

# A Comparison of Case-Control and Case-Only Designs to Investigate Gene-Environment Interactions Using Breast Cancer Data

Jafar Hassanzadeh<sup>1</sup>, Rahmatollah Moradzadeh<sup>2</sup>, Abdolreza Rajaei fard<sup>1</sup>, Sedigheh Tahmasebi<sup>3</sup>, Parvaneh Golmohammadi<sup>2</sup>

## Abstract

**Background:** The traditional methods of studying the gene-environment interactions need a control group. However, the selection of an appropriate control group has been associated with problems. Therefore, new methods, such as case-only design, have been created to study such interactions. The objective of this study was to compare the case-only and case-control designs using data from patients with breast cancer.

**Methods:** The interaction of genetic and environmental factor as well as the ratio of control to population odds ratio was calculated for case-only (300 patients with breast cancer) and case-control (300 cases of breast cancer and 300 matched controls) designs.

**Results:** The confidence intervals and  $-2\log$  likelihood in all variables in case-only design was smaller than those in the matched case-control design. In case-only design, the standard errors of some variables such as age at menarche, the first delivery at the age of 35 yrs and more or no delivery, the history of having live birth, use of oral contraception pills, breastfeeding history were less than those in the matched case-control design.

**Conclusion:** The findings indicate that the case-only design is an efficient method to investigate the interaction of genetic and environmental factors.

Please cite this article as: J. Hassanzadeh, R. Moradzadeh, A.R. Rajaei fard, S. Tahmasebi, P. Golmohammadi. A Comparison of Case-Control and Case-Only Designs to Investigate Gene-Environment Interactions Using Breast Cancer Data. *Iran J Med Sci.* 2012;37(2):112-118.

**Keywords** • Case-control • breast cancer • gene-environment interaction

<sup>1</sup>Department of Epidemiology, School of Health and Nutrition, Shiraz University of Medical Sciences, Shiraz, Iran.

<sup>2</sup>Department of Epidemiology, School of Nursing and Midwifery, Islamic Azad University, Sanandaj, Iran.

<sup>3</sup>Department of General Surgery, Mottahari Clinic, Shiraz University of Medical Sciences, Shiraz, Iran.

## Correspondence:

Rahmatollah Moradzadeh MSc,  
Department of Epidemiology, School of Nursing and Midwifery, Islamic Azad University, Sanandaj Branch, Pasdaran Avenue, Sanandaj, Iran.

**Tel:** +98 871 2284495

**Fax:** +98 871 3288684

**Email:** moradzadehr@yahoo.com

Received: 20 April 2011

Revised: 31 October 2011

Accepted: 10 November 2011

## Introduction

Many common diseases are caused by the interaction of genetic and environmental factors.<sup>1</sup> Interaction is used in epidemiology to describe a situation in which two or more risk factors adjust the effects of each other with respect to the level of a certain outcome.<sup>2</sup> In other words, when the incidence of a disease in the presence of two factors is different from the incidence of that disease in the presence of either factor alone, the interaction of the two factors is understood.<sup>3</sup>

Although, the case-control designs are appropriate for examining gene-environment interactions, they have some limitations including the high cost and time needed to select the control group, a big sample size for estimating interactions, and the limitations in selecting an appropriate control group.<sup>4</sup> Recently, several modern methods have been created to study the diseases genetic factors, which are based on using the internal control group instead of external ones. One of these methods, are case-only designs in

which researchers use individuals to assess the magnitude of a relationship between a specific exposure and genotype susceptibility.<sup>5</sup> This design does not have a lot of limitations, which exist in the analogous case-control studies.<sup>4</sup>

Breast cancer is the most-frequently diagnosed cancer in women,<sup>6</sup> and is a worldwide concern.<sup>7</sup> It constitutes about one-third of all cancers among women.<sup>8</sup> Approximately, one out of nine women is affected by breast cancer throughout her life.<sup>9</sup> The well-established risk factors of breast cancer such as age at menarche, age at the first delivery, age at menopause, and alcohol consumption may be the criteria for cumulative exposure of breast epithelium to estrogenic substances.<sup>10</sup> Previous studies have shown that the family history and its genetic polymorphisms may be a guide to constitute the familial patterns of estrogen endogenous level.<sup>10</sup>

Several studies have assessed the fertility factors and familial predisposition to breast cancer to investigate the gene-environment interactions.<sup>4,11-14</sup> In these studies, pedigree information, which is the family history of breast cancer in the first degree relatives of case groups, was used as a criterion for replacing the genetic susceptibility. The case-only designs may be used to examine genetic-drug interactions, survival, and some other studies. A major issue in genetic epidemiology is that diseases results from interactions between genetic and environmental factors.<sup>5</sup> The aim of this study was to compare the case-only and case-control designs using the data related to patients with breast cancer in the city of Shiraz, Iran.

## Materials and Methods

The study included patients with breast cancer referring to Shahid Mottahari Breast Cancer Clinic in Shiraz. Patients had been identified through screening programs for breast cancer in Health Care Centers, and had been referred to Shahid Mottahari Breast Cancer Clinic. A questionnaire comprising of demographic data, reproductive factors, care, and treatment had been completed for every patient. By the time of the present study, two thousands questionnaires about patients with breast cancer had been completed by physicians and nurses in the clinic. Quanto1.2 Software (January 2007),<sup>15</sup> was used to determine the sample size of 300 subjects. The samples were selected randomly among the files of patients registered at the center from 2003-2007. The study sample consisted of individuals who had been residing in the city of Shiraz for at least five years. The Ethics Committee of Shiraz University of Medical Sciences approved the study.

The data of the 300 patients were used as cases in case-only design and case-control design. Moreover, for case-control design the

files of 300 people without breast or ovary cancer referring to the other divisions of Shahid Mottahari Clinic such as internal and surgical divisions were selected using convenient sampling method. Referrals to the other parts of the clinic may be considered population-based, because almost all socioeconomic groups refer to the clinic for specialist medical care.

### *Matched Case-Control Design*

An important concern in case-control studies is the difference that may exist between the subjects of case and control groups in terms of individual and exposure variables other than those being studied. A method to overcome this problem is the design of the study in a way that one can match the subjects of case and control groups in terms of the factors considered. For this reason individual matching was conducted. As age is a major risk factor for breast cancer, and recognized as a potential altering variable, matching was conducted on it. On the other hand, matching increases the efficiency of the study.<sup>16</sup>

In case-control studies, the main effects of non-matched variables and the interactions between these effects may be assessed. Therefore, we used conditional logistic regression analysis. First the case group files were selected randomly from Cancer Registry Center in Breast Cancer Clinic. Then for each case a matching control was selected using the ages of cases. For each case, a control woman, who had an age of  $\pm 3$  years difference from the case's age and had no ovary or breast cancer was selected.

### *Case-Only Design*

The use of case-only design, to assess gene-environment interaction was suggested by Piegorsch et al.<sup>17</sup> The selection of cases in case-only designs is the same as that in other case-control studies. Although with case-only design alone one cannot assess the independent effects of exposure and genotype, independence assumption of environmental exposure and genotype are the basic premise of this design.<sup>5</sup>

The entire sample was used for case-only analysis consisting of people with the disease.<sup>17</sup>

The size achieved from the case-only technique is interpreted as a deviation from the multiplicative relationship.<sup>5</sup> A sizable number of prevalent diseases are the result of the interactions between the genetic and environment factors.<sup>1</sup> The major advantage of case-only designs application in genetic epidemiology and assessment of the gene-environment interaction is the simplicity of collecting the required data, and decrease of calculations and financial costs. It seems that case-only designs are more efficient than the

traditional case-control designs, because they show more precise estimations, which is the result of less dispersion and more homogeneity. Therefore, in order to explore a specific odds ratio (OR) for interaction, the case-only designs need fewer cases than case-control studies. Moreover, the control group often has less motivation to participate in the study; therefore, the case-only design helps in minimizing the potential bias of participants. In case-only designs, data analysis is performed in a more straightforward way than in case-control designs. Although the case-only designs is not population-based, it uses simple sampling methods.<sup>18</sup>

The standard case-control analysis often has a weak power to explore multiplicative interactions, which are the results of the low numbers of cases and controls in matrix cells of genotype and exposure. The assumption of independence of gene-environment association results in a stronger estimation of interaction. However, the violation of this assumption results in an increased Type II error.<sup>19</sup>

The case-only design OR is calculated by multiplying the interaction (OR<sub>int</sub>) and OR of control group. If the independence assumption of gene and exposure in control group is valid and the disease is rare, the case-only OR measures interactional effect in a multiplicative model similar to the conventional case-control studies.<sup>17</sup>

To impose independence assumption, Weinberg and Umbach suggested a Maximum Likelihood Method based on log-linear model. They have shown that their method may need less than half of the individuals who do not have the gene-environment independence assumption.<sup>20</sup> In the studies of gene-environment interactions a specific genotype might be used. When the genetic marker data is not available, the family history data may be used as a proxy for genetic susceptibility; however, such a use may result in the possibility of significant misclassification.<sup>21</sup>

#### Independence Assumption

As Nicolle et al. stated clearly, the independence between gene and environment is central to valid interpretation of a case-only study.<sup>17</sup> In practice, controlling non-independency is not always easy. For example, the control of non-independence assumptions requires the knowledge of non-independence sources, which can be difficult or impossible to locate in some situations. It is difficult to control for sources of bias in cohort and case-control studies, therefore, it may also be difficult to control for the sources of bias in case-only studies. However, sensitivity analysis method, the benefits of which have been shown in case-control and cohort studies, may be used

in case-only studies.

As non-independence can be calculated in analysis, the case-only design may be a useful epidemiological instrument for examining gene-environment interactions.<sup>17</sup> In the following, a formula has been provided to describe the situation in which OR is concluded for the gene-environment associations. The formula can be used to estimate gene environment OR in source population.

$$G-E_{ORinD} = G-E_{ORinPop} \times \left( \frac{\left( \frac{1}{P(D/G-E-)} - 1 \right) \times \left( \frac{1}{P(D/G-E-)} - R_E \right)}{\left( \frac{1}{P(D/G-E-)} - R_G \right) \times \left( \frac{1}{P(D/G-E-)} - R_E \right)} \right)$$

In a previous study,<sup>22</sup> gene-environment independence assumptions was examined by Correlation or Chi-Square tests. In the present study, such an assumption was investigated using standard statistical multivariable techniques.<sup>17</sup>

In the equation, G-EOR<sub>inD</sub>-1 represents the OR of gene-environment in the control group and the G-EOR<sub>inPop</sub> represents the OR of gene-environment in the population. If C/PROR is equal to one, the control group can be used to estimate the OR of interaction of genetic and environmental factors in the population. In the above equation, P(D/G-E-) represents the disease baseline risks, which shows the likelihood of disease occurrence in people, who do not have the gene and environment factors. The RR<sub>GE</sub> or OR<sub>GE</sub> represents the OR of the disease in those who have both, the gene and the environmental factors (ie, two-factor interactions). The RR<sub>G</sub> or OR<sub>G</sub> represents the OR of the disease in people who have only the gene factor and RR<sub>E</sub> or OR<sub>E</sub> represents the OR of the disease in people who have only the environmental factor.<sup>17</sup>

The case-control and case-only designs were compared in terms variations in standard error, -2Log likelihood, and 95% confidence interval for of the gene-environment interactions.<sup>5</sup> Statistical analyses were performed using STATA 8.0 statistical software. The measured independent variables included continuous use of oral contraceptives for the past five years, breastfeeding history, number of pregnancies, age at menarche, age at the first delivery, and the history of breast cancer in the family. The family history of breast cancer in the first degree relatives including mother or sister was used as a proxy for the disease susceptibility.<sup>11</sup> Conditional Logistic Regression was used for data analysis of matched case-control design.

## Results

The age of the case group was  $51.8 \pm 0.5$  years. The OR of interaction between family history of breast cancer in the first degree relatives and other variables such as the first delivery at the age of 35 years and more, or no delivery, the history of having live birth, breastfeeding history, and oral contraception use were not statistically significant.

### Assessing Independency Assumption and Case-Only Study

The findings regarding the independency assumption of gene-environment are shown in table 1. There was independence between the gene-environment factors in the control group, therefore, in the second step we could use interaction analysis using case-only study for all variables.

The findings of assessment of GE-OR in the control group as a surrogate of GEOR in the population are shown in table 1. All control/population ratio of odds ratio were approximately close to 1.

### The Efficiency Comparison of Case-Control and Case-Only Designs

Odds ratio for GE interaction and the comparison of efficiency of case-control and case-only design are shown in table 2. Although all ORs are not significant, the standard error in estimating

the interaction of family history of breast cancer in the first degree relatives and other variables such as age at menarche, the first delivery at the age of 35 years and more, or no delivery, the history of having live birth, oral contraceptive use, or breastfeeding history in the case-only design were less than those found in the case-control design.

Confidence intervals in estimating the interaction of family history of breast cancer in the first degree relatives and other variables including in the case-only design were less wide than those in the case-control design. Moreover, the log likelihood for assessing the interaction of family history of breast cancer in the first degree relatives and other variables in the case-only design were smaller than those in the case control design.

The P values obtained in statistical analysis of the interaction of family history of breast cancer in the first degree relatives and variables like age at menarche, the first delivery at the age of 35 years and more, or no delivery, the history of having live birth, or breastfeeding history in the case-only design were smaller than those in the case-control design (table 2).

## Discussion

Based on a previous study,<sup>24</sup> the major risk of breast cancer among patients without the gene and the environmental factors, P (D/G-E-) was equal to

**Table 1:** The ratio of control to population odds ratio for the relation of family history to other variables of the participants

	RR <sub>GE</sub>	RR <sub>G</sub>	RR <sub>E</sub>	C/PROR
Family history and age at menarche	1.43 <sup>12</sup>	1.43 <sup>12</sup>	0.88 <sup>12</sup>	0.9988
Family history and age at the first delivery	1.03 <sup>4</sup>	1.43 <sup>12</sup>	1.02 <sup>12</sup>	1.0028
Family history and live birth history	0.77 <sup>12</sup>	1.43 <sup>12</sup>	0.93 <sup>12</sup>	1.0039
Family history and breastfeeding history	0.22 <sup>12</sup>	1.43 <sup>12</sup>	0.64 <sup>12</sup>	1.0056
Family history and contraceptive pills usage	1.01	1.43 <sup>12</sup>	0.83 <sup>23</sup>	1.0017

RR<sub>GE</sub>: Relative risk of joint effects of genetic and environmental factors; RR<sub>G</sub>: Relative risk of genetic factor; RR<sub>E</sub>: Relative risk of environmental factors; C/PROR: The ratio of odds ratio of control group to that of population. If C/PROR is equal to one, the control group can be used to estimate the odds ratio of interaction of genetic and environmental factors in the population.

**Table 2:** The efficiency comparison of case-control and case-only designs in estimating the genetic-environment interactions

	Design	Odds ratio ±SE	Confidence interval	-2likelihood ratio
Interaction of family history and age of more than 12 yrs at menarche	Case-control	0.85±1.05	0.08-9.6	384
	Case-only	0.76±0.43	0.24-2.34	247.8
Interaction of family history and the first delivery at ages between 25-34 yrs	Case-control	0.37±0.34	0.06-2.25	365
	Case-only	0.88±0.37	0.38-2.04	233.4
Interaction of family history and the first delivery at the age of 35 yrs and more or no delivery	Case-control	87160.61±	-.0	365
	Case-only	0.95±0.74	0.21-4.4	233.4
Interaction of family history and the history of having live birth	Case-control	5.27e+32	-.0	388.2
	Case-only	0.95±0.74	0.21-4.39	232.8
Interaction of family history and breastfeeding history	Case-control	1.39±1.62	0.14-13.62	420.2
	Case-only	0.8±0.32	0.4-1.7	266.8
Interaction of family history and oral contraceptive use	Case-control	1.01±0.009	0.996-1.00	437.8
	Case-only	0.999±0.002	0.993-1.00	267

There was no statistically significant interaction between the family history and other variables.

0.0066. This value is one of the assumptions of equation 1 for the calculation of gene-environment independence presented in table 1. Consistent with the findings of previous studies,<sup>11,12</sup> our study showed that in interaction analysis of family history of breast cancer and age at menarche, confidence intervals, standard error, and -2log likelihood in the case-only design were better than those in the case-control design. Moreover, similar to finding of Ardalan and colleagues<sup>12</sup> the P value in the case-only design was lower than that in the case-control design. However, such findings does not agree with those of Becher et al.<sup>11</sup> who showed otherwise.

Consistent with previous studies, the independence assumption of gene with age at menarche was established.<sup>11,12</sup> The confidence intervals and -2log likelihood of the interaction of family history of breast cancer in the first degree relatives and the first delivery at the ages between 25-34 yrs old in case-only design were better than those in the case-control design. However, standard error in the case-control design was somewhat higher.

Similar to the findings of Yavari et al.<sup>4</sup> the present study established the independence assumption of gene-age at the first delivery. They conducted the independence assumption based on independence test and suggested the standard statistical multivariable techniques to investigate the independence assumption of gene-environment. Therefore, similar to a study done in Japan,<sup>14</sup> we used this method to investigate the independence assumption of gene-environment. There was not significant interaction between the family history and the first delivery at the age of 35 yrs and more or no delivery. However, our study show that the sample size of 300 subjects to investigate interaction in case control studies is not enough, and that case-only designs might be a better design to examine such an interaction in such a sample size.

The independence assumption of the two factors was not established in a study by Ardalan and colleagues,<sup>12</sup> and therefore was removed from the case-only analysis. In agreement with the findings of Yavari et al.<sup>4</sup> the present study showed that case-only design had more efficiency than that of the matched case-control designs.

The interaction of family history of breast cancer and the history of having live birth was not statistically significant in the case-control design. Since the independence assumption was not established, the interaction of these two factors in case-only design was not estimated in the study by Yavari and colleagues,<sup>4</sup> or was not reported in Becher et al. studies.<sup>11</sup> The present study showed that the confidence interval, standard error and -2log likelihood estimations in case-only design were smaller than those in

case-control analogues. These findings are similar to that of Ardalan and colleagues,<sup>12</sup> and is an indication of more efficacy of case-only designs than the case-control ones.

The confidence interval, standard error, -2log likelihood and p-value estimations in case-only design were smaller than those in the case-control design. This shows that the case-only design had more efficacy than the case-control one. The confidence interval, standard error, and -2log likelihood estimations in the case-only design were smaller than those in the case-control design. Such findings are a sign that case-only designs are more efficacious than the case-control design. However, the P value in the case-only design was higher than that in the case-control design.

The reason for not detecting the interaction between the family history of breast cancer and other variables might be the small sample size of the participants in case-control design study. In case-only studies, the basic assumption of gene-environment independence in non-diseased group should be established. In the studies before 2004, this assumption used to be established using classic statistical tests such as Chi Square, Correlation Coefficient and so on. However, such an approach was also criticized,<sup>25</sup> for the lack of ability of the use of GE-OR in control groups to estimate the GE-Or in populations. Gatto et al.<sup>17</sup> described a modification of the methods. In order to establish the independence assumption, they introduce the standard statistical multivariable techniques, which have resolved the previous shortcoming of the design, and have led to widespread use of the techniques.

The studies of Yavari et al.<sup>4</sup> and Becher et al.<sup>11</sup> used classic methods to test the independence assumption. Therefore, this study used the new method to estimate the independence assumption. The present study did not identify genes involved in breast cancer; therefore, similar to previous studies,<sup>4,11,12</sup> the family history of breast cancer, as an alternative for genetic mutations, was used. Our approach is similar to other studies using the familial history as alternatives for genetic factors to study the interactions of genetic factors in pulmonary,<sup>14</sup> and colon,<sup>13</sup> cancers.

## Conclusion

Matched case-control studies require a sample higher than 300 subjects, which was used in the present study, to examine a logical interaction. Considering the smaller standard error and -2log likelihood ratio of the case-only design than those of case-control design, we might be able to suggest that the case-only design is a better method to examine the interactions between the genetic and

environmental variable involved in breast cancer.

### Acknowledgement

This study was supported by a grant from Vice-chancellor for Research, Shiraz University of Medical Sciences. This publication constitutes a part of Rahmatollah Moradzadeh's MSc thesis (grant no 87-3992). The authors thank Mr Haidar Nadrian for his kind cooperation in translating the paper to English.

**Conflict of Interest:** None declared

### References

- Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered.* 2007;63:111-9. doi: 10.1159/000099183. PubMed PMID: 17283440.
- Szklo M, F Javier Nieto. *Epidemiology beyond the basics.* 2nd ed. USA: Jones and Bartlett Publishers; 2007. p. 183.
- Last JM. *A Dictionary of Epidemiology.* Translated By: Nasery Q. 1st ed. Tehran: Tehran University Published, 1989. p. 106 [Persian]
- Yavari P, Pourhosseingholi MA. Estimation of the Gene-Environment Interaction in Breast Cancer Patients. *Iranian Journal of Epidemiology.* 2006;2:49-52.[Persian]
- Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology.* 3rd ed. Lippincott Williams & Wilkins; 2008. p. 573-5.
- Attar Parsaie F, Golchin M, Asvadi I. A STUDY OF RELATIONSHIP BETWEEN DEMOGRAPHICS, LIFE-STYLE, STRESSFUL LIFE-EVENTS AND BREAST CANCER IN WOMEN. *MEDICAL JOURNAL OF TABRIZ UNIVERSITY OF MEDICAL SCIENCES.* 2001;50:15-21.[Persian]
- Qin LQ, Xu JY, Tezuka H, Li J, Arita J, Hoshi K, Sato A. Consumption of commercial whole and non-fat milk increases the incidence of 7,12-dimethylbenz(a)anthracene-induced mammary tumors in rats. *Cancer Detect Prev.* 2007;31:339-43. doi: 10.1016/j.cdp.2007.04.010. PubMed PMID: 17935906.
- Fathi Najafi T, Jabbarzadeh Ganjeh S, Mojahedi Rezaeian S, Mazloom S.R. Assessment of Certain Breast Cancer Risk Factors During Reproductive Age in Women in Mashhad(2002-2003). *Razi Journal of Medical Sciences.* 2004;11:577-85.[Persian]
- Azizi F, Hatami H, Janghorbani M. *Epidemiology and control of common diseases in Iran.* 2nd ed. Tehran: Khosravi; 2002. p. 201-5. [Persian]
- Vogel VG, Taioli E. Have we found the ultimate risk factor for breast Cancer? *J Clin Oncol.* 2006;24:1791-4. doi: 10.1200/JCO.2005.05.4122. PubMed PMID: 16567765.
- Becher H, Schmidt S, Chang-Claude J. Reproductive factors and familial predisposition for breast cancer by age 50 years. A case-control-family study for assessing main effects and possible gene-environment interaction. *Int J Epidemiol.* 2003;32:38-48. doi: 10.1093/ije/dyg003. PubMed PMID: 190006.
- Ardalan A, Holakouie naieni K, Mahmoudi M, Majdzadeh R, Derakhshandeh Peykar P. Comparison of the power and statistical efficiency of case-control and case-only designs in estimating the gene-environment interaction in breast cancer. *Journal of School of Public Health and Institute of Public Health Research.* 2006;4:7-20.[Farsi]
- Fernandez E, Gallus S, La Vecchia C, Talamini R, Negri E, Franceschi S. Family history and environmental risk factors for colon cancer. *Cancer Epidemiol Biomarkers Prev.* 2004;13:658-61. PubMed PMID: 15066934.
- Tsuchiya M, Iwasaki M, Otani T, Nitadori J, Goto K, Nishiwaki Y, et al. Breast cancer in first-degree relatives and risk of lung cancer: assessment of the existence of gene sex interactions. *Jpn J Clin Oncol.* 2007;37:419-23. doi: 10.1093/jjco/hym048. PubMed PMID: 17586847.
- Dempfle A, Scherag A, Hein R, Beckmann L, Chang-Claude J, Schäfer H. Gene-environment interactions for complex traits: definitions, methodological requirements and challenges. *Eur J Hum Genet.* 2008;16:1164-72. doi:10.1038/ejhg.2008.106. PubMed PMID:18523454.
- Gordis L. *Epidemiology.* 2nd ed. Philadelphia: W.B. Saunders Company; 2000. p. 148-9.
- 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 of Epidemiol.* 2004;33:1014-24. doi: 10.1093/ije/dyh306. PubMed PMID: 15358745.
- Ahrens W, Pigeot I. *Handbook of epidemiology.* Germany: Springer-Verlag Berlin Heidelberg; 2005. p. 338-41.
- Mukherjee B, Ahn J, Gruber SB, Rennert G, Moreno V, Chatterjee N. Tests for gene-environment interaction from case-control data: a novel study of type I error, power and

- designs. *Genet Epidemiol.* 2008;32:615-26. doi: 10.1002/gepi.20337. PubMed PMID: 18473390.
- 20 Goldstein AM, Andrieu N. Detection of interaction involving identified genes: available study designs. *J Natl Cancer Inst Monogr.* 1999;49:54. doi: 10.1093/oxfordjournals.jncimonographs.a024226. PubMed PMID: 10854486.
- 21 Khoury MJ, Beaty TH. Applications of the case-control method in genetic epidemiology. *Epidemiol Rev.* 1994;16:134-50. PubMed PMID: 7925722.
- 22 Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol.* 2002;155:478-84. doi: 10.1093/aje/155.5.478. PubMed PMID: 11867360.
- 23 Al-Shaibani H, Bu-Alayyan S, Habiba S, Sorkhou E, Al-Shamali N, Al-Qallaf B. Risk factors of breast cancer in Kuwait: Case-Control study. *IJMS.* 2006;31:61-4.
- 24 Motamed N, Hadi N, Talei A. A Survey on risk factors of breast cancer in women over 35 years of age (Shiraz -2000). *Journal of Zanjan University of Medical Sciences and Health Services.* 2004;12:25-32.[Farsi]
- 25 Hopper JL. Commentary: Case-control-family designs: a paradigm for future epidemiology research? *Int J Epidemiol.* 2003;32:48-50. doi: 10.1093/ije/dyg114. PubMed PMID: 12690007.