



Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research
Doberaner Strasse 114 · D-18057 Rostock · GERMANY
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;
<http://www.demogr.mpg.de>

MPIDR WORKING PAPER WP 2001-012
MAY 2001

A Case-only Approach for Assessing Gene-sex Interaction in Human Longevity

Qihua Tan (tan@demogr.mpg.de)
Anatoli I. Yashin (yashin@demogr.mpg.de)
Else M. Bladbjerg
Moniek P.M. de Maat
Karen Andersen-Ranberg
Bernard Jeune
Kaare Christensen
James W. Vaupel (jwv@demogr.mpg.de)

© Copyright is held by the authors.

Working papers of the Max Planck Institute for Demographic Research receive only limited review. Views or opinions expressed in working papers are attributable to the authors and do not necessarily reflect those of the Institute.

A Case-only Approach for Assessing Gene-sex Interaction in Human Longevity

Qihua Tan¹, Anatoli I. Yashin*¹, Else M. Bladbjerg², Moniek P.M. de Maat², Karen Andersen-Ranberg³, Bernard Jeune³, Kaare Christensen³, James W. Vaupel¹

¹ Max-Planck Institute for Demographic Research, Rostock, Germany

² Department of Thrombosis Research, University of Southern Denmark and Department of Clinical Biochemistry, Ribe County Hospital in Esbjerg, Denmark

³ Epidemiology, Institute of Public Health, and Aging Research Center, University of Southern Denmark-Odense University, Denmark

*Address for correspondence

Prof. Anatoli I. Yashin

Max-Planck Institute for Demographic Research

Doberaner Str. 114

18057 Rostock, Germany

Tel. +49 381 2081 106 Fax +49 381 2081 406

e-mail: Yashin@demogr.mpg.de

ABSTRACT

As one aspect of the complex feature of longevity, gene-sex interaction plays an important role in influencing human life span. With advances in molecular genetics, more studies aimed at assessing gene-sex interaction are expected. New and valid statistical methods are needed. In this paper, we introduce a nontraditional approach, the case-only design, which was originally proposed for assessing gene and disease associations, to detect gene-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. Important features of the application in human longevity studies are highlighted and discussed. Since centenarians constitute a special population representing successful ageing, the easily applicable case-only approach will be an important tool for screening potential major genes that contribute to human longevity.

Key words: case-only study, centenarian, gene-sex interaction

Introduction

A great deal of interest has been generated in the study of genetic influences on human longevity taking advantage of the rapid development in molecular genetics (1). In the recent literature, the gene-sex interaction arises as an important phenomenon in the genetic modulation of human life span. For example, Ivanova et al. (2) reported the correlation between HLA-DR7 and longevity with elevated frequency of DR7 in long-lived men. De Benedictis et al. (3) observed a significant decrease in the frequency of the THO (tyrosine hydroxylase) 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 (2,3,4). 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-sex interaction that potentially contribute to human life span heterogeneity (5,6). Comparisons between the two approaches have been made (7,8,5) revealing better performance for the latter since 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 paper, we are going to introduce a nontraditional approach, the case-only method, which was originally designed for analysing gene-environment interaction in disease etiology (9,10) to detect gene-sex interaction in human longevity when treating centenarians 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-sex interaction (9,11). We are going to show that the same method is also a valid approach that gives consistent results on reported gene-sex

interactions from previous gene longevity association studies. Special issues that come up when applying the model to the longevity studies will be discussed, with advantages of the application highlighted.

Methods

Suppose that we are interested in analysing the genetic influence on longevity that is sex-dependent using case-control design. We can display our data in a 2×4 table (Table 1). In Table 1, we classify the genotype as presence (1 or 2 alleles) and absence (0 allele) of the susceptible gene. We also assign 1 for males and 0 for females. Using females without the susceptible allele as the reference group, we can calculate the odds ratio for the other groups with different sex and genotype compositions. In Table 1, OR_{01} is the odds ratio for female carriers as compared to female non-carriers of the allele or, in other words, the effect of the gene alone; OR_{10} is the odds ratio for sex; OR_{11} is odds ratio for the combined effect for both sex and genotype. The case-only odds ratio (Table 2) can be calculated as the cross-product

$$OR_{ca} = \frac{a_{11}a_{00}}{a_{10}a_{01}} = \frac{OR_{11}}{OR_{10}OR_{01}} OR_{co} \quad (1)$$

where OR_{co} is the control-only odds ratio, $\frac{b_{11}b_{00}}{b_{10}b_{01}}$. Assuming that the susceptible genotype and sex are independent and that the event (longevity) the interaction is associated with is rare, then OR_{co} becomes unity (see Appendix), and we can rewrite (1) as

$$OR_{ca} = \frac{OR_{11}}{OR_{10}OR_{01}}. \quad (2)$$

(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-sex interaction. The null hypothesis for this approach is $H_0:OR_{ca} = 1$. Any statistically significant deviation of OR_{ca} from unity indicates that there is a gene-sex interaction that contributes to and modifies the probability of achieving longevity. By comparing variances of the MLE of the logarithm of (1') and (2'), Piegorsch et al. (9) concluded that the case-only study has increased precision in estimating interactions since the variance corresponding to (1') involves an extra component for OR_{co} , the control-only odds ratio.

Statistical test for the null hypothesis can be conducted by employing 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 sample size is large, the LLR is approximately distributed as χ^2 on 1 degree of freedom. Alternatively, we can apply the standard χ^2 statistic to test the null hypothesis (Table 2), i.e.

$$\chi^2 = \frac{n_{..}(a_{11}a_{00} - a_{10}a_{01})^2}{n_{1.}n_{2.}n_{.1}n_{.2}}$$

with one degree of freedom. Here, $n_{1.}$, $n_{2.}$, $n_{.1}$, $n_{.2}$ is the

marginal sum of the observations by sex and by genotype, $n_{..}$ is the total sum of the observations (cases). The χ^2 statistic with continuity correction can also be applied but is not recommended (12). 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 exponentiate the boundaries to get the confidence intervals for

$$OR_{ca}. \text{ The standard error of } \ln(OR_{ca}) \text{ is given by } \sqrt{\frac{1}{a_{00}} + \frac{1}{a_{01}} + \frac{1}{a_{10}} + \frac{1}{a_{11}}}. \text{ Since the}$$

natural log of odds ratio is more normally distributed than the odds ratio itself, we can use the critical values of the standard normal distribution for calculating the intervals.

Examples

(1) Cardiovascular disease associated genes and longevity

In order to explore if genes involved in the etiology of cardiovascular disease are also associated with human longevity, a cross-sectional investigation was conducted in Denmark with blood samples taken from controls (healthy young subjects) and centenarians (13). Results from the data analysis show that there is no significant influence on longevity from each of the genes alone (13,6). However, reanalysis of the data with consideration of gene-sex interaction has revealed that 3 genetic polymorphisms, angiotensinogen M/T235, FVII R/Q353 and FVII-323ins10, manifest significant sex-dependent effects that are favourable to the longevity of males (6). By setting the risk of female carriers to 1, the relative risks for male carriers of the angiotensinogen M235, FVII Q353, and FVII –323p10 alleles were estimated as 0.67, 0.68 and 0.69, respectively. All are highly significant. Applying the case-only approach to centenarian data on the 3 alleles, we got consistent results as shown in Table 3. The odds ratios for the 3 alleles are 2.519 for angiotensinogen M235 ($p=0.013$), 2.353 for FVII Q353 ($p=0.038$), 2.305 for FVII-323p10 ($p=0.037$). All are significantly different from unity, which means that male centenarians are more likely to carry these mutants than female centenarians. The results suggest that these mutant alleles convey gene-sex interactions that are beneficial to male survival. We also tried to do the same calculation on other genes tested in the study but no further gene-sex interaction was found.

Although the results are compatible, the previous analysis by Tan et al. (6) relies on the proportional hazard assumption, which may not be an appropriate approach since it ignores the possible antagonistic genetic effect in the process of ageing. This problem is avoided in the case-only study because we are drawing our conclusions only from the centenarians.

(2) HLA-DR7 and human longevity

In an effort to examine the contribution of HLA-DRB1 polymorphisms to human longevity, a case-control study 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 percent in the control group, but 21.9 percent in male and 14 percent in female centenarians. By comparing allele frequency differences between centenarians and the control group, the HLA-DR7 allele frequency was found significantly increased in male but not in female centenarians (OR=1.72, 95%CI: 1.19-2.5, p=0.004 for males) with conclusion 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=0.013) with a 95% confidence interval from 1.129 to 2.923. This result again indicates that HLA-DR7 allele favours male longevity, which is compatible with the conclusion drawn from the case-control approach.

Discussions

As one aspect of the complex feature of human longevity, the gene-sex interaction is an important phenomenon that needs to 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-sex interactions in human longevity. Since the case-only approach does not employ 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 since 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). However, there are important assumptions that underlie the application of the model.

First, in order to apply this method, researchers must assume that sex and the genotype are independent. This assumption holds for any autosomal genes since 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-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-sex interaction should be rare. The study of longevity fits into this assumption since longevity by definition is

always a rare event. It was estimated that there were about 44 centenarians per million population in the developed countries in 1990 (14). Although the number is increasing very rapidly (15), according to a United Nations' prediction, in the year 2050 centenarians will comprise only about 1 percent of the total population in Japan, which has the highest number of centenarians in the world. Schmidt and Schaid (16) showed that the cross-product computed from case-only data may be substantially smaller than the odds ratio calculated from case-control study. This would underestimate the true effect when risk of event associated with the gene is relatively high. Fortunately, we do not have to worry about the problem in longevity study since we are always dealing with the small proportion in the population who managed to achieve extraordinary long lives.

Since the context of centenarian study fits into the case-only approach for assessing gene-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. No effect of gene or sex alone is estimated in this approach. 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. 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 localising the specific gene loci.

The case-only approach was originally derived to assess gene-environment interactions in the etiology of complex diseases. Can we simply apply it for analysing gene-environment interaction in longevity? In the context of human longevity, we think that such an application should be made with caution. For illustration, we simulated case-only data from a centenarian study conducted in Italy where subjects were collected from the south and the north. Suppose that we are interested in analysing the gene-environment interaction for a susceptible genotype but with different genotype frequencies in the south (0.1) and the north (0.2) due to different ethnic origins. However, in a case-only study this difference is not known, and we assume that it is ignored. We also assume that the genotype alone is neutral and has no gene-environment interaction affecting longevity. We simulated 13,300 centenarians collected from the south and the north (Table 5). The genotype frequency among the centenarians is estimated as 0.1 in the south and 0.2 in the north, which does not show any deviation from the frequencies set for the simulation. However, when the case-only approach aimed at detecting gene-environment interaction is applied, we got an OR of 0.454 (95% CI: 0.406-0.499, $p=0$). The result appears to be a strong gene-environment interaction that favours northerners, which is contrary to the real situation. This example demonstrates the importance of the basic assumptions that underlie the case-only approach. In order to study gene-environment interaction in longevity, especially when the environment relates to geographical allocation, it is of crucial importance to check the ethnic origin of the populations to ensure that the assumptions are satisfied.

The choice of subject in case-only study should follow the usual rules of case selection for any case-control study (17). However, since centenarians constitute the special population representing successful ageing, the genetic studies on centenarians

could help to find important genes that contribute to human health and longevity. In this sense, the case-only approach is a promising tool for this purpose.

Appendix

In accordance with Table 1, let $S = 1$ stand for males and $S = 0$ for females, $G = 1$ for carriers of the susceptible genotype and $G = 0$ for non-carriers of the genotype. Let L stand for cases (longevity) and \bar{L} for controls. By treating sex as outcome and following Piegorsch et al. (9), the control-only odds ratio is

$$OR_{co} = \frac{P(S = 1|\bar{L}, G = 1)P(S = 0|\bar{L}, G = 0)}{P(S = 1|\bar{L}, G = 0)P(S = 0|\bar{L}, G = 1)}. \quad (3')$$

The independence between genotype and sex means

$$\frac{P(S = 1|G = 1)}{P(S = 0|G = 1)} = \frac{P(S = 1|G = 0)}{P(S = 0|G = 0)}. \quad (4')$$

In (4')

$$P(S = 1|G = 1) = P(S = 1|\bar{L}, G = 1)P(\bar{L}|G = 1) + P(S = 1|L, G = 1)P(L|G = 1).$$

Since longevity is a rare event, $P(L|G = 1) \approx 0$ and $P(\bar{L}|G = 1) \approx 1$. Then we have

$$P(S = 1|G = 1) \approx P(S = 1|\bar{L}, G = 1).$$

yield

$$\frac{P(S = 1|\bar{L}, G = 1)P(S = 0|\bar{L}, G = 0)}{P(S = 1|\bar{L}, G = 0)P(S = 0|\bar{L}, G = 1)} \approx 1.$$

Acknowledgements

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

References

1. De Benedictis G, Tan Q, Jeune B, et al. Recent advances in human gene-longevity association studies. *Mechanisms of Ageing and Development* 2001; in Print.
2. Ivanova R, Henon N, Lepage V, Charron D, Vicaut E, Schachter F. HLA-DR alleles display sex-dependent effects on survival and discriminate between individual and familial longevity. *Hum Mol Genet.* 1998; 7, 187-194.
3. De Benedictis G, Carotenuto L, Carrieri G, et al. Gene/longevity association studies at four autosomal loci (REN, THO, PARP, SOD2). *Eur J Hum Genet.* 1998; 6:534-541.
4. Heijmans BT, Gussekloo J, Kluit C, et al. Mortality risk in men is associated with a common mutation in the methylene-tetrahydrofolate reductase gene (MTHFR). *Eur J Hum Genet.* 1999;7(2):197-204.
5. Tan Q, De Benedictis G, Yashin AI, et al. Measuring the genetic influence in modulating human life span: Gene-environment and gene-sex interactions. *Biogerontology* 2001;2(3), in print.
6. Tan Q, Yashin AI, Bladbjerg EM, de Maat M, Andersen-Ranberg K, Jeune B, Christensen K, Vaupel JW. Variations of Cardiovascular Disease Associated Genes Exhibit Sex-dependent Influence on Human Longevity. *Experimental Gerontology* 2001;36(8), in print.
7. Yashin AI, De Benedictis G, Vaupel JW, et al. Genes, Demography, and Life Span: The Contribution of Demographic Data in Genetic Studies on Aging and Longevity. *Am J Hum Genet.* 1999; 65:1178-1193.
8. Yashin AI, De Benedictis G, Vaupel JW, et al. Genes and longevity: Lessons from studies on centenarians. *Journal of Gerontology.* 2000; 55:B1-B10.

9. Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Stat Med.* 1994; 13(2):153-62.
10. 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(3):207-13.
11. Yang Q, Khoury MJ, Flanders WD. Sample size requirements in case-only designs to detect gene-environment interaction. *Am J Epidemiol.* 1997; 146(9):713-20.
12. Fisher LD, Van Belle G. *Biostatistics: A methodology for the health sciences.* N.Y.: John-Wiley & Sons; 1993: 229.
13. Bladbjerg EM, Andersen-Ranberg K, de Maat MP, Kristensen SR, Jeune B, Gram J, Jespersen J. Longevity is independent of common variations in genes associated with cardiovascular risk. *Thromb Haemost.* 1999; 82(3), 1100-5.
14. Kannisto V. Development of oldest-old mortality, 1950-1990: evidence from 28 developed countries. Odense: Odense University Press; Odense, 1994:59-66.
15. Vaupel JW, Carey JR, Christensen K, et al. Biodemographic trajectories of longevity. *Science.* 1998; 280:855-860.
16. Schmidt S, Schaid DJ. Potential misinterpretation of the case-only study to assess gene-environment interaction. *Am J Epidemiol.* 1999; 150(8):878-85.
17. Khoury MJ. Genetic epidemiology. In: Rothman KJ, Greenland S, editors. *Modern Epidemiology.* Philadelphia: Lippincott-Raven Publisher; 1998:615-617.

Table 1. Gene-sex interaction based on case-control design

Sex*	Genotype**	Cases	Controls	OR
0	0	a_{00}	b_{00}	$OR_{00} = 1$
0	1	a_{01}	b_{01}	$OR_{01} = a_{01}b_{00} / a_{00}b_{01}$
1	0	a_{10}	b_{10}	$OR_{10} = a_{10}b_{00} / a_{00}b_{10}$
1	1	a_{11}	b_{11}	$OR_{11} = a_{11}b_{00} / a_{00}b_{11}$

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

Table 2. The case-only 2×2 table classified by genotype and sex

Sex*	Genotype**		Sum
	0	1	
0	a_{00}	a_{01}	$n_{1.}$
1	a_{10}	a_{11}	$n_{2.}$
Sum	$n_{.1}$	$n_{.2}$	$n_{..}$

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

Table 3. Case-only approach for assessing gene-sex interactions in the Danish centenarian study

Gene	Male*		Female		Sum	OR	95% CI	<i>p</i>
	0	1	0	1				
Angiotensinogen								
M235	12	34	64	72	182	2.519	1.203-5.275	0.013
FVII Q353	34	12	120	18	184	2.353	1.032-5.362	0.038
FVII-323p10	33	13	117	20	183	2.305	1.038-5.119	0.037

* 1, present; 0, absent.

Table 4. Case-only approach for assessing gene-sex interactions for HLA-DR7 allele in French centenarians

Sex*	Genotype**		Sum
	0	1	
0	328	116	444
1	54	35	89
Sum	382	151	533

The numbers are calculated according to the allele frequency and total number of centenarians from Ivanova et al. Human Molecular Genetics, 1998, Vol. 7, No.2.

* 1, male; 0, female. ** 1, present; 0, absent.
 OR=1.816, 95%CI: 1.129-2.923, p=0.013

Table 5. Case-only approach for assessing gene-environment interaction for simulated data

Area	Genotype*		Sum
	0	1	
North	6378	1617	7995
South	4757	548	5305
Sum	11135	2165	13300

* 1, present; 0, absent.

OR=0.454, 95% CI: 0.406-0.499, p=0