A Case-Only Approach for Assessing Gene by Sex Interaction in Human Longevity

Qihua Tan,1 Anatoli I. Yashin,1 Else M. Bladbjerg,2 Moniek P.M. de Maat,2 Karen Andersen-Ranberg,3 Bernard Jeune,3 Kaare Christensen,3 and James W. Vaupel1

1Max Planck Institute for Demographic Research, Rostock, Germany. 2Department of Thrombosis Research, University of Southern Denmark–Odense University and Department of Clinical Biochemistry, Ribe County Hospital, Esbjerg, Denmark. 3Department of Epidemiology, Institute of Public Health, and Aging Research Center, University of Southern Denmark–Odense University.

As one aspect of the complex feature of longevity, gene by sex interaction plays an important role in influencing human life span. With advances in molecular genetics, more studies aimed at assessing gene by sex interaction are expected. New and valid statistical methods are needed. In this article, we introduce a nontraditional approach, the case-only design, which was originally proposed for assessing gene and disease associations, to detect gene by sex interaction in human longevity. Applications of this method to data collected from centenarian studies show that it can produce consistent results as compared with results obtained from case-control and other approaches. The method cannot be used as a substitute for traditional case-control studies since it is limited to the detection of interactions only. However, the easily applicable case-only approach can be an important tool for screening many potential genes that contribute to human longevity.

A GREAT deal of interest has been generated in the study of genetic influences on human longevity because of rapid developments in molecular genetics (1). In recent literature, the gene by sex interaction arises as an important phenomenon in the genetic modulation of human life span. For example, Ivanova and colleagues (2) reported on the correlation between HLA-DR7 and longevity, with an elevated frequency of DR7 in long-lived men. De Benedictis and colleagues (3) observed a significant decrease in the frequency of the tyrosine hydroxylase (THO) large allele group (alleles 9, 10-1, 10) in male Italian centenarians, but not in females. As concerning statistical analysis, a conventional gene frequency method based on the case-control design has been used (2–4). A relative risk approach that combines sex-specific population survival distributions has been proposed and applied to data from cross-sectional studies to detect the risk of gene alone as well as the risk of gene by sex interaction that potentially contribute to human life-span heterogeneity (5,6). Comparisons between the two approaches have been made (5,7,8) that reveal a better performance for the latter, because it makes full use of individual genetic as well as survival information. However, both approaches have difficulty in dealing with crucial issues that originate from the cross-sectional design. Spurious conclusions could be made when improper control is chosen. In this article, we introduce a nontraditional approach, the case-only method, which was originally designed for analyzing gene by environment (9–11) and gene by gene (12) interactions in disease etiology to detect gene by sex interaction in human longevity when centenarians were treated as cases. This approach appears to have greater precision and requires fewer case subjects than the traditional case-control study when the primary interest is in gene by sex interaction (9,11,13). We show that the same method is also a valid approach that gives consistent results on reported gene by sex interactions from previous gene longevity association studies. Special issues that come up when the model is applied to the longevity studies are discussed, with advantages of the application highlighted.

METHODS

Suppose that we are interested in analyzing the genetic influence on longevity that is sex dependent by using case-control design. We can display our data in a $2 \times 4$ table (Table 1). In Table 1, we classify the genotype as presence (1 or 2 alleles) or absence (0 allele) of the susceptible gene. We also assign 1 for males and 0 for females. As concerning statistical analysis, a conventional gene frequency method based on the case-control design has been used (2–4). A relative risk approach that combines sex-specific population survival distributions has been proposed and applied to data from cross-sectional studies to detect the risk of gene alone as well as the risk of gene by sex interaction that potentially contribute to human life-span heterogeneity (5,6). Comparisons between the two approaches have been made (5,7,8) that reveal a better performance for the latter, because it makes full use of individual genetic as well as survival information. However, both approaches have difficulty in dealing with crucial issues that originate from the cross-sectional design. Spurious conclusions could be made when improper control is chosen. In this article, we introduce a nontraditional approach, the case-only method, which was originally designed for analyzing gene by environment (9–11) and gene by gene (12) interactions in disease etiology to detect gene by sex interaction in human longevity when centenarians were treated as cases. This approach appears to have greater precision and requires fewer case subjects than the traditional case-control study when the primary interest is in gene by sex interaction (9,11,13). We show that the same method is also a valid approach that gives consistent results on reported gene by sex interactions from previous gene longevity association studies. Special issues that come up when the model is applied to the longevity studies are discussed, with advantages of the application highlighted.
Table 1. Gene by Sex Interaction Based on Case-Control Design

<table>
<thead>
<tr>
<th>Sex *</th>
<th>Genotype †</th>
<th>Cases</th>
<th>Controls</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>a₀₀</td>
<td>b₀₀</td>
<td>O₁₀ = 1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>a₀₁</td>
<td>b₀₁</td>
<td>O₁₀ = a₁₀b₀₀/a₀₀b₀₁</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>a₁₀</td>
<td>b₁₀</td>
<td>O₁₀ = a₁₀b₀₀/a₀₀b₁₀</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>a₁₁</td>
<td>b₁₁</td>
<td>O₁₁ = a₁₁b₀₀/a₀₀b₁₁</td>
</tr>
</tbody>
</table>

Note: OR = odds ratio.
*1, male; 0, female.
†1, present; 0, absent.

where OR ca is the control-only odds ratio, b₁₁/b₁₀/b₀₁. Assuming that the susceptible genotype and sex are independent and that the event (longevity) the interaction is associated with is rare, then OR ca becomes unity (see the Appendix), and we can rewrite Equation (1) as

\[
\text{OR}_{ca} = \frac{\text{OR}_{11}}{\text{OR}_{10}\text{OR}_{01}}.
\]

Here Equation (2) means that the case-only odds ratio measures the departure from multiplicative joint effect of the genotype and sex or, in other words, the effect of gene by sex interaction. The null hypothesis for this approach is H₀: OR ca = 1. Any statistically significant deviation of OR ca from unity indicates that there is a gene by sex interaction that contributes to and modifies the probability of achieving longevity. By comparing variates of the maximum likelihood estimate of the logarithm of Equations (1) and (2), Pie-gorgsch and colleagues (9) concluded that the case-only study has increased precision in estimating interactions because the variance corresponding to Equation (1) involves an extra component for OR ca, the control-only odds ratio.

A statistical test for the null hypothesis can be conducted by using the log likelihood ratio (LLR) test by calculating twice the difference between the log likelihoods at OR ca = 1 and at OR ca estimated. When the sample size is large, the LLR is approximately distributed as χ² on one degree of freedom. Alternatively, we can apply the standard χ² statistic to test the null hypothesis (Table 2), that is, \[ χ² = -2\left[\ln(a₁₁a₀₀ − a₀₁a₁₀) − \ln(n₁n₂n₁n₂)\right] \] with one degree of freedom. Here n₁, n₂, n₁, n₂ is the marginal sum of the observations by sex and by genotype, n is the total sum of the observations (cases). The χ² statistic with continuity correction can also be applied but is not recommended (14). For a significant OR ca, confidence intervals can be constructed. The procedure is first to construct a confidence interval for the natural log of OR ca and then to exponentiate the boundaries to get the confidence intervals for OR ca.

Table 2. Case-Only 2 × 2 Table Classified by Genotype and Sex

<table>
<thead>
<tr>
<th>Sex *</th>
<th>0</th>
<th>1</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>a₀₀</td>
<td>a₀₁</td>
<td>n₀</td>
</tr>
<tr>
<td>1</td>
<td>a₁₀</td>
<td>a₁₁</td>
<td>n₁</td>
</tr>
<tr>
<td>Sum</td>
<td>n₀</td>
<td>n₁</td>
<td>n</td>
</tr>
</tbody>
</table>

*1, male; 0, female.
†1, present; 0, absent.

Table 3. Case-Only Approach for Assessing Gene by Sex Interactions in the Danish Centenarian Study

<table>
<thead>
<tr>
<th>Gene</th>
<th>Male 0</th>
<th>Female 0</th>
<th>Sum</th>
<th>OR</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angiotensinogen</td>
<td>12</td>
<td>34</td>
<td>64</td>
<td>72</td>
<td>182</td>
<td>2.519</td>
</tr>
<tr>
<td>FVII Q353</td>
<td>34</td>
<td>12</td>
<td>120</td>
<td>18</td>
<td>184</td>
<td>2.353</td>
</tr>
<tr>
<td>FVII–323p10</td>
<td>33</td>
<td>13</td>
<td>117</td>
<td>20</td>
<td>183</td>
<td>2.305</td>
</tr>
</tbody>
</table>

Notes: 1, present; 0, absent; OR = odds ratio; CI = confidence interval.
was set up in France (2). A total of 533 centenarians (89 males and 444 females) were pooled for the study (Table 4). The allele frequency of HLA-DR7 was observed at 14% in the control group but at 21.9% in male and 14% in female centenarians. When allele frequency differences were compared between centenarians and the control group, the HLA-DR7 allele frequency was found to be significantly increased in male but not in female centenarians; OR = 1.72, 95% confidence interval (CI) 1.19–2.5, \( p = .004 \) for males. The conclusion was that the DR7 allele has a beneficial effect on male longevity (2). Here we examine the sex-dependent influence of HLA-DR7 by applying the case-only approach. Table 4 is arranged in complete conformity with Table 2, with numbers calculated according to the allele frequency and total number of centenarians available from the article (2). The case-only odds ratio is estimated as 1.816 (\( p = .013 \)) with a 95% CI from 1.129 to 2.923. This result again indicates that the HLA-DR7 allele favors male longevity, which is compatible with the conclusion drawn from the case-control approach.

### Discussion

As one aspect of the complex feature of human longevity, the gene by sex interaction is an important phenomenon that should be addressed. The rapid advance in molecular technology is leading to the relative ease of searching for a large number of DNA markers at several candidate gene loci. With large amounts of individual genetic information available, new and efficient statistical methods are needed to help search for important genes that play crucial roles in the various pathways constituting the network of human longevity. In this regard, the easily applicable case-only approach can serve as a valid and useful way for screening gene by sex interactions in human longevity. Because the case-only approach does not use control subjects, crucial issues in the choice of an appropriate control group that have been perplexing with regard to case-control study are avoided (10). This is important because the improper choice of a control group could lead to spurious conclusions that distort the study. In addition, this approach has greater precision in estimating interactions than the traditional case-control design (9,11,13). However, there are important assumptions that underlie the application of the model (9,10,16,17).

First, in order to apply this method, researchers must assume that sex and the genotype are independent. This assumption holds for any autosomal genes because their segregation does not depend on sex. However, one must note that for sex-linked genes, such an application definitely violates the basic assumption. When our primary interest is gene by sex interaction, such a valid assumption is much preferable to the weak assumptions underlining the traditional case-control study, such as no population stratification (for example, cases and controls are ethnically different), no cohort effects, and so on.

Second, the event (longevity) associated with gene by sex interaction should be rare. The study of longevity fits into this assumption because longevity by definition is always a rare event. It was estimated that there were approximately 44 centenarians per million population in the developed countries in 1990 (18). Although the number is increasing very rapidly (19), according to a United Nations’ prediction, in the year 2050 centenarians will comprise approximately 1% of the total population in Japan, which has the highest number of centenarians in the world. Schmidt and Schaid (20) showed that the cross-product computed from case-only data may be substantially smaller than the odds ratio calculated from a case-control study. This would underestimate the true effect when the risk of event associated with the gene is relatively high. Fortunately, we do not have to worry about the problem in a longevity study because we are always dealing with the small proportion in the population who managed to achieve extraordinarily long lives.

Because the context of the centenarian study fits into the case-only approach for assessing gene by sex interaction, we believe that the method is valid and should be promoted. However, one has to keep in mind that the case-only approach makes sense only when the primary interest is in estimating the possible sex-dependent effect from the susceptible gene. The odds ratio estimated from the case-only approach measures only the departure of the overall effect of both the gene in question and sex from the multiplicative effect. Because no effect of gene or sex alone is estimated in this approach, it does not provide any definite information on the survival of particular genotype carriers. Thus, although the case-only approach for assessing gene by sex interaction in longevity can be used as a tool for preliminary screening of many candidate genes, it cannot be used as a substitute for the traditional case-control studies. A better strategy could be that after screening the candidate genes for gene by sex interaction, we can then fit survival models to the data and estimate survival functions for males and for females separately (5,6). In this way, application of this easily applicable method can help to rapidly increase efficiency for future longevity studies. Also, it is necessary to point out that, like other association studies, the case-only approach also has difficulty in the situation when linkage disequilibrium exists (10,16). The detected interaction could be due to the fact that the marker is in linkage disequilibrium with the real gene that is relevant to longevity. Nevertheless, such an association approach can complement future research aimed at localizing the specific gene loci. At this point, one also has to be aware that, by longevity relevant genes, we only mean those genes whose action increases physiological ca-
pacity or reserve and thus indirectly increases the potential of longevity (21,22), because there is no gene that is solely responsible for longevity.

The importance of independence between exposure (sex in our context) and genotype in applying this method has been addressed by previous researchers (10,17). Recently, Albert and colleagues (23) showed that inferences from the case-only design can be highly distorted when there is departure from the independence assumption. Although in the context of gene by sex interaction, such an assumption holds provided the genes of interest are autosomal, caution has to be paid when we want to apply the same approach to study gene by environment interactions (Table 5). In the latter situation, especially when the environment relates to the geographical allocation, it is important to check the ethnic origin of the populations to make sure that the assumption is satisfied. The choice of subject in the case-only study should follow the usual rules of case selection for any case-control study (16,24). However, given the importance of centenarian studies, the case-only approach is a promising tool for finding important genes that contribute to human health and longevity.

Acknowledgments

We are thankful to G. De Benedictis from the Cell Biology Department, University of Calabria, Rende, Italy and to Jinhua Zhao from the Department of Psychological Medicine, Institute of Psychiatry, St. Bartholomew’s Hospital, London, UK and Royal London School of Medicine and Dentistry for useful discussions. We also thank Dr. Karl Brehmer for help in preparing the manuscript.

Address correspondence to Anatoli I. Yashin, Max Planck Institute for Demographic Research, Doberaner Strasse 114, 18057 Rostock, Germany. E-mail: Yashin@demogr.mpg.de

Table 5. Case-Only Approach for Assessing Gene by Environment Interaction for Simulated Data

<table>
<thead>
<tr>
<th>Area</th>
<th>Genotype*</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>Sum</td>
<td></td>
</tr>
<tr>
<td>North</td>
<td>6378</td>
<td>1617</td>
<td>7995</td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>4757</td>
<td>548</td>
<td>5305</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>11135</td>
<td>2165</td>
<td>13300</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Odds ratio = 0.454; 95% confidence interval = 0.406–0.499; p = 0.001

*1, present; 0, absent.

References


Received April 17, 2001
Accepted November 8, 2001
Decision Editor: John A. Faulkner, PhD

Appendix

In accordance with Table 1, let \( S = 1 \) stand for males and \( S = 0 \) for females, and let \( G = 1 \) stand for carriers of the susceptible genotype and \( G = 0 \) for noncarriers of the genotype. Let \( L \) stand for cases (longevity) and \( \bar{L} \) for controls. By treating sex as an outcome and following Piepersch and colleagues (9), we find that the control-only odds ratio is

\[
OR_{co} = \frac{P(S = 1 | \bar{L}, G = 1)P(S = 0 | \bar{L}, G = 0)}{P(S = 1 | L, G = 0)P(S = 0 | L, G = 1)}
\]  

The independence between genotype and sex means
In Equation (4),
\[
\frac{P(S = 1| G = 1)}{P(S = 0| G = 1)} = \frac{P(S = 1| G = 0)}{P(S = 0| G = 0)}.
\]

Then we have \(P(S = 1| G = 1) \approx P(S = 1| \overline{L}, G = 0)\). Do the same for the rest in Equation (4) and substitute to yield
\[
\frac{P(S = 1| G, L, L, G = 1)}{P(S = 0| G, L, L, G = 1)} = \frac{P(S = 1| \overline{L}, G, L, G = 1)}{P(S = 0| \overline{L}, G, L, G = 1)} = 1.
\]