J Hum Genet (2006) 51:387–390
DOI 10.1007/s10038-006-0371-z

SHORT COMMUNICATION

Shin-ichi Kuno · Shiori Furihata · Toshikazu Itou
Kayoko Saito · Naoyuki Kamatani

Unified method for Bayesian calculation of genetic risk

Received: 7 November 2005 / Accepted: 25 December 2005 / Published online: 10 February 2006
© The Japan Society of Human Genetics and Springer-Verlag 2006

Abstract Bayesian inference has been used for genetic risk calculation. In this traditional method, inheritance events are divided into a number of cases under the inheritance model, and some elements of the inheritance model are usually disregarded. We developed a genetic risk calculation program, GRISK, which contains an improved Bayesian risk calculation algorithm to express the outcome of inheritance events with inheritance vectors, a set of ordered genotypes of founders, and mutation vectors, which represent a new idea for description of mutations in a pedigree. GRISK can calculate genetic risk in a common format that allows users to execute the same operation in every case, whereas the traditional risk calculation method requires construction of a calculation table in which the inheritance events are variously divided in each respective case. In addition, GRISK does not disregard any possible events in inheritance. This program was developed as a Japanese macro for Excel to run on Windows.

Keywords Genetic disease · Genetic risk · Counseling · Bayesian theorem · GRISK

S. Kuno (✉) · N. Kamatani
Institute of Rheumatology, Tokyo Women's Medical University,
Tokyo, Japan
E-mail: omoikane@tri-kobe.org
Tel.: +81-78-3032760
Fax: +81-78-3032761

S. Kuno
Translational Research Informatics Center,
Foundation for Biomedical Research and Innovation,
1-5-4 Minatojima-minamimachi, Chuo-ku,
Kobe 650-0047, Japan

S. Kuno
Clinical Genome Informatics Center,
Kobe University Graduate School of Medicine,
Kobe, Japan

S. Furihata
Japan Biological Information Research Center,
Japan Biological Informatics Consortium,
Tokyo, Japan

S. Furihata · T. Itou
Mitsubishi Research Institute, Tokyo, Japan

K. Saito
Institute of Medical Genetics,
Tokyo Women's Medical University,
Tokyo, Japan

N. Kamatani
Division of Genomic Medicine, Department of Advanced
Biomedical Engineering and Science, Tokyo Women's Medical
University, Tokyo, Japan

Introduction

Consultants dealing with genetic diseases require information regarding their risk of crisis of patients and that of their progeny. Presentation of such risks is an important role of genetic counseling. However, the risk of crisis is only an estimate, and there is no way to provide absolutely certain information. However, counselors must provide satisfactory answers to their patients. The Bayesian risk calculation method, which utilizes phenotype data for members of the pedigree, has provided such answers (Bridge 1997; Hodge 1998). In Bayesian risk calculation, expectations of penetrance are calculated with the posterior probabilities of the pedigree members' genotypes for the trait locus. Thus, Bayesian calculation of genetic risk essentially calculates the posterior probabilities of genotypes in the trait locus. In the Bayesian risk calculation method, possible inheritance events are divided into a number of hypotheses under the inheritance model, and genetic risk is calculated with a probability table. Therefore, it is necessary to construct a respective probability table for each case. In addition, some rare events, e.g., homozygosity for the disease-causing allele in diseases with dominant inheritance patterns, are usually disregarded for ease of calculation. We are now developing a new risk calculation program, GRISK, containing an improved Bayesian risk calculation algorithm to express the outcome of inheritance events with inheritance vectors, a set of ordered genotypes of founders, and mutation vectors, representing a new idea for the description of mutation in a pedigree. GRISK can perform genetic risk calculations in a common format, and provides almost the same calculation results as obtained by the traditional method; the results are slightly different due to the disregarding of rare events in the traditional method. As GRISK calculates genetic risk without disregarding any events, the risk calculated using this program is more reliable than that determined by the traditional method. GRISK should ease the burden on counsellors of calculating genetic risk and supply more reliable calculation results.

Algorithm

Inheritance event

As mentioned above, Bayesian calculation of genetic risk essentially calculates the posterior probabilities of genotypes in the trait locus. The genotype of a member is defined as a subset of the sample space of inheritance events. Therefore, it is necessary to define the sample space of inheritance events for Bayesian calculation of genetic risk. The general sample space of inheritance events in a pedigree is defined as all simultaneous events of the ordered genotypes and the phenotypes of all members. However, the phenotype data used in the genetic risk calculation method are the phenotype data of the pedigree's members. Therefore, the sample space of inheritance events in genetic risk calculation is defined as all ordered genotypes of all members in the pedigree under given phenotypes. However, calculations of the prior probability of the pedigree member's genotypes are usually complicated because the genotypes of non-founders are dependent on those of the founders. On the other hand, a combination of an inheritance vector (v) and a set of the ordered genotypes of founders (g) also determines a set of ordered genotypes of all members of a pedigree. Therefore, the outcome of an inheritance event in a pedigree can also be defined as a simultaneous event of v and g. In Genehunter (Kruglyak et al. 1996), an outcome of the inheritance event in a pedigree is defined as described above under all phenotypes of all members in the pedigree (Φ), although some steps are taken to decrease the amount of calculation. However, the definition of outcome in Genehunter cannot describe events with mutations, although mutations must be considered in the calculation of genetic risk. To resolve this problem, we introduced a mutation vector (w) to describe outcomes with mutations. In our method, the outcome of an inheritance event is defined as the simultaneous event of v, g, and w under given Φ. The sample space is then the collection of all possible outcomes defined as described above. The prior probability of an outcome is the product of the prior probabilities of v, g, and w, because these factors are independent of each other.

Mutation vector

The mutation vector is defined as follows: (1) each element is assigned to each allele transmission, and (2) the value of the element is 1 or 0 when a mutation has or has not occurred in allele transmission, respectively. In a pedigree including n_m allele transmissions from a male with a mutation rate, μ_m, and n_f transmissions from a female with a mutation rate, μ_f, the prior probability of w [$\Pr(w)$] including x_m mutations in transmission from the male and x_f mutations in transmission from the female is:

$$\Pr(w) = \mu_m^{x_m}(1-\mu_m)^{n_m-x_m}\mu_f^{x_f}(1-\mu_f)^{n_f-x_f}.$$

If the mutation rate is extremely low, the hypothesis that a lone mutation at most occurs in a pedigree may be appropriate. This simplification can markedly reduce the quantity of calculations. GRISK includes two select modes for mutation vectors to allow multiple mutations in a pedigree or not.

Some outcomes in which a mutation occurs in the disease-causing allele are reverse mutations. As the reverse mutation rate is considered to be much lower than the mutation rate (Muller and Oster 1957; Schlager and Dickie 1971), they are disregarded in this algorithm. Therefore, outcomes with such reverse mutations are excluded.

Bayesian inference

The posterior probability of the outcome is calculated with the likelihood and the prior probability of the outcome by the Bayesian theorem. Equal distributions are assigned to the prior probabilities of the inheritance vectors [$\Pr(v)$] in cases without information regarding the marker loci (cases with information regarding the marker loci are discussed below). The equal distribution in the pedigree containing N non-founders is:

$$\Pr(v) = 2^{-2N}.$$

The prior probabilities of the permutation of the ordered genotypes of the whole F founders [$\Pr(g)$] including k disease-causing allele with a relative frequency of p are calculated as:

$$\Pr(g) = p^k(1-p)^{2F-k}.$$

The phenotype (ϕ_I) is given in the Ith member of the pedigree, and the genotype (γ_I) is determined by v, g, and w. The conditional probability of the phenotype under the genotype of the Ith member, $\Pr(\phi_I|\gamma_I)$, is the penetrance in the symptomatic member, or 1 minus penetrance in the asymptomatic member. Then, the likelihood of the outcome, $\Pr(\Phi|v,g,w)$, is calculated as:

$$\Pr(\Phi|v,g,w) = \prod_i \Pr(\phi_I|\gamma_I).$$

The posterior probability of the outcome, $\Pr(v,g,w|\Phi)$, is then calculated as:

$$\Pr(v,g,w|\Phi) = \frac{\Pr(\Phi|v,g,w)\Pr(v)\Pr(g)\Pr(w)}{\sum\limits_v\sum\limits_g\sum\limits_w \Pr(\Phi|v,g,w)\Pr(v)\Pr(g)\Pr(w)}.$$

Genetic risk calculation

γ_I is a set of corresponding outcomes. For the posterior probabilities of each member's genotype, $\Pr(\gamma_I)$ is calculated, the posterior probabilities of the outcomes, $\Pr(v,g,w|\Phi)$, coinciding with γ_I are summed:

$$\Pr(\gamma_I) = \sum_v\sum_g\sum_w \Pr(v,g,w|\Phi)\Pr(\gamma_I|v,g,w)$$

$\Pr(\gamma_I\ |v,g,w)$ is 1 or 0 in outcomes that do or not coincide with γ_I , respectively. The symptomatic risk in the *I*th individual is calculated from the expectation of penetrance of the respective genotypes in the trait locus.

X-linked inheritance disorder

In loci responsible for disorders with X-linked inheritance, (1) Y chromosomes derived from grandfathers and X chromosomes from grandmothers are definitely transmitted to sons and daughters in transmission from their fathers, respectively; (2) the males' alleles on Y chromosomes transmitted from fathers must not be causing the disorder; and (3) mutations to disorder-causing alleles must not occur on Y chromosomes transmitted from fathers to sons. Therefore, disorders showing X-linked inheritance are characterized by v coinciding with (1), g coinciding with (2), and w coinciding with (3). To calculate the genetic risk of such disorders with X-linked inheritance, outcomes including v, g, or w that do not coincide with these characterizations are excluded.

Utilization of the genotype data of trait loci

In diseases in which the trait loci are known, some members of the pedigree might provide the genotype data of the trait locus. In such cases, utilization of the genotype data of the trait locus should give a more accurate risk calculation. GRISK includes a mode to utilize such data. In this mode, outcomes that do not coincide with a given set of genotype data are excluded.

Utilization of the results of biochemical tests

In the case of diseases caused by genomic deficiencies in metabolic enzymes, prediction of future symptoms is possible via biochemical tests for the activities of the enzymes involved. Members of the pedigree with biochemical tests are assigned to another liability class to allow the results of these tests to be reflected in the risk calculation. Continuous quantitative results of the biochemical tests are classified into two categories, normal and abnormal, based on a threshold. The probabilities of entering the abnormal range of biochemical test results in respective genotypes are defined as penetrance in this liability class. The members of pedigrees with the results of biochemical tests are given another phenotype, normal or abnormal, to replace the actual phenotype, symptomatic or asymptomatic. In this method of calculating genetic risk, the results of biochemical tests are not continuously reflected in risk calculation. However, it is unusual for the results of biochemical tests to relate linearly to symptom probability. Therefore, this method is sufficient to allow utilization of the results of biochemical tests.

Marker loci information

In the traditional method of Bayesian risk calculation, marker loci information is incorporated into the risk calculations to construct haplotypes consisting of trait and marker loci. As combinations of haplotypes are augmented exponentially with increases in marker loci, utilization of information for many marker loci is difficult in the traditional method. On the other hand, inheritance events are represented as inheritance vectors in GRISK. Thus, this program can incorporate marker loci information by the forward–backward conditioning approach employed in Hidden Markov Models (HMM; Rabiner 1989; Kruglyak et al. 1996). As the calculations are augmented only arithmetically with increases in number of marker loci in this method, it is easy to utilize information for many marker loci with our method. As mentioned above, equal distribution is assigned to the prior probabilities of the inheritance vectors in risk calculations without marker information. On the other hand, in risk calculations with marker information, the distribution calculated by the forward–backward conditioning approach is used as prior probabilities instead of equal distribution.

Prenatal diagnosis

The main purpose of genetic risk calculation is prenatal diagnosis in the pedigree with the inheritance disease. The phenotype of the fetus cannot be determined in prenatal diagnosis. However, GRISK can deal with the pedigree of the fetus as described below. The fetus, without phenotype information, is temporarily assigned the asymptomatic phenotype. The fetus is then divided into another liability class in which all genotypes have a penetrance of 0. Pedigree members without phenotype information are also dealt with as above.

Implementation

GRISK was developed as a Japanese macro for Excel to run on Windows. Therefore, the Japanese version of Excel is necessary to run this program on Windows XP. GRISK requires data input to the disease data table and the pedigree data table. The disease data table includes disease gene frequency, mutation rate, penetrance, marker number, recombination fraction, threshold of biochemical tests, and optional information. The pedigree data table includes member's ID, father's ID, mother's ID, sex, phenotype, marker genotype, and the results of biochemical tests. GRISK also has a function to draw a family tree from the pedigree data table. Inquiries regarding GRISK should be addressed to kamatani@ior.twmu.ac.jp.

Discussion

To date, many genes related to hereditary disorders have been detected. In such disorders, the risk of symptoms can be predicted by analysis of the individual's genotype for the genes involved. However, the inference of genetic risk is frequently required in disorders for which the cause has not yet been determined or in cases in which genetic analyses are difficult. In such cases, Bayesian calculation of genetic risk has been used. In Bayesian calculation of genetic risk, all possible hypotheses of inheritance events must be constructed from a given set of pedigree data, and the posterior probabilities of the respective hypotheses are calculated. In the traditional method of Bayesian genetic risk calculation, the whole sample space of possible inheritance events is divided with tables. However, the table framework must be changed for each pedigree and each disorder. In addition, to prevent complication of the table framework, rare events, such as homozygosity for the disease-causing allele in dominant diseases, are usually disregarded. On the other hand, the essential point of GRISK is that the whole sample space of the inheritance events is divided into most subdivisional outcomes with inheritance vectors, ordered genotypes of founders, and mutation vectors. Therefore, a common framework can be used in all pedigrees and disorders. However, the use of inheritance and mutation vectors gives GRISK a disadvantage in calculations in large pedigrees. The amount of calculation is augmented exponentially with increases in the number of pedigree members. Therefore, GRISK is best suited for calculations in small pedigrees.

GRISK has the advantage that it does not disregard any possible events in inheritance. The genetic risks calculated by GRISK were almost same as those determined by the traditional Bayesian method; slight differences were found as the traditional method disregards rare events. The genetic risks calculated by GRISK, which disregards no events, should be more accurate than those determined by the traditional method. However, future comparative evaluation between GRISK and risk calculation programs with the traditional method using simulation or real data will be necessary to ascertain the accuracy of the calculation and to evaluate other performance parameters.

GRISK also has another advantage with regard to utilization of marker loci information. In the traditional Bayesian risk calculation method, the amount of calculation is augmented exponentially with increases in the number of marker loci. On the other hand, the calculations are augmented only arithmetically with increasing marker loci number in GRISK because this program uses HMM marker loci information. Therefore, GRISK is suited for calculations with many marker loci. The utilization of many marker loci should be effective in the case of single nucleotide polymorphisms less informative than microsatellites.

In conclusion, GRISK should help counselors calculate genetic risk and supply more reliable calculation results than the traditional Bayesian risk calculation method.

Acknowledgements This work was supported by grants from New Energy and Industrial Technology Development Organization.

References

Bridge PJ (1997) The calculation of genetic risks, 2nd edn. The Johns Hopkins University Press, Baltimore

Hodge SE (1998) A simple, unified approach to Bayesian risk calculations. J Genetic Counsel 7:235–261

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Muller HJ, Oster II (1957) Principles of back mutation as observed in Drosophila and other organisms. Adv Radiobiol 1957:407–415

Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 77:257–286

Schlager G, Dickie MM (1971) Natural mutation rates in the house mouse. Estimates for five specific loci and dominant mutations. Mutat Res 11:89–96