Meta Gene 5 (2015) 1–8

Contents lists available at ScienceDirect

Meta Gene

Genetic model selection for a case–control study and a meta-analysis

Nobuyuki Horita *, Takeshi Kaneko

Department of Pulmonology, Yokohama City University Graduate School of Medicine, Yokohama, Japan

ARTICLE INFO

Article history:
Received 12 April 2015
Revised 29 April 2015
Accepted 29 April 2015
Available online 22 May 2015

ABSTRACT

A case–control study often compares the prevalence of a specific disease among persons with normal alleles and persons with variant alleles, which generates an odds ratio (OR). The most common type of allele variation, single-nucleotide polymorphism, consists of a major allele (M) and a minor allele (m). Thus, the genotype can be a major allele homozygote (MM), a heterozygote (Mm) or a minor allele homozygote (mm). Odds are given for each genotype, and a pair of odds generates an OR. Summarizing data using two-by-two contingency is the simplest method of estimating an OR. Thus, dominant, multiplicative, recessive, and over-dominant models are often used. Traditionally, researchers used to calculate ORs using many models and then select the best model from among these calculated ORs. This may cause problems due to multiple comparisons. Therefore, we should choose the best model before calculating the OR for each model. In this article, we will discuss how to choose the best model among many subject-level models when evaluating the impact of the MM/Mm/mm genotype on the disease prevalence.

Introduction

Heredity involves the passing of traits, and occasionally the risk of diseases, to offspring from their parents. This phenomenon was known long before DNA was discovered in the 20th century. Mendelian inheritance is observed for some rare diseases. On the other hand, most common diseases do not present typical Mendelian

Abbreviations: OR, odds ratio; M, Major allele; m, minor allele; MM, homozygote of major allele; Mm, heterozygote; mm, homozygote of minor allele.

* Corresponding author at: 3-9 Fukuura, Kanazawa-ku, Yokohama, 236-0004, Japan. Tel.: +81 45 787 2630; fax: +81 45 786 3444.
E-mail addresses: nobuyuki_horita@yahoo.co.jp (N. Horita), takeshi@med.yokohama-cu.ac.jp (T. Kaneko).

http://dx.doi.org/10.1016/j.mgene.2015.04.003
2214-5400/

inheritance. According to the common disease–common variant hypothesis, some of those common variants lead to susceptibility to complex polygenic diseases. Each variant of each gene that influences a complex disease will have a small effect on the disease phenotype and susceptibility (Marian, 2012; Pritchard and Cox, 2002). Case–control studies, often in the form of genome-wide association studies or meta-analysis, have been conducted to discover causative variants and to evaluate the impact of gene polymorphism on a specific disease.

A case–control study often compares the prevalence of a specific disease among persons with normal alleles and persons with variant alleles, which generates an odds ratio (OR). The most common type of allele variation, single-nucleotide polymorphism, consists of a major allele (M) and a minor allele (m). Thus, the genotype can be a major allele homozygote (MM), a heterozygote (Mm) or a minor allele homozygote (mm). Odds are given for each genotype, and a pair of odds generates an OR (Table 1). Summarizing data using two-by-two contingency is the simplest method of estimating an OR. Therefore, the three kinds of genotypes are often transformed into two variables. For example, a dominant model compares MM versus Mm + mm, and a recessive model compares MM + Mm versus mm. An over-dominant model assumes the heterozygote has the strongest impact and compares MM + mm versus Mm. On the other hand, co-dominant models including additive and multiplicative models hypothesize that MM, Mm, and mm are associated with the lowest, the intermediate, and the highest risk, respectively, or they are associated with the highest, the intermediate, and the lowest risk, respectively (Thakkinstian et al., 2005; Attia et al., 2003). While these models above discuss a subject-level phenomenon, the allelic model evaluates the impact of individual alleles on the disease. This allelic model produces an OR similar to that estimated from the multiplicative model (Thakkinstian et al., 2005; Attia et al., 2003).

Traditionally, researchers used to calculate ORs using many models and then select the best model from among these calculated ORs (Thakkinstian et al., 2005; Attia et al., 2003). This may increase the possibility of type I error due to multiple comparisons (Bagos, 2013). Therefore, we should choose the best model before calculating the OR for each model. Although model selection for genome-wide study was explained by Bagos (Bagos, 2013), another method for model selection for case–control study has been anticipated. In this article, we will discuss how to choose the best model among many subject-level models when evaluating the impact of the MM/Mm/mm genotype on the disease prevalence.

Methods and examples

In this article, for the additive model, we supposed the impact of Mm allele was estimated from the additive mean of impacts of MM and mm alleles. Similarly, for the multiplicative model, we supposed the impact of the Mm allele was estimated from the multiplicative mean of impacts of the MM and mm alleles. Although we knew that some researchers use the wording “log-additive model” instead of “multiplicative model” which is defined above and the wording “additive model” instead of “multiplicative model” which is defined above, we did not use these wordings for the current article.

Table 1
Genotypes, odds and odds ratio (OR).

	MM	Mm	mm
Number of cases	a	b	c
Number of controls	d	e	f
Odd	a/d	b/e	c/f

M: Major allele.
m: minor allele.
MM: homozygote of major allele.
Mm: heterozygote.
mm: homozygote of minor allele.
a, b, c, d, e, f: number of subjects.
We defined OR1 and OR2 as follows:OR1 = odd_{Mm} / odd_{MM} = bd / aeOR2 = odd_{mm} / odd_{Mm} = ce / bf.
Therefore, odd_{Mm} = b / e = odd_{MM} × OR1ori odd_{mm} = c / f = odd_{MM} × OR1ori × OR2ori.