RESEARCH COMMUNICATION

Applicability of Allele/genotype Frequency from Documented Controls for Case-Control Studies on Genotypes among Japanese: MTHFR C677T as an Example

Kenta Iida¹, Koutaro Tomita², Rieko Okada², Sayo Kawai², Emi Morita², Asahi Hishida², Mariko Naito², Kenji Wakai², Nobuyuki Hamajima²*

Abstract

In a case-control study, controls have to be selected from the population where cases are identified. However, there is an idea to make common controls applicable for different case-control studies, which could reduce study costs. This study compared methylenetetrahydrofolate reductase C677T among subjects from different studies, to examine the applicability of combined subjects as controls. Case-control studies and cross-sectional studies with more than 100 Japanese controls or subjects were selected from PubMed in December 2008. Between 1996 and 2008, 31 eligible studies with 14,510 subjects in total were published; the 677T allele frequency varied from 0.300 to 0.438. The genotype frequencies were all in Hardy-Weinberg equilibrium. The average weighted with the number of subjects was 0.385. The 95% confidence interval (95% CI) of 10 studies did not include the weighted average. The study whose proximal limit of 95% CI was furthest from the weighted average was removed, and then the weighted average was recalculated. Through the process, 7 studies were excluded, resulting in the remaining 24 studies having a 95% CI including the weighted average (0.391) with 10,854 subjects. Of the 7 excuded studies, one was from patients enrolled in a clinical study (0.431), two were relatively isolated (0.300 and 0.438), two were published ten years ago (0.315 and 0.334), one was from a study whose allele frequency of female subjects was 0.382 (0.436), and one had a large sample size (0.360). This example demonstrated that the allele frequency of MTHFR 677T was in common among 24 out of 31 Japanese case-control studies, suggesting that the calculated allele frequency could be used for Japanese case-control studies, whose cases are sampled in a general population.

Key Words: Case-control study - polymorphism - allele frequency - methylenetetrahydrofolate reductase - Japanese

Asian Pacific J Cancer Prev, 10, 231-236

Introduction

In a case-control study, controls have to be selected from the population where cases are identified. If controls are selected in a different population, the prevalence of the factor under study may differ between cases and controls, which causes selection bias in the association between the factor and disease risk under study. For example, in a case-control study on lung cancer and smoking habits, cases derived from a population with a high smoking rate and controls derived from a population with a low smoking rate will produce a high odds ratio (OR), even if the smoking is not associated with lung cancer risk in each population.

Since control selection and information collection from the selected controls need costs, there was an idea to make a commonly applicable control group for different case-control studies. In Japan, case-control studies on intractable diseases were challenged to make common controls for several different diseases, although epidemiologists realized the problems and limitations concerning the controls because of potential confounding. Recent epidemiologic studies on genetic polymorphisms propose new situations concerning controls. In general, allele/genotype frequencies are independent of sex, age, and exposure of hazardous agents, but the difference in the frequency may occur among different ethnic groups.

In multi-ethnic populations, population stratification may cause bias in the estimation of genotype OR through the confounding of ethnicity.

In a population consisting almost completely of a single ethnic group, however, the effects of population stratification may be negligible (Yamaguchi-Kabata et al., 2008). If so, allele/genotype frequencies of documented controls may be commonly used for different case-control studies on the associations with disease risks. This study
compared an allele frequency among subjects from different studies, to examine the applicability of combined subjects as a control group. In addition, the random variation of the allele frequency according to the sample size was demonstrated to realize how large the random variation is in comparison with a possible shift of the documented allele frequency from general populations. Methylenetetrahydrofolate reductase (MTHFR) C677T polymorphism was used as an example.

MTHFR catalyses the irreversible conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate, which serves as a methyl donor in the reaction converting homocysteine to methionine. The MTHFR C677T polymorphism is a C to T transition at base pair 677 (exon 4), leading to an amino acid substitution (alanine to valine) at codon 222. The TT homozygotes presented a significantly elevated homocysteine concentration, as compared with the CC homozygotes. Accordingly, the polymorphism has been considered a possible genetic risk factor for folate-related diseases including cardiovascular diseases (Trabetti, 2008).

Materials and Methods

Studies on MTHFR C677T polymorphism for Japanese were listed from PubMed in December 2008. Case-control studies with more than 100 controls and cross-sectional studies with more than 100 subjects were selected. The reports based on the same dataset were excluded. The allele frequency was calculated from the genotype frequency. Hardy-Weinberg equilibrium was examined for all studies, using a $\chi^2$ test. The 95% confidence interval (CI) of the allele frequency was calculated based on a binomial distribution. The random variation of the point estimate for a given allele frequency in the population was calculated according to sample size. These calculations were conducted by computer program STATA Version 7.0 (STATA Corporation, College Station, TX).

Results

As shown in Table 1, 31 eligible studies were found to be published between 1996 and 2008. The total numbers of subjects were 14,510. Six studies were for males and four studies were for females. The residence of subjects distributed from Hokkaido to Kyushu, although 6 studies were conducted in Aichi, 5 studies in Tokyo, and 5 studies in Hyogo. Three studies covered throughout Japan. The minimum of age was 18 years and the maximum was 91 years. A study by Nishio et al in 1996 was smallest (129 subjects), and a study by Yamada et al in 2006 was the largest (2,291 subjects).

The genotype frequencies were in Hardy-Weinberg equilibrium for all studies. The allele frequency of the T allele varied from 0.300 to 0.438 among 31 studies. The

![Table 1. Studies on Frequencies of Methylenetetrahydrofolate Reductase Genotype among Japanese](image-url)
Allele/genotype Frequency from Documented Controls - MTHFR C677T as an Example

The average weighted with the number of subjects was 0.385 (95%CI, 0.379-0.390). The 95%CI of 10 studies did not include the weighted average; 6 studies (Miyano et al., 2000; Kohara et al., 2003; Hiraoka et al., 2004; Lwin et al., 2006; Yamada et al., 2006; Arai et al., 2007) had a higher frequency and 4 studies (Maruyama et al., 2004; Moriyama et al., 2002; Morita et al., 1998; Morita et al., 1997) had a lower frequency.

The study whose proximal limit of 95%CI was furthest from the weighted average was removed, then the weighted average was recalculated. Through this process, 7 studies were excluded, resulting that remaining 24 studies had a 95%CI including the weighted average (0.391) calculated from 24 studies with 10,854 subjects (Figure 1). The 95%CI for 0.391 was 0.384-0.397.

The characteristics of the subjects for the 7 studies were as follows. The study subjects by Arai et al were participants of Serum Lipid Survey 2000 from 9 hospitals (Arai et al., 2007). They were randomly selected from 2,267 patients with 202.1 mg/dl of cholesterol on average, who were not treated with lipid-altering medication. The proportion of the patients with hyperlipidemia or other diseases was not described.

Lwin et al conducted a cross-sectional study for residents aged 40-69 years at Shiso, Hyogo Prefecture in 1999 and 2000, which was published for males in 2002 (Lwin et al., 2006) and for females in 2002 (Lwin et al., 2002). The T allele was 0.382 (95%CI, 0.339-0.427) among 242 females, but 0.436 (95%CI, 0.398-0.474) among 335 males.

Maruyama et al enrolled health checkup examinees (270 males and 340 females) aged 30-69 years in a relatively isolated rural area, Kisei-cho, Mie prefecture from 1997 to 2002 (Maruyama et al., 2004). Excluding those under medication such as anti-hyperlipemic, anti-diabetes mellitus, anti-inflammatory, antiepileptic, antacids, thiazide diuretic and anti-rheumatoid arthritis drugs, 147 males and 293 females were analyzed for the polymorphism, giving that the allele frequency was 0.300. Ishikawa et al also reported the allele frequency to be 0.337 among 132 male non-drinkers in an unspecified area in Mie prefecture (Ishikawa et al., 2006).

Moriyama et al conducted a cross-sectional study in 1996, whose subjects were health checkup examinees in Kochi City, Kochi prefecture of Shikoku island. The subjects consisted of 329 males and 653 females aged 20 to 74 years. The allele frequency was 0.360, but the relatively large sample size produced a 95%CI of 0.339-0.382, which did not include the weighted average. The difference from the final weighted average 0.391 was not large (0.031).

In the study by Miyano et al, the subjects were 307 postmenopausal healthy women aged 46 to 91 years living in Nagano prefecture. The source of the enrollment framework was not described (Miyano et al., 2000).

In the study by Morita et al in 1998, the control subjects (174 males and 151 females) aged 46 to 89 years were enrolled from health checkup examinees in Tokyo, having a lower frequency, 0.315 (Morita et al., 1998). In another Morita’s study in 1997, the study subjects were 778 male health checkup examinees aged 26 to 86 years in Tokyo area (Morita et al., 1997). The allele frequency was

Figure 1. Allele Frequency of Methylenetetrahydrofolate Reductase 677T among 31 Studies for Japanese
similarity lower (0.334) than the average.

Table 2 shows the 50%CI and 80%CI of the sample allele frequency for given population allele frequencies (0.1, 0.2, 0.3, 0.4, and 0.5) according to sample size (2N, where N means the number of subjects). When the sample size is only 50 (25 persons), the point estimate distributes within the difference of 0.032 (lower limit in case of population frequency = 0.1) to 0.057 (upper limit in case of population frequency = 0.5) around the population frequency under the possibility of 50%. Under the possibility of 80%, it distributes 0.051 to 0.099 around the population frequency. In case of a sample size of 100 (50 persons), it distributes with differences of 0.022 to 0.039 and 0.037 to 0.069, respectively. They are thus very large differences compared with the possible variation among general populations in areas or regions of Japan.

Discussion

We identified 31 eligible studies reporting the MTHFR C677T genotype frequency among putative general populations in Japan. The 677T allele frequency varied from 0.300 to 0.438. The average weighted with the number of subjects was 0.385. The 95%CI of 10 studies did not include the weighted average. When extreme 7 studies were excluded, remaining 24 studies had a 95%CI including the weighted average (0.391) calculated from the 24 studies with 10,854 subjects. This example demonstrated that the allele frequency of MTHFR 677T could be common among 24 out of 31 Japanese studies. As demonstrated in Table 2, the random variation of the allele frequency in the samples seems larger than the possible shift from general Japanese populations in case of a small-size control group.

The method to remove the extreme studies may affect the final dataset to calculate the weighted average. We calculated it in a different method; the weighted average was calculated based on the studies excluding the study under comparison. This alternative method provided the same 24 studies as the final dataset. Although the final dataset was the same in this study, the same dataset may not always be guaranteed in other situations, especially when the study number is small.

Allele frequency varies substantially in many polymorphisms among different ethnic groups. MTHFR allele was reported to be 0.41 in Japanese Americans, 0.36 in Whites, 0.42 in Latinos, 0.13 in African American, and 0.22 in Native Hawaiians (Le Marchand et al., 2005). However, no studies reported a marked difference in the allele frequency measured with the same method by the same researchers among different areas in Japan. In the present study on MTHFR C677T, 7 papers reported a significantly different allele frequency from the weighted average (0.391). The possible reasons are 1) extreme random variation, 2) genotyping errors, and 3) actual difference, although we cannot realize the true reason(s) for each study. For the studies in Nagano prefecture and Kisei-cho of Mie prefecture, the actual difference is possible because of the relatively isolated areas without persons moving in. For the others, the actual reasons are unknown.

When several datasets with a large sample are available, which should be selected or combined for controls? In this analysis, 0.391 was obtained based on 10,854 subjects. JSNP database provided 0.393 for the T allele, based on 934 inhabitants in Osaka, although the genotype frequency was not in Hardy-Weinberg equilibrium (p=0.0078). JBIC dataset reported 0.403 based on 744 Japanese subjects. Another estimate will be soon available from the Japan Multi-institutional Collaborative Cohort Study (J-MICC Study) launched in 2005 (J-MICC Study Group, 2007) based on more than 4,000 subjects from 10 areas. Accordingly, the rule may be necessary; the allele frequency derived 1) from the source with the most similar characteristics such as area, 2) from that with the allele frequency giving the most conservative results, and 3) from the combined estimate of all available sources.

In order to estimate ORs of genotypes, a genotype frequency of controls is needed. It has to be discussed which is most appropriate, the weighted average of genotype frequency or genotype frequency calculated on the weighted average of allele frequency. Since the average of allele frequency is more stable (1 degree of freedom) than the average of genotype frequency (2 degrees of freedom) in terms of random variation, it would be better that the genotype frequency is calculated based on the adopted allele frequency average, that is, \( p^2 \) for major homozygotes, \( 2p(1-p) \) for heterozygotes, and \( (1-p)^2 \) for the minor homozygotes, where \( p \) is the averages of major allele.

The applicability of the documented allele frequency for controls has to be examined in terms of characteristics of the cases. It is not applicable under the situation that the case group is sampled in an isolated area, where the allele frequency in the non-disease population is expected or known to be different. Generally speaking, the sampling framework on sex, age, occupation, and lifestyle are not related to the applicability.

In conclusion, this example demonstrated that the allele frequency of MTHFR 677T could be common among 24 out of 31 Japanese studies, suggesting that a common allele frequency could be used for Japanese case-control studies, whose cases are sampled in a general population. In the other polymorphisms, the allele frequency documented for many subjects from general populations would be similarly applicable for case-control
studies on the association with genotypes.

Acknowledgements

This study was supported in part by a Grant-in-Aid for Scientific Research on Special Priority Areas of Cancer from the Japanese Ministry of Education, Culture, Sports, Science and Technology (No. 17015018).

References


