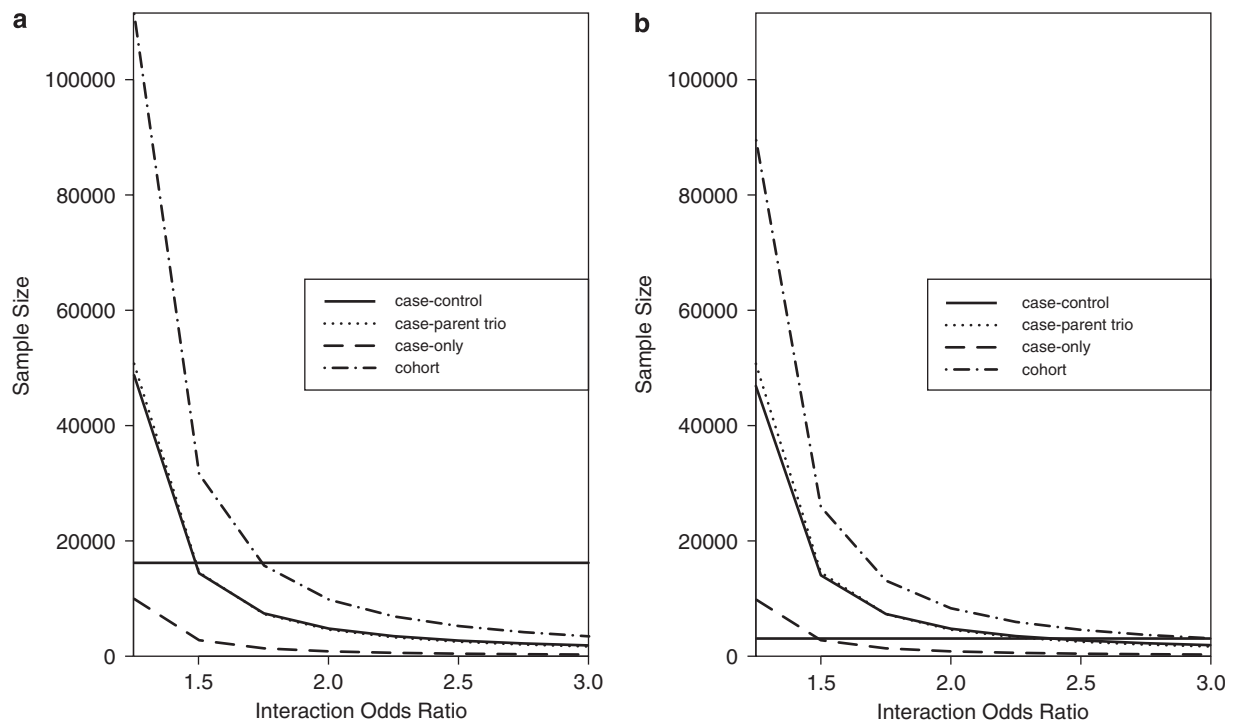
<i>Study design</i>	<i>Main advantage</i>	<i>Disadvantage</i>	<i>In which situation is this design most suitable?</i>
<i>Family-based designs</i>			
Trio design: <sup>26,27</sup> Case and both parents, evaluate transmission disequilibrium, depending on exposure	Protects against confounding due to population stratification	High ascertainment costs; often impossible for late onset diseases	Early onset diseases when population stratification is a concern
Sib design: <sup>27,28</sup> case and unaffected siblings	Protects against population stratification	Potentially difficult to recruit enough suitable families	Late onset diseases when population stratification is a concern
<i>Population-based designs</i>			
Cohort: compare disease frequency between genotype-exposure groups	Reduces selection and stratification bias compared to case–control design; prospective measurement of exposure	Expensive and time consuming; investigation of very rare diseases may be impossible with realistic sample sizes	Very reliable results due to low risk of biases, therefore suitable for confirmatory studies; useful for common diseases; can use existing cohorts with additional DNA collection
Case–control (retrospective): compare exposure rate and genotype frequencies between cases and controls	Simple, relatively inexpensive and less time consuming compared to cohort studies	Possible stratification bias due to population stratification. Higher risk of measurement error in exposure	Only reasonable population-based design for very rare diseases; very suitable for first exploratory studies
Case-control (nested, prospective) <sup>29</sup>	Less expensive than full cohort design, reduced biases compared to retrospective case-control design		For rare or common diseases in existing cohorts; confirmatory studies
Case-only: evaluate differences in exposure between genotype groups in cases	Simple, relatively inexpensive and less time consuming compared to cohort studies; no controls needed; high power for $G \times E$ interaction test	Assumption of independence of $G$ and $E$ not assessable in the case-only sample, therefore prone to bias and confounding; joint effect of $G$ and $E$ cannot be estimated	Fast, uncomplicated exploratory studies, results need to be confirmed with other study designs
Clinical trial, randomization stratified by genotype	Gives reliable results of clinical relevance of a presumed gene-by-drug interaction	Increased sample size and higher cost over traditional clinical trial	If preliminary studies suggest differential drug effects by genotype; basis for personalized medicine
Exposed-only for infectious diseases	No need to include a presumably unexposed group		Infectious disease genetics

polymorphism in codon 129 of the prion protein gene PRNP.<sup>47</sup> In these examples, individuals with certain genotypes have a much lower risk for infection or progression to serious disease. Infectious disease studies usually include only individuals at high risk of infection (assumed to be exposed). Here, the aim is an investigation of potential differences in disease prevalence between genotype groups similar to the usual genetic association or linkage studies without explicit consideration of  $G \times E$  interaction in the statistical analysis. Such differences can then be interpreted as  $G \times E$  interactions, since the genotype alone cannot lead to an infectious disease. Similarly, some pharmacogenetic studies for licensed drugs aim at identifying individuals at risk for serious side effects

or increased efficacy by exclusively including drug-treated patients. In this design it is impossible to distinguish between genetic effects and  $G \times E$  interaction. More suitable is a design that includes pharmacogenetic aspects in randomized clinical trials by giving placebo or active drug stratified according to genotype.<sup>48,49</sup>

### Sample size and power

Depending on the strength of the interaction and exposure and allele frequencies, sample size requirements to detect a statistically significant  $G \times E$  interaction may be substantially larger than the sample sizes to identify a  $G$



**Figure 2** Sample size requirements for 80% power to detect a gene–environment ( $G \times E$ ) interaction for different study designs depending on the strength of the interaction. Sample sizes for case–control, case–parent trio, and case-only designs were calculated using Quanto<sup>59</sup> (<http://hydra.usc.edu/gxe>), assuming an analysis by (conditional) logistic regression. For the cohort design, sample sizes are estimated using Power<sup>58</sup> (<http://dcegqa.cancer.gov/bb/tools/power>), which is based on a prospective binary response model. Shown are the number of individuals required to detect a significant  $G \times E$  interaction effect at  $\alpha = 0.01$  with a power of 80%. Solid lines represent the case–control design, dotted lines the trio design, dashed line the case-only design and dotted-dashed lines the cohort design. The horizontal solid line represents the sample size required for 80% power to detect a genetic main effect using a case–control design. The interaction odds ratio was varied between 1.25 and 3 whereas the main effects of the genetic and environmental risk factors were 1.2 (a) and 1.5 (b). The disease model was defined by a recessive disease allele with frequency 0.3. The environmental risk factor had a prevalence of 30%. The baseline risk of the disease was 10%. The samples sizes to detect the genetic main effect, which were constant in the two scenarios, were 43 045, 16 196 and 19 860 in (a) for the cohort, case–control and trio design, and 7712, 3070 and 3423 in (b), respectively. For a dominant disease allele, similar relations between required sample sizes are observed for the different designs.

or E marginal effect. Some illustrative examples for association studies of a candidate gene are shown in Figure 2, which give the required samples sizes for four different study designs (case–control, trio, case-only and cohort) for varying effect sizes of the  $G \times E$  interaction. Only for very weak marginal effects ( $OR = 1.2$ , a) and at least moderate interactions ( $OR > 1.5$ ), the interaction is detectable with a smaller sample size than the marginal effect. But even for slightly larger marginal effects ( $OR = 1.5$ , b) and weak to moderate interactions ( $OR < 2$ ), the sample size required to detect the interaction can be several fold higher than that required for detecting the marginal genetic effect. These examples are based on a level of significance (0.01) that might be used in a confirmatory study for testing one well-defined *a priori* hypothesis (eg one polymorphism within one gene). Sample sizes would be much higher for (exploratory) studies such as genome-wide association scans with hundreds of thousands of markers, as the correction for multiple testing requires much smaller levels

of significance and thus much larger samples. In addition, these studies rely on linkage disequilibrium between the genotyped markers and potentially untyped disease alleles, and such indirect association studies may need much larger sample sizes.<sup>50</sup> Especially for  $G \times E$  interactions that might realistically be even smaller, large cohorts such as BioBank UK (planned with 500 000 individuals over 10 years<sup>33</sup>), EPIC<sup>51</sup> and the Multi-ethnic Cohort<sup>52</sup> will be necessary. Although a sample size of 500 000 might be useful for common diseases such as type II diabetes, it will still be insufficient for rarer diseases with prevalence less than approximately 1%, for which case–control studies might be the only feasible approach.<sup>21</sup>

Note that sample size and power calculations are also possible for other study designs, for example for association studies of quantitative traits,<sup>53</sup> categorical or continuous exposure variables<sup>54</sup> as well as for pharmacogenetic study designs.<sup>55–57</sup> Freely available software programs such as Power,<sup>58</sup> Quanto<sup>59</sup> or a Stata program by Saunders *et al*<sup>60</sup> may be used if required.

### Methodological challenges and perspectives

In summary, the methodological requirements for a  $G \times E$  interaction study are greatly driven by the research question. We thus conclude by addressing five common caveats that need to be considered: the study aims, the conduct of a study, reporting and interpretation of results, extending inferences and clinical relevance.

First, one should distinguish between primarily exploratory (ie hypothesis-generating) or confirmatory (hypothesis testing) study aims. In our opinion, genome-wide association studies and small initial studies can only be considered exploratory. The latter will often be performed, for example because of difficult or time-consuming phenotyping, limited availability of the required biological material (eg tissue samples) and financial constraints. Both approaches are important and valid first steps in research but their exploratory nature has to be kept in mind. Therefore, such smaller studies will be valuable for generating hypotheses that should then be tested for confirmation in adequately powered, presumably larger studies. On the other hand, inadequate sample sizes lead to underpowered studies that give rise to both false-negative and false-positive findings especially at the hypothesis-generating stage. Biological relationships cannot be inferred from genetic–epidemiological studies, and further functional experiments are necessary for this.

Second, a well-designed confirmatory study of  $G \times E$  interaction should be based on a justifiable *a priori* hypothesis of an interaction between a plausible or established gene with known function and a known environmental risk factor with some link to gene function, for which a reasonable biological interaction mechanism exists. Only prespecified (prior to data collection) hypotheses and statistical tests can be interpreted as confirmatory. Ideally, there is evidence from formal genetic studies (eg twin studies or segregation analyses) of an interaction between the exposure and genetic factors. Next, an appropriate study design (see above) must be chosen and a sufficient sample size needs to be pheno- and genotyped. Then, an adequate statistical analysis is needed (including a multiple comparison procedure for control of the type I error if more than one statistical test is conducted).

Third, reporting and interpretation of detected  $G \times E$  interactions should be faithful and balanced. Reporting should center on what range of true effects would be compatible with the observed effects (using confidence intervals of effect estimates) and it should be discussed whether these could be of a clinically relevant size. By contrast, less emphasis should be on the results of significance tests (*P*-values) as these will be misleading if the reader is unaware of the multiple tests performed. To avoid publication bias, all test results (or at least the number of tests performed) must be reported, not only interactions that are nominally significant (eg at a 5% level). Overreporting and overinterpretation of results will

lead to inconsistent and inconclusive results.<sup>2,61</sup> And even in case of careful descriptions, effect estimates in initial reports tend to be biased<sup>5,6</sup> and may vary between different populations with different allele and exposure frequencies.

Fourth, if some evidence for a  $G \times E$  interaction is observed, its biological plausibility should be critically discussed and potential confounders or intermediate pathways have to be explored. Here, one has to keep in mind that conclusions dealing with a certain biological mechanism cannot be confirmed or rejected by statistical arguments based on epidemiological data alone.<sup>20</sup> Only in light of additional lines of evidence, such as functional experiments, may the inferences toward causality be extended.

Finally, even though the potential clinical relevance or impact of a reported  $G \times E$  interaction may be discussed, these implications should be evaluated in subsequent studies designed for that special purpose. At this subsequent stage, the choice of the appropriate phenotype(s) is of special importance and clinically relevant end points and disease-related phenotypes, such as myocardial infarction, need to be studied before study results are embedded in public health programs or exploited for personalized medicine and individualized lifestyle recommendations.<sup>1,5</sup> Note that physiological and biochemical phenotypes (endophenotypes), such as lipid levels, IgE levels and so on may be closer to the underlying gene action and may thus be more appropriate for elucidating the biological mechanism underlying an interaction. Such biomarkers are, however, at most surrogate risk factors for a disease. Clinical relevance by contrast requires that the predictive or discriminative power of the genotype for the clinically defined disease (eg death due to myocardial infarction) or treatment success (eg extended survival time) has to be sufficiently high. Predominantly, this will be the case for strong qualitative interactions.

When these challenging requirements are fulfilled, research on  $G \times E$  interactions can yield valuable insights into the etiology of complex diseases. Ultimately, this knowledge may contribute to more effective strategies for prevention and treatment.

### Acknowledgements

This work was funded by the Bundesministerium für Bildung und Forschung through the German National Genome Net (NGFN2, grant numbers 01GR0460 and 01GR0461).

### References

- 1 Ordovas JM, Mooser V: Nutrigenomics and nutrigenetics. *Curr Opin Lipidol* 2004; **15**: 101–108.
- 2 Brennan P: Gene–environment interaction and aetiology of cancer: what does it mean and how can we measure it? *Carcinogenesis* 2002; **23**: 381–387.
- 3 Gardiner SJ, Begg EJ: Pharmacogenetic testing for drug metabolizing enzymes: is it happening in practice? *Pharmacogenomics* 2005; **15**: 365–369.

- 4 Rees JL: The genetics of sun sensitivity in humans. *Am J Hum Genet* 2004; **75**: 739–751.
- 5 Hauser ER, Allen AS: Where the rubber meets the road in pharmacogenetics: assessment of gene-environment interactions. *Am Heart J* 2003; **146**: 929–931.
- 6 Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG: Replication validity of genetic association studies. *Nat Genet* 2001; **29**: 306–309.
- 7 Rose G: Sick individuals and sick populations. *Int J Epidemiol* 2001; **30**: 427–432.
- 8 Willett WC: Balancing life-style and genomics research for disease prevention. *Science* 2002; **296**: 695–698.
- 9 Roses AD: Pharmacogenetics and drug development: the path to safer and more effective drugs. *Nat Rev Genet* 2004; **5**: 645–656.
- 10 Hillman MA, Wilke RA, Yale SH et al: A prospective, randomized pilot trial of model-based warfarin dose initiation using CYP2C9 genotype and clinical data. *Clin Med Res* 2005; **3**: 137–145.
- 11 Voora D, Eby C, Linder MW et al: Prospective dosing of warfarin based on cytochrome P-450 2C9 genotype. *Thromb Haemost* 2005; **93**: 700–705.
- 12 Goldstein DB, Tate SK, Sisodiya SM: Pharmacogenetics goes genomic. *Nat Rev Genet* 2003; **4**: 937–947.
- 13 Cordell HJ: Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet* 2002; **11**: 2463–2468.
- 14 Rothman KJ, Greenland S: *Modern Epidemiology*. Philadelphia: Lippincott-Raven, 1998.
- 15 Yang Q, Khoury MJ: Evolving methods in genetic epidemiology. III. Gene-environment interaction in epidemiologic research. *Epidemiol Rev* 1997; **19**: 33–43.
- 16 Ottman R: An epidemiologic approach to gene-environment interaction. *Genet Epidemiol* 1990; **7**: 177–185.
- 17 Thomas DC: Temporal effects and interactions in cancer: implications of carcinogenic models; in Prentice RL, Whittemore AS (eds): *Environmental Epidemiology: Risk Assessment*. Philadelphia: Society for Industrial and Applied Mathematics, 1982, pp 107–121.
- 18 Yusuf S, Wittes J, Probstfield J, Tyroler HA: Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991; **266**: 93–98.
- 19 Tukey JW: One degree of freedom for non-additivity. *Biometrics* 1949; **5**: 232–242.
- 20 Thompson WD: Effect modification and the limits of biological inference from epidemiologic data. *J Clin Epidemiol* 1991; **44**: 221–232.
- 21 Clayton D, McKeigue PM: Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet* 2001; **358**: 1356–1360.
- 22 Davey Smith G, Ebrahim S: 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; **32**: 1–22.
- 23 Gauderman WJ, Thomas DC: The role of interacting determinants in the localization of genes. *Adv Genet* 2001; **42**: 393–412.
- 24 Dizier MH, Selinger-Leneman H, Genin E: Testing linkage and gene × environment interaction: comparison of different affected sib-pair methods. *Genet Epidemiol* 2003; **25**: 73–79.
- 25 Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ: Exploiting gene-environment interaction to detect genetic associations. *Hum Hered* 2007; **63**: 111–119.
- 26 Schaid DJ: Case-parents design for gene-environment interaction. *Genet Epidemiol* 1999; **16**: 261–273.
- 27 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; **149**: 693–705.
- 28 Gauderman WJ, Witte JS, Thomas DC: Family-based association studies. *J Natl Cancer Inst Monogr* 1999; **26**: 31–37.
- 29 Hunter DJ: Gene-environment interactions in human diseases. *Nat Rev Genet* 2005; **6**: 287–298.
- 30 Wacholder S, Garcia-Closas M, Rothman N: Study of genes and environmental factors in complex diseases. *Lancet* 2002; **359**: 1155.
- 31 Burton P, McCarthy M, Elliott P: Study of genes and environmental factors in complex diseases. *Lancet* 2002; **359**: 1155–1156.
- 32 Stene LC: Study of genes and environmental factors in complex diseases. *Lancet* 2002; **359**: 1156.
- 33 Banks E, Meade T: Study of genes and environmental factors in complex diseases. *Lancet* 2002; **359**: 1156–1157.
- 34 Khoury MJ, Flanders WD: Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: case-control studies with no controls!. *Am J Epidemiol* 1996; **144**: 207–213.
- 35 Schmidt S, Schaid DJ: Potential misinterpretation of the case-only study to assess gene-environment interaction. *Am J Epidemiol* 1999; **150**: 878–885.
- 36 Gatto NM, Campbell UB, Rundle AG, Ahsan H: Further development of the case-only design for assessing gene-environment interaction: evaluation of and adjustment for bias. *Int J Epidemiol* 2004; **33**: 1014–1024.
- 37 Garcia-Closas M, Thompson WD, Robins JM: Differential misclassification and the assessment of gene-environment interactions in case-control studies. *Am J Epidemiol* 1998; **147**: 426–433.
- 38 Vineis P: A self-fulfilling prophecy: are we underestimating the role of the environment in gene-environment interaction research? *Int J Epidemiol* 2004; **33**: 945–946.
- 39 Umbach DM, Weinberg CR: Designing and analysing case-control studies to exploit independence of genotype and exposure. *Stat Med* 1997; **16**: 1731–1743.
- 40 Liu X, Fallin MD, Kao WH: Genetic dissection methods: designs used for tests of gene-environment interaction. *Curr Opin Genet Dev* 2004; **14**: 241–245.
- 41 Chatterjee N, Kalaylioglu Z, Carroll RJ: Exploiting gene-environment independence in family-based case-control studies: increased power for detecting associations, interactions and joint effects. *Genet Epidemiol* 2005; **28**: 138–156.
- 42 Abel L, Dessein AJ: Genetic epidemiology of infectious diseases in humans: design of population-based studies. *Emerg Infect Dis* 1998; **4**: 593–603.
- 43 Hill AV: Genetics and genomics of infectious disease susceptibility. *Br Med Bull* 1999; **55**: 401–413.
- 44 Clementi M, Di Gianantonio E: Genetic susceptibility to infectious diseases. *Reprod Toxicol* 2006; **21**: 345–349.
- 45 Smith MW, Dean M, Carrington M et al: Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression. Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC), ALIVE Study. *Science* 1997; **277**: 959–965.
- 46 Pouniotis DS, Proudfoot O, Minigo G, Hanley JL, Plebanski M: Malaria parasite interactions with the human host. *J Postgrad Med* 2004; **50**: 30–34.
- 47 Brown P, Cervenakova L, Goldfarb LG et al: Iatrogenic Creutzfeldt-Jakob disease: an example of the interplay between ancient genes and modern medicine. *Neurology* 1994; **44**: 291–293.
- 48 Cardon LR, Idury RM, Harris TJ, Witte JS, Elston RC: Testing drug response in the presence of genetic information: sampling issues for clinical trials. *Pharmacogenetics* 2000; **10**: 503–510.
- 49 Kelly PJ, Stallard N, Whittaker JC: Statistical design and analysis of pharmacogenetic trials. *Stat Med* 2005; **24**: 1495–1508.
- 50 Hein R, Beckmann L, Chang-Claude J: Sample size requirements for indirect association studies of gene-environment interactions (G × E). *Genet Epidemiol* 2008; **32**: 235–245.
- 51 Riboli E, Hunt KJ, Slimani N et al: European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002; **5**: 1113–1124.
- 52 Kolonel LN, Altshuler D, Henderson BE: The multiethnic cohort study: exploring genes, lifestyle and cancer risk. *Nat Rev Cancer* 2004; **4**: 519–527.



- 53 Luan JA, Wong MY, Day NE, Wareham NJ: Sample size determination for studies of gene–environment interaction. *Int J Epidemiol* 2001; **30**: 1035–1040.
- 54 Foppa I, Spiegelman D: Power and sample size calculations for case–control studies of gene–environment interactions with a polytomous exposure variable. *Am J Epidemiol* 1997; **146**: 596–604.
- 55 Judson R: Using multiple drug exposure levels to optimize power in pharmacogenetic trials. *J Clin Pharmacol* 2003; **43**: 816–824.
- 56 Singer C, Grossman I, Avidan N, Beckmann JS, Pe'er I: Trick or treat: the effect of placebo on the power of pharmacogenetic association studies. *Hum Genomics* 2005; **2**: 28–38.
- 57 Elston RC, Idury RM, Cardon LR, Lichter JB: The study of candidate genes in drug trials: sample size considerations. *Stat Med* 1999; **18**: 741–751.
- 58 Garcia-Closas M, Lubin JH: Power and sample size calculations in case–control studies of gene–environment interactions: comments on different approaches. *Am J Epidemiol* 1999; **149**: 689–692.
- 59 Gauderman WJ: Sample size requirements for matched case–control studies of gene–environment interaction. *Stat Med* 2002; **21**: 35–50.
- 60 Saunders CL, Bishop DT, Barrett JH: Sample size calculations for main effects and interactions in case–control studies using Stata's `nchi2` and `npnchi2` functions. *Stata J* 2003; **3**: 47–56.
- 61 Ryan SG: Regression to the truth: replication of association in pharmacogenetic studies. *Pharmacogenomics* 2003; **4**: 201–207.