Published in final edited form as:

Annu Rev Gerontol Geriatr. 2013 ; 33(1): 217–232.

Extended Fixed Attribute Dynamics Method and an Illustrative Application in Bio-demographic-genetic Analysis on Longevity

Yi Zeng, Ph.D [Professor]

Center for Study of Aging and Human Development, Medical School of Duke University, Center for Healthy Aging and Development Studies, National School of Development, Peking University.
zengyi68@gmail.com

Abstract

This chapter presents and discuss the extended Fixed Attributes Dynamics (FAD) method to estimate the independent effect of a genetic variant (or other fixed attribute) in the absence of another relevant genetic variant (or another relevant fixed attribute), the joint effects when both are present and the effects of interactions between them. We present an illustrative application of the extended FAD method for estimating the general, independent and joint effects of the FOXO genotypes on longevity, based on the genotypic data of 760 centenarians from the Chinese Longitudinal Healthy Longevity Survey, and the 1,060 middle-age controls. We also discuss the strengths and limitations of the FAD method.

Keywords

Fixed Attributes Dynamics method; Bio-demographic analysis; FOXO genes; G×G and G×E interactions; longevity

1. Introduction

The Fixed Attribute Dynamics (FAD) method belongs to the family of methods of case-control population association analysis, and it is designed to investigate the association of fixed attributes (including genotypes) on longevity. The initial use of the FAD method (known as the Survival Attribute Assay then) by Vaupel (1992) and Yashin et al (1999) required knowledge of cohort age-specific mortality rates. The FAD method extended by Zeng and Vaupel (2004) does not require these rates, which are not available in many practical applications, especially for many developing countries. Zeng and Vaupel (2004) derived the following formula and empirically justified it:

The odds ratio of survival from age x to $x+n$ (RS) for those with the fixed attribute (e.g., a genotype) to those without the attribute can be estimated as:

$$RS = \frac{(1 - p(x))p(x+n)}{p(x)(1 - p(x+n))} \quad (1)$$

where $p(x)$ is the proportion of individuals who are x years old and have the fixed attribute.

Note that Formula (1) provides estimates of the general impact of one fixed attribute (e.g. a genetic variant) on survival from age x to $x+n$, regardless of other relevant genetic variants or other fixed attributes. This chapter aims to present and discuss the extended FAD method and an illustrative application to estimate the independent effect of a genetic variant (or other fixed attribute) in the absence of another relevant genetic variant (or another relevant fixed attribute), the joint effects when both are present and the effects of interactions between them. We present the derivation, justification and discussion of the extended FAD method in the next section. Section three presents an illustrative application of the FAD method for estimating the general, independent and joint effects of the FOXO genotypes on longevity. We conclude the chapter by discussing the strengths and limitations of the FAD method in the last section.

2. The extended FAD method

Zeng et al. (2010) extended the FAD method outlined above to provide a formula to estimate the independent effect of a genetic variant (or other fixed attribute) in the absence of another relevant genetic variant (or another relevant fixed attribute), the joint effects when both are present and the effects of interactions between them.

Let $i=0, 1$ represent not-carrying and carrying the genetic variant (or other fixed attribute), abbreviated as “non-carrier” and “carrier” hereafter; and $j=0, 1$ represent not having and having another fixed attribute (including genetic variant), abbreviated as “not-exposed” and “exposed” hereafter. Let $p_{11}(x)$ be the proportion of x -year-old individuals who are both carrier and exposed among all persons aged x ; similarly define $p_{10}(x)$, $p_{01}(x)$ and $p_{00}(x)$ as the proportions who are a carrier and not-exposed, non-carrier and exposed, and neither carrier nor exposed. Note that the sum of $p_{11}(x)$, $p_{10}(x)$, $p_{01}(x)$ and $p_{00}(x)$ is equal to one.

Let $s_{11}(x)$, $s_{10}(x)$, $s_{01}(x)$ and $s_{00}(x)$ denote the conditional survival probability from age x to $x+n$ for those who are both carrier and exposed, carrier and not-exposed, non-carrier and exposed, and neither carrier nor exposed, respectively; $N(x)$, the total number of persons aged x ; $S(x)$, the overall average conditional survival probability from age x to $x+n$ for all persons aged x .

Because $N(x) p_{ij}(x) s_{ij}(x) = N(x) S(x) p_{ij}(x+n)$, it follows that

$$s_{ij}(x) = S(x) \left(\frac{p_{ij}(x+n)}{p_{ij}(x)} \right). \quad (2)$$

We choose the group of individuals who are neither a carrier nor exposed as a reference group, and it similarly follows that

$$s_{00}(x) = S(x) \left(\frac{p_{00}(x+n)}{p_{00}(x)} \right). \quad (3)$$

Dividing Formula (2) by Formula (3) gives the odds ratio of survival from age x to age $x+n$ of those with the i - j combination of the gene-fixed exposure to those who are neither a carrier nor exposed:

$$RS_{ij/00} = \frac{s_{ij}(x)}{s_{00}(x)} = \frac{p_{00}(x)p_{ij}(x+n)}{p_{ij}(x)p_{00}(x+n)}. \quad (4)$$

If the odds ratio of survival $RS_{ij/00}$ expressed in Formula (4) is significantly greater (or smaller) than 1.0, the i - j combination of the gene-fixed exposure is positively (or negatively) associated with longevity, and the difference between the odds ratio of survival and 1.0 measures the magnitude of the effects on longevity. Standard statistical significance tests including estimates of P values and 95% confidence intervals (Newman, 2001) can be performed to assess whether the estimates of an odds ratio of survival (RS in Formula (1) and $RS_{ij/00}$ in Formula (4)) is significantly different from 1.0, using a simple procedure in standard statistical software (e.g., STATA, SAS or SPSS). Note that $RS_{10/00}$ and $RS_{01/00}$ represent the independent impact of the genetic variant (or another fixed exposure) on survival from age x to $x+n$, in the absence of another fixed exposure (including genetic variant) and that $RS_{11/00}$ represents the joint impact of the genetic variant and the fixed exposure on survival from age x to $x+n$ when both of them present.

In the genetic applications of Formula (4), individuals are identified as “carriers” or “non-carriers,” and a carrier could have 1 or 2 copies of the minor allele of the SNPs of the gene under study. Hence we are implicitly assuming that the minor allele is dominant rather than recessive or additive. A modified version of the FAD approach/application is required to analyze recessive or additive models.

Gene-environment interaction ($G \times E$) is defined as “a different effect of an environmental exposure on disease and health risk in persons with different genotypes,” or, alternatively, “a different effect of a genotype on disease and health risk in persons with different environmental exposures” (Ottman, 1996; IOM, 2006). Similarly, gene-gene interaction ($G \times G$) is defined as “a different effect of a genotype on disease and health risk in persons with and without another genotype.”

Applying the standard definition and method for estimating the $G \times E$ or $G \times G$ interactions (IOM, 2006), we present below procedures for estimating the effects of interactions between the genetic variant and the fixed exposure (including another genetic variant) on survival from age x to $x+n$, using the odds ratios of survival ($RS_{ij/00}$, $i=0, 1$; $j=0, 1$) derived by Formula (4):

On a multiplicative scale, the interaction effect (IS^m) is estimated as:

$$IS^m = RS_{11/00} / (RS_{10/00} \times RS_{01/00}). \quad (5)$$

On an additive scale, the interaction effect (IS^a) is estimated as:

$$IS^a = RS_{11/00} / (RS_{10/00} + RS_{01/00} - 1), \quad (6)$$

If IS^m or IS^a is significantly greater or smaller than one, the interaction between the genetic variant and the fixed exposure (including another genetic variant) positively or negatively affects survival from age x to $x+n$.

Note that assuming multiplicative versus additive interactions is not purely a statistical question but rather one based on what we know about the social environment, behavior, and biology (IOM, 2006). In other words, it is content based, and we should determine whether we adopt a multiplicative or additive assumption case by case based on empirical data or through learning from literature.

Unlike the formulas (1) and (4), we cannot simply use the standard statistical significance tests (Newman, 2001) to assess whether the interaction effects estimated by formulas (5) and (6) are statistically significant because the functional distributions of formulas (5) and (6) are unknown, at least to our knowledge. In theory, we may use multivariate Delta method to do the statistical significance tests. However, in this case the Delta method is not feasible, because we have eight variables (i.e., $P_{ij}(x)$ and $P_{ij}(x+n)$); the sample size may usually not be large enough to estimate the 72 parameters (8 variances plus 64 ($= 8 \times 8$) covariants) in most applications including ours. Based on a careful literature search, we believe that a bootstrap simulation approach is most appropriate for estimating the confidence intervals and P values of IS^m and IS^a (Andersson, et al. 2005; Cameron, et al. 2005).

The basic and extended FAD methods expressed in formulas (1) and (4) can be used to analyze the follow-up data from the same cohort, or cross-sectional data from different cohorts of long-lived cases and younger controls. If cohort follow-up data are used, one needs to assume that the loss to follow-up from age x to $x+n$ is not correlated with respondents' combination of the genetic variant and the fixed exposure under study. In the case of analyzing the cross-sectional data, we actually connect the younger cohort and the long-lived cohort into what the demographic text books call "a hypothetical cohort," and estimate the odds ratio of survival based on proportions of the genetic variant and fixed exposure among the long-lived and younger cohorts (i.e., at advanced and younger ages in the "hypothetical cohorts"). It is important to note that, if one uses cross-sectional data, both the FAD method expressed in Formulas (1) and (4) and the conventional case-control statistical-genetic approach using logistic regression assume that the initial distributions of the genetic variant, the fixed exposure and their combinations, and their association with longevity do not differ substantially between cohorts of long-lived and younger controls, i.e., there is no substantial problem of population stratification (see more detailed discussions in the Appendix on potential bias if this basic assumption is violated). Serious attention must be paid to this issue in applying either the conventional statistical-genetic approach or the FAD method for identifying genetic contributions to longevity. To control for population stratification in the FAD analysis in our illustrative application using the cross-sectional data of centenarians and middle-age controls to be presented in the next section, we have restricted our analysis to a single ethnic group (Han Chinese) residing in Southern China, and classified the samples by gender as well.

The basic and extended FAD methods are based on the fundamental demographic insight that the prevalence of a fixed attribute (e.g., genetic variant, birth weight, childhood conditions, serious disease(s) suffered earlier in life) and combinations of two fixed attributes in a population can change with age even though no individual can change his or her fixed attributes. Therefore, much can be learned about the impact of the fixed attribute (including genetic variant) and the combinations of the fixed attributes on survival.

Note that it is mathematically and numerically verified that, using the same genotypic datasets of the long-lived cases and younger controls, the estimates of the odds ratios of survival and their associated *P* values and 95% confidence intervals and the interaction effects by our FAD method expressed in formulas (1), (4), (5) and (6) are exactly the same as the corresponding numerical results produced by the logistic regression not adjusted for covariates, which is a conventional approach in genetic analysis on the association of a genetic variant with longevity. This fact verifies the statistical properties of our FAD method. Although the FAD method produces the same numerical results as does the logistic regression, the odds ratios of survival based on the FAD method quantify the magnitude of the genetic effects on survival probability. In other words, the FAD method provides clearer demographic interpretations and presents meaningful demographic syntheses of the case-control genetic analysis on longevity. As the two approaches produce exactly the same numerical results, one may adopt the FAD's demographic interpretation of the odds ratio, namely, the effects on survival probability from the average age of the younger controls to the average age of the long-lived cases, even if one still uses the conventional approach of logistic regression. From this point of view, we may consider that the FAD method contributes an additional new insight of demographic interpretation of the magnitude of the genetic effects on survival probability to the conventional approach of logistic regression in case-control genetic analysis on longevity.

3. Illustrative applications of the FAD method for estimating the general, independent and joint effects of the FOXO genotypes on longevity

3.1. Data sources

Our analyses are mainly based on DNA samples from 578 female centenarians and 182 male centenarians who survived to age 100+ when they participated in the interviews in baseline or the follow-up wave(s) of the Chinese Longitudinal Healthy Longevity Survey (CLHLS) study (Zeng et al, 2001; Zeng et al., 2008). The DNA samples of the middle-age controls (378 women and 682 men) were collected from the routine clinic health examinations involving ordinary people. These 760 centenarians and the 1,060 middle-age controls were all from Southern China¹, and the internationally standardized techniques and procedures used to produce these genotypic data have already been described (Li et al., 2009). More

¹China is traditionally and culturally considered to have Southern and Northern parts, although there is no legitimate or official administrative meaning to this division. The South and North are divided by the geographical line of the Qin Ling mountain and Huai River. The line crosses through four provinces of Shaanxi, Henan, Anhui and Jiansu. Among the 22 provinces where CLHLS was conducted, Liaoning, Jilin, Heilongjiang, Hebei, Beijing, Tianjing, Shanxi, and Shangdong belong to Northern China; Shanghai, Zhejiang, Fujian, Jiangxi, Hubei, Hunan, Guangdong, Guangxi, Sichuan, Chongqing belong to Southern China; and the northern and southern sides of the Qin Ling mountain-Huai River line in Shaanxi, Jiangsu, Anhui, Henan belong to Northern and Southern China, respectively.

specifically, we use the same genotypic data for the SNPs rs2755209 and rs2755213 of *FOXO1A* gene and the three SNPs rs2253310, rs2802292, rs4946936 of *FOXO3A* gene as those described in Li et al. (2009), to investigate different topics following different methodological approaches. The rs17630266 SNP of *FOXO1A* which was found not to be associated with longevity will not be investigated further here. The minor allele frequency (MAF) distributions of the SNPs, the single SNP association analysis, genotype association analysis with recessive and additive models, the linkage disequilibrium and haplotype association analysis have already been reported (Li et al., 2009), and thus there is no need to be repeated here. To ease the presentation, we abbreviate in this article the two SNPs rs2755209 and rs2755213 of *FOXO1A* gene as 1A-209 and 1A-213; we abbreviate the three SNPs rs2253310, rs2802292, rs4946936 of *FOXO3A* gene as 3A-310, 3A-292, and 3A-936, respectively.

To define the genotypes to be analyzed in this chapter, we adopt the dominant model in which individuals who carry 1 or 2 copies of the minor allele are coded as “Yes” (carrier), and the individuals who do not carry the minor allele are coded as “No” (non-carrier). More specifically, carrying or not-carrying the minor allele of one of the two SNPs of *FOXO1A* or of one of the three SNPs of *FOXO3A* is determined as “Yes (carrier)” or “No (non-carrier)” for each of the centenarians and controls.

3.2. General effects of the genotypes of *FOXO1A* and *FOXO3A* on long-term survival probability from middle age to age 100+

Table 1 presents the estimates of frequency distribution of the genotypes and the general effects of the genotypes of *FOXO1A* and *FOXO3A* on long-term survival probability from middle age to age 100+, regardless of carrying or not-carrying another relevant genotype. The results show that, compared to non-carriers, the odds ratio of survival from middle ages to age 100+ for women who are carriers of the minor alleles of 1A-209 and 1A-213 are 0.70 ($P=0.0073$) and 0.71 ($P=0.0152$), respectively; the corresponding odds ratio for men are 0.73 ($P=0.093$) and 0.65 ($P=0.025$). These estimates imply that the probability of survival from middle-age to age 100+ for women who are a carrier of the minor alleles of the 1A-209 and 1A-213 are 30 and 29 percent lower than those who are non-carriers, and the estimates are statistically significant; those men who carry the minor alleles of 1A-209 and 1A-213 may have 27 and 35 percent lower probability of survival from middle-age to age 100+, although the estimate for 1A-209 is not statistically significant (see Table 1). Compared to non-carriers, the odds ratio of survival for women and men who carry the minor alleles of 3A-310, 3A-292, and 3A-936 are 1.62–1.67 ($P=0.0002\sim0.0004$) and 1.61–1.73 ($P=0.001\sim0.005$), respectively. These estimates imply that, on average, middle-age women and men who carry the minor allele of any one of the three SNPs of *FOXO3A* may have 62–67 and 61–73 percent higher probability of survival to age 100+, with all estimates highly statistically significant.

3.3. The independent and joint effects of the genotypes of *FOXO1A* and *FOXO3A* on long-term survival from middle age to age 100+

The analyses concerning *FOXO1A* presented in Table 1 employing Formula (1) of the FAD method are based on comparing the frequencies of carrying the minor allele of the SNPs of

FOXO1A (regardless of whether carrying the minor allele of the SNPs of *FOXO3A*) between centenarians and younger controls. So are the analyses concerning *FOXO3A* presented in Table 1. Although they are useful in exploring the general impacts of one genotype on longevity regardless of carrying or not-carrying the other relevant genotype, these analyses cannot reveal the independent, joint and interactive effects of the genotypes of *FOXO3A* and *FOXO1A*, which influence survival in opposite directions.

Applying our extended FAD method expressed in Formula (4), Tables 2 and 3 present the estimates of odds ratios of survival ($RS_{ij/00}$) from middle-age to age 100+ reflecting the independent and joint effects of the genotypes of *FOXO3A* and *FOXO1A*. The genotypes presented in Tables 3 and 4 are defined by a combination of carrying or not-carrying minor alleles of one of the two SNPs of *FOXO1A* and minor alleles of one of the three SNPs of *FOXO3A*, with the reference group who carry neither the minor allele of the SNP of *FOXO1A* nor the minor allele of the SNP of *FOXO3A*. We choose these combinations, because we aim to capture the independent effects ($RS_{10/00}$ and $RS_{01/00}$), joint effect ($RS_{11/00}$) of the two genes of *FOXO1A* and *FOXO3A* which affect longevity in opposite directions, and the gene by gene interactions. We did not choose the combination of the pairs within the three SNPs of *FOXO3A* gene and the combinations of the pairs within the two SNPs of the *FOXO1A* gene, mainly because such haplotype-based analysis have already been reported (Li et al., 2009).

Compared to those who are non-carriers of the minor allele of one of the three SNPs of *FOXO3A* and non-carriers of the minor allele of *1A-209*, the negative and independent effects of carrying the minor allele of *1A-209* on survival in women are all highly significant ($P=0.002\sim0.006$) (see Table 2), consistent with the general effects presented in 1st row of the panel for women in Table 1. The corresponding negative and independent effects of *1A-213* in women are moderately significant ($P=0.029\sim0.043$) or not significant ($P=0.061$) (see Table 3), which are consistent in terms of direction and magnitude of the odds ratios but with lower level of significance compared with the general effects (see 2nd row of the panel for women in Table 1). Other estimates of the independent effects of the three SNPs of *FOXO3A* in women are not significant which is remarkably different from the general effects regardless of the carrying or not-carrying the minor allele of the three SNPs of *FOXO3A* estimated by the basic FAD analysis (see Table 1) and the conventional genetic analysis (e.g., Willcox et al., 2008; Li et al., 2009).

In men the patterns of the independent effects of *FOXO1A* and *FOXO3A* are substantially different from those of women. Compared to men who are non-carriers of the minor allele of one of the three SNPs of *FOXO3A* and non-carriers of the minor allele of the corresponding SNP of *FOXO1A*, the independent effects of the two SNPs of *FOXO1A* in men are all non-significant but the independent effects of the three SNPs of *FOXO3A* are all highly significant ($P=0.002\sim0.005$; see panels for Men in Tables 2 and 3), which, unlike the results for women, are consistent with the general effects.

Dual presence of the minor allele of one of the three SNPs of *FOXO3A* and the minor allele of one of the two SNPs of *FOXO1A* may increase the chance of survival from middle age to age 100+ by 12–19 percent in women and by 14–30 percent in men, compared to those who

neither carry the minor allele of the corresponding SNP of *FOXO3A* nor carry the minor allele of the corresponding SNP of *FOXO1A*, but all of the estimates associated with the dual presence (i.e., joint effects) of *FOXO3A* and *FOXO1A* are not statistically significant (see the italic and bold numbers in Tables 3 and 4). These results indicate that the positive effects of *FOXO3A* and the negative effects of *FOXO1A* substantially compensate each other, while the impact of *FOXO3A* is somewhat stronger.

3.4. The effects of interaction between *FOXO1A* and *FOXO3A* on survival from middle age to age 100+

The absolute values of the correlation coefficients between each of the two SNPs of the *FOXO1A* and each of the three SNPs of the *FOXO3A* are very small (about half <0.01 and another half $0.01-0.05$). Moreover, *FOXO1A* and *FOXO3A* are located on different chromosomes (*FOXO1A* on chromosome 13, *FOXO3A* on chromosome 6), and they are physically independent in terms of linkage disequilibrium. Consequently, the assumption of independence between *FOXO1A* and *FOXO3A* is clearly met. Thus, we employed Formula (5) of the FAD approach described in Section 2, which is based on the multiplicative assumption to estimate the effects of interactions between *FOXO1A* and *FOXO3A* on long-term survival; *P* and 95% CI are estimated by the bootstrap simulation method. The results show that all interactions between each of the two SNPs of the *FOXO1A* and each of the three SNPs of the *FOXO3A* are not statistically significant (data not shown).

4. Strengths and limitations of the FAD method

The standard method of survival analysis using cohort follow-up data is limited in investigating the effects of the fixed attributes (including genetic variants) on long-term survival, because it requires prospective tracking of the mortality of study participants, with a follow-up period for estimating long-term survival probabilities that may be prohibitively expensive and time-consuming (Hill, 1999). The FAD method circumvents this major limitation of standard survival analysis because it does not require expensive follow-up mortality data of study participants and it can be applied to both cohort and cross-sectional datasets if the basic assumption of no serious problems of population stratification is met. However, the FAD method also has a major limitation: it permits an analysis to ascertain the 'de facto' association of a fixed attribute (including genes) or the combination of gene-fixed exposure on longevity, while usually controlling for age and sex. In theory, we can further control for other covariates by extending the FAD method through cross-tabulations of the study subjects by confounding factors which are also fixed attributes and meet the assumption of no substantial problems of population stratification, but this may cause problems with small sample sizes of the sub-groups.

Thus, we may apply both FAD methods to investigate the long-term survival (e.g. from middle-age to ages 100+) and multivariate survival analysis to study the factors affecting survival in shorter periods while controlling for various confounders whenever possible to complement each other and to maximize the utility of the available data sources (see Zeng et al., 2010).

Acknowledgments

This study is jointly funded by National Natural Science Foundation of China (71110107025 and 71233001), the 211 program of Peking University and the U.S. National Institute of Aging (AG023627). I am very grateful for the research collaborations of Xiaoli Tian and his lab, Lingguo Cheng, Huashuai Chen, Elizabeth Hauser and James W. Vaupel. I appreciate the thoughtful comments provided by the editors.

Appendix: Discussions of the basic assumption inherent in the FAD method and the conventional case-control statistical-genetic approach

Note that the derivation of the formula (4) in Section 2 is based on the framework of cohort survivorship. If the formula (4) is applied to follow-up data from the same cohort counted at different time points (i.e. at ages x and $x+n$), one needs to assume that the loss-to-follow-up from age x to $x+n$ is not correlated with respondents' combination of the genetic variant and the fixed exposure under study. However, the cohort long-term follow-up data, such as the $p_{ij}(x)$ at middle-age and age 100+ from the same cohort, are usually not available.

Consequently, one may use cross-sectional data from different cohorts of long-lived cases (e.g., centenarians) and younger controls (e.g., middle-age group). In this case, we connect the younger cohort and the long-lived cohort into "a hypothetical cohort," and we assume that the initial distributions of the combinations of the genetic variant and the fixed exposure at age x of the cohorts of long-lived individuals (denoted as $p^1_{ij}(x)$) and younger controls (denoted as $p^2_{ij}(x)$) do not differ substantially. We then compare the observed $p^1_{ij}(x+n)$ of the long-lived individuals (cases) and the observed $p^2_{ij}(x)$ of the middle-age controls, considering the cases and controls as members of a hypothetical cohort, and derive the estimate of $RS_{ij/00}$. This basic assumption inherent in the FAD method is the same as the one adopted in the conventional case-control statistical-genetic approach using cross-sectional data from long-lived individuals (as cases) and younger controls, namely, assuming that there is no substantial problem of population stratification. We further explore in the following discussion the implications of this basic assumption adopted in both FAD and the conventional case-control statistical-genetic approaches, and its potential bias if the basic assumption is violated.

Let $p^1_{ij}(0)$ and $p^2_{ij}(0)$ denote the initial distribution of the combination of the genetic variant and the fixed exposure at birth among the individuals born $x+n$ years ago and x years ago, respectively; $N^1_{ij}(0)$ and $N^2_{ij}(0)$, the initial size (number of births) of the cohorts of the individuals born $x+n$ years ago and x years ago, respectively; $S^1(0 \rightarrow x)$ and $S^2(0 \rightarrow x)$, the overall average probability of survival from age 0 to age x among all of the individuals born $x+n$ years ago and x years ago, respectively; C^1_{ij} and C^2_{ij} , the ratio of the probability of survival from age 0 to x among persons with the i - j combination of the genetic variant and the fixed exposure to the overall average probability of survival from age 0 to x for all of the cohort members born $x+n$ years ago and x years ago, respectively; $p^1_{ij}(x)$ and $p^2_{ij}(x)$, the proportions of persons aged x with the combination of the genetic variant and the fixed exposure among all individuals aged x in the cohorts born $x+n$ years ago and x years ago, respectively.

$$p^1_{ij}(x) = \frac{N^1(0)p^1_{ij}(0)[S^1(0 \rightarrow x)C^1_{ij}]}{N^1(0)S^1(0 \rightarrow x)} = p^1_{ij}(0)C^1_{ij} \quad (\text{A-1})$$

Similarly,

$$p^2_{ij}(x) = \frac{N^2(0)p^2_{ij}(0)[S^2(0 \rightarrow x)C^2_{ij}]}{N^2(0)S^2(0 \rightarrow x)} = p^2_{ij}(0)C^2_{ij} \quad (\text{A-2})$$

As discussed earlier, when the extended FAD method is applied to the cross-sectional data from the long-lived individuals aged $x+n$ and the younger controls aged x , one in fact assumes that: $p^1_{ij}(x) = p^2_{ij}(x)$ or $p^1_{ij}(0)C^1_{ij} = p^2_{ij}(0)C^2_{ij}$, namely, assuming:

1. the initial distributions of the combinations of the genetic variant and the fixed exposure of the two cohorts born $x+n$ years ago and x years ago are the same, i.e. $p^1_{ij}(0) = p^2_{ij}(0)$;
2. the associations between the genetic variant and longevity in the two cohorts are the same, $C^1_{ij} = C^2_{ij}$.

If these assumptions are met, the odds ratio of survival $RS_{ij/00}$ produced by the formula (4) based on the cross-sectional data adequately provides the estimate of the effects of the i - j combination of the genetic variant and the fixed exposure on longevity. Based on Eqs. (A-1) and (A-2), we observe that, theoretically speaking, the estimates of the odds ratio of survival are not directly affected by the differences in the initial sizes (numbers of births) of the cohorts (i.e. $N^1(0)$ and $N^2(0)$) and the overall average probabilities of survival from birth to age x among all of the individuals born $x+n$ years ago and x years ago ($S^1(0 \rightarrow x)$ and $S^2(0 \rightarrow x)$), as they are canceled out during the process of derivation. However, there would be bias in the estimates of the odds ratio if the basic assumptions (1) and (2) described above are violated due to population stratification. The magnitude of the bias depends on size direction of the difference between $p^1_{ij}(0)$ and $p^2_{ij}(0)$ and between C^1_{ij} and C^2_{ij} .

In our study, we control for population stratification in the FAD analysis using the cross-sectional data of centenarians and middle-age controls by restricting our analysis to a single ethnic group (Han Chinese) residing in Southern China, and classifying the samples by gender as well. We believe that this may reduce and hopefully minimize the bias, but we are aware of that there may still be potential bias, which needs to be evaluated in the future when more sophisticated cohort data sources become available.

References

- Andersson T, Alfredsson L, Källberg H. Calculating Measures of Biological Interaction. *European Journal of Epidemiology*. 2005; 20(7):575–579. [PubMed: 16119429]
- Cameron, AC.; Trivedi, PK. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press; 2005.
- Hill ME. Multivariate survivorship analysis using two cross-sectional samples. *Demography*. 1999; 36:497–503. [PubMed: 10604077]

- Hernandez, LM.; Blazer, DG., editors. IOM (Institute of Medicine, National Academy of Sciences) Committee. Genes, behavior, and the social environment: Moving beyond the nature/nurture debate. Washington, D.C.: National Academia of Sciences; 2006.
- Li Y, Wang W, Cao H, Lu J, Wu C, Hu F, Guo J, Zhao L, Yang F, Zhang Y, Li W, Zheng G, Cui H, Chen X, Zhu Z, He H, Dong B, Mo X, Zeng Yi, Tian Xiao-Li. Genetic association of FOXO1A and FOXO3A with longevity trait in Han Chinese populations. *Human Molecular Genetics*. 2009 in Press (Xiao-Li Tian and Yi Zeng are co-corresponding authors).
- Newman, Stephen C. Biostatistical methods in epidemiology. John Wiley and Sons Inc; 2001. Wiley series in probability and statistics
- Ottman R. Gene-environment interaction: definition and study designs. *Preventive Medicine*. 1996; 25:764–770. [PubMed: 8936580]
- Perls TT, Wilmoth J, Levenson R, Drinkwater M, Cohen M, Bogan H, Joyce E, Brewster S, Kunkel L, Puca A. Life-long sustained mortality advantage of siblings of centenarians. *Proc Natl Acad Sci USA*. 2002; 99:8442–8447. [PubMed: 12060785]
- Vaupel, JW. Analysis of population changes and differences; Presented at the annual meeting of the Population Association of America; April 1992; 1992.
- Willcox BJ, Donlon TA, He Q, Chen R, Grove JS, Yano K, Masaki KH, Willcox DC, Rodriguez B, Curb JD. FOXO3A genotype is strongly associated with human longevity. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105:13987–13992. [PubMed: 18765803]
- Yashin AI, Benedictis GD, Vaupel JW, Tan Q, Andreev KF, Iachine IA, Bonafe M, Luca MD, Valensin S, Carotenuto L, Franceschi C. Gene, demography and life Span: The contribution of demographic data in genetic studies of aging and longevity. *American Journal of Human Genetics*. 1999; 65:1178–1193. [PubMed: 10486337]
- Zeng Y, Vaupel JW, Xiao Z, Zhang C, Liu Y. The healthy longevity survey and the active life expectancy of the oldest old in China. *Population: An English Selection*. 2001; 13(1):95–116.
- Zeng, Yi; Vaupel, JW. Association of Late Childbearing With Healthy Longevity among The Oldest-Old in China. *Population Studies*. 2004; 58(1):37–53. [PubMed: 15204261]
- Zeng, Yi; Poston, Dudley; Vlosky, Denese Ashbaugh; Gu, Danan, editors. *Healthy Longevity in China: Demographic, Socioeconomic, and Psychological Dimensions*. Dordrecht: Springer Publisher; 2008.
- Zeng, Yi; Gu, Danan. “Reliability of Age Reporting among the Chinese Oldest-old in the CLHLS Datasets”, Chapter 4. In: Zeng, Yi; Poston, Dudley; Vlosky, Denese Ashbaugh; Gu, Danan, editors. 2008. *Healthy Longevity in China: Demographic, Socioeconomic, and Psychological Dimensions*. Dordrecht: Springer Publisher; 2008.
- Zeng, Yi; Cheng, L.; Chen, H.; Cao, H.; Hauser, E.; Liu, Y.; Xiao, Z.; Tan, Q.; Tian, X.; Vaupel, JW. Effects of FOXO Genotypes on Longevity: A Bio-demographic Analysis. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*. 2010; 65A(12):1285–1299. (Yi Zeng and Xiao-Li Tian are co-corresponding authors) [PMC PMC2990269].

Table 1

Estimates of the *general effects* of the genotypes, using the FAD method (formula (1))

Genotypes	Women				Men			
	Cent. %car.	Contr-%car.	RS (95% CI)	P	Cent. %car.	Contr-%car.	RS (95% CI)	P
Carrying minor allele of 1A-209	42.3	51.3	0.70 (0.53–0.91)	0.0073	41.8	49.5	0.73 (0.51–1.05)	0.093
Carrying minor allele of 1A-213	57.8	65.8	0.71 (0.54–0.91)	0.0152	58.8	68.6	0.65 (0.45–0.95)	0.025
Carrying minor allele of 3A-310	51.9	39.3	1.67 (1.27–2.18)	0.0002	55.5	43.0	1.65 (1.19–2.30)	0.003
Carrying minor allele of 3A-292	51.7	39.8	1.62 (1.24–2.12)	0.0004	56.3	42.7	1.73 (1.24–2.41)	0.001
Carrying minor allele of 3A-936	49.5	37.1	1.66 (1.26–2.18)	0.0003	51.9	40.2	1.61 (1.16–2.24)	0.005

Notes: (1) Cent. %car. – % of centenarians who carry the minor allele; Contr %car. – % of middle-age controls who carry the minor allele. (2) Sample size – centenarians: 578 women and 182 men; middle-age controls: 378 women and 682 men, all Han Chinese from Southern China.

Table 2

Estimates of the *independent and joint effects* of the genotypes (combinations of SNP 1A-209 of *FOXO1A* with the three SNPs of *FOXO3A*), using the extended FAD method (formula (4))

Carrier or non-carrier of the minor allele of the SNPs of <i>FOXO3A</i>	Non-carrier of minor allele of 1A-209 (<i>i</i> = 0)			Carrier of minor allele of 1A-209 (<i>i</i> = 1)		
	Cent. %genotype	Contr. %genotype	<i>RS</i> _{ijm0} (95% CI)	<i>P</i>	Cent. %genotype	Contr. %genotype
Women						
3A-310, non-carrier (<i>j</i> =0)	27.8	26.6	1.00		20.2	34.0
3A-310, carrier (<i>j</i> =1)	29.9	21.7	1.32(0.89–1.94)	0.148	22.1	17.7
3A-292, non-carrier (<i>j</i> =0)	27.7	26.6	1.00		20.5	33.6
3A-292, carrier (<i>j</i> =1)	30.1	21.9	1.31(0.89–1.93)	0.148	21.7	17.9
3A-936, non-carrier (<i>j</i> =0)	28.8	28.2	1.00		21.6	34.6
3A-936, carrier (<i>j</i> =1)	28.9	19.9	1.41(0.95–2.10)	0.073	20.7	17.3
Men						
3A-310, non-carrier (<i>j</i> =0)	23.3	29.3	1.00		20.6	27.8
3A-310, carrier (<i>j</i> =1)	35.0	22.3	1.96(1.23–3.15)	0.003	21.1	20.6
3A-292, non-carrier (<i>j</i> =0)	23.8	29.6	1.00		19.9	27.8
3A-292, carrier (<i>j</i> =1)	34.8	22.0	1.96(1.23–3.15)	0.003	21.5	20.6
3A-936, non-carrier (<i>j</i> =0)	26.6	31.6	1.00		21.5	28.3
3A-936, carrier (<i>j</i> =1)	32.0	19.9	1.92(1.20–3.06)	0.004	19.9	20.2

Notes: (1) The sample size is the same as that listed in Table 1.

(2) The odds ratio 1.00 refers to the reference group with a genotype without carrying the minor allele of any of the two SNPs paired for the analysis; the italic and bold numbers in this table present the estimates of the *joint effects* of carrying the minor allele of SNP 1A-209 and carrying the minor allele of one of the three SNPs of *FOXO3A*; and the rest present the estimates of the *independent effects* of carrying the minor allele of one of the three SNPs of *FOXO3A* but not-carrying the minor allele of the SNP 1A-209, or carrying the minor allele of the SNP 1A-209 but not-carrying the minor allele of one of the three SNPs of *FOXO3A*.

Table 3

Estimates of the *independent and joint effects* of the genotypes (combinations of SNP 1A-213 of *FOXO1A* with the three SNPs of *FOXO3A*), using the extended FAD method (formula (4))

Carrier or non-carrier of the minor allele of the SNPs of <i>FOXO3A</i>	Non-carrier of minor allele of 1A-213 (<i>i</i> = 0)			Carrier of minor allele of 1A-213 (<i>i</i> = 1)		
	Cent. % genotype	Contr % genotype	RS_{ij00} (95% CI)	<i>P</i>	Cent. % genotype	Contr % genotype
Women						
3A-310, non-carrier (<i>j</i> =0)	19.2	18.8	1.00		28.7	41.7
3A-310, carrier (<i>j</i> =1)	23.0	14.6	1.55(0.97–2.48)	0.052	29.1	24.9
3A-292, non-carrier (<i>j</i> =0)	19.0	18.8	1.00		29.1	41.3
3A-292, carrier (<i>j</i> =1)	23.2	14.8	1.55(0.97–2.47)	0.052	28.7	25.1
3A-936, non-carrier (<i>j</i> =0)	20.1	19	1.00		30.3	43.7
3A-936, carrier (<i>j</i> =1)	22.1	13.8	1.53(0.95–2.47)	0.066	27.5	23.5
Men						
3A-310, non-carrier (<i>j</i> =0)	13.9	17.5	1.00		30.0	39.5
3A-310, carrier (<i>j</i> =1)	27.2	14.7	2.34(1.30–4.25)	0.002	28.9	28.3
3A-292, non-carrier (<i>j</i> =0)	14.3	17.7	1.00		29.3	39.7
3A-292, carrier (<i>j</i> =1)	27.1	14.5	2.29(1.28–4.14)	0.003	29.3	28.1
3A-936, non-carrier (<i>j</i> =0)	17.2	19.3	1.00		30.9	40.6
3A-936, carrier (<i>j</i> =1)	24.3	12.8	2.14(1.21–3.81)	0.005	27.6	27.3

Notes: (1) the same as the note (1) in Table 1.

(2) the same as the note (2) in Table 2, except replacing 1A-209 by 1A-213.