

On the Different Ways to Handle the Trend of Disease Risk in Genetic Association Tests

Tapati Basak

Department of Statistics, Jahangirnagar University, Savar, Bangladesh

Email: tapati.basak@juniv.edu

How to cite this paper: Basak, T. (2022) On the Different Ways to Handle the Trend of Disease Risk in Genetic Association Tests. *Open Journal of Statistics*, 12, 521-531. <https://doi.org/10.4236/ojs.2022.124031>

Received: July 11, 2022

Accepted: August 16, 2022

Published: August 19, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Genetic association studies usually apply the simple chi-square (χ^2)-test for testing association between a single-nucleotide polymorphism (SNP) and a particular phenotype, assuming the genotypes and phenotypes are independent. So, the conventional χ^2 -test does not consider the increased risk of an individual carrying the increasing number of disease responsible allele (a particular genotype). But, the association tests should be performed with the consideration of this disease risk according to the mode of inheritance (additive, dominant, recessive). Practical demonstration of the two possible methods for considering such order or trends in contingency tables of genetic association studies using SNP genotype data is the purpose of this paper. One method is by pooling the genotypes, and the other is scoring the individual genotypes, based on the disease risk according to the inheritance pattern. The results show that the p -values obtained from both the methods are similar for the dominant and recessive models. The other important features of the methods were also extracted using the SNP genotype data for different inheritance patterns.

Keywords

Association, Chi-Square, Trend Test, SNP, Genotype

1. Introduction

The disease aetiology has been well understood from the blessings of the genome-wide association study (GWAS). GWAS detects the association between genetic variants and disease traits using samples from a given population. The findings of GWAS data have opened a new clinical insight. This leads to novel bioinformatic advances in processing and interpreting GWAS summary data that enabled the detection of novel disease variants and gene loci [1]-[8].

Testing association is a crucial part of GWAS as the positive genes are reported from this inferential procedure in a case-control study design. The contingency table tests for individual single-nucleotide polymorphism (SNP) are carried out here, where, the individual genotype counts are handled with the phenotype (case-control) [3] [5] [6] [9] [10] [11] [12]. Generally, the two degrees of freedom (d.f.) simple chi-square (χ^2)-test is applied for the contingency table for testing association [13].

But, this conventional χ^2 -test does not consider risk of developing disease for carrying a particular genotype. That is, instead of assuming that the conditional probability of being affected with disease having a specific genotype, an independence association between the genotype and phenotype is assumed. For this independence assumption, the simple χ^2 -test has no logic of ordering or trend of genotype based on the disease risk. Because, different people having different genotype combinations will produce varying risks of developing a particular disease as frequency of risk allele will differ with respect to the difference of genotype.

Defining the penetrance function is the way to model the relationship between SNPs and disease risk by considering such order [14] [15] [16]. This function measures the probability for occurring a particular phenotype for a given genotype [11] [12] [17] [18]. For each inheritance pattern (recessive, dominant, additive), the penetrance can be defined by a mathematical model [10] [11] [17].

There are two ways to include this trend or order of genotypes in the contingency table. One is to rearrange or pool of genotype counts of the table with the consideration of alternative model of penetrance [12] [19], and the other is applied by specifying a score vector for each of the models.

The main objective of this paper is to demonstrate a practical application of the two different ways of considering the order or the trend of the genotypes in GWAS association tests for SNP genotype deoxyribonucleic acid (DNA) sequencing data.

The organization of this paper is as follows. Section 2 presents how the SNP genotype data can be organized in a contingency table. Different ways of testing association for both the ordered and unordered genotype data are outlined in Section 3. The description of the genotype data used in this analysis is provided in Section 4. Section 5 presents the results obtained from the analysis of SNP genotype data using the tests described in Section 3. Finally, some concluding remarks are given in Section 6.

2. Tabular Presentation of SNP Genotype Data: Contingency Table

The contingency table is the most popular tabular method for presenting the genotype data obtained from a case-control study. For any SNP, the data can be summarized in a 2×3 contingency table as in **Table 1**. Assume that, “ M ” and “ m ” are the two alleles for a SNP. The generated data consist of six counts of the

numbers of genotypes (M/M , M/m and m/m) in a case control study, where, n_1 , n_2 , n_3 , n_4 , n_5 , n_6 are the genotype counts observed in the cases and controls, respectively. Here, sample size (table sum) $=N$, total of cases $=n_{\text{cases}}$, total of controls $=n_{\text{controls}}$ and the total number of M/M genotypes observed is $n_{M/M}$ and so on. Thus, **Table 1** is a tabular presentation of the genotype data with the binary status of any particular phenotype (case-control: disease status) [11].

Table 1. Contingency table for any SNP (For “ M ” and “ m ” alleles).

	M/M	M/m	m/m
Cases	n_1	n_2	n_3
Controls	n_4	n_5	n_6

3. Contingency Table Analysis

3.1. Testing for the Unordered Genotypes

The 2×3 contingency table of genotype counts (**Table 1**) can be directly analyzed by a statistical test that is applied to sets of categorical data to evaluate how likely it is that any observed difference between the sets arose by chance. This observed-expected test statistic has a chi-squared χ^2 -distribution with two (2) d.f. [13] [20].

This χ^2 -statistic tests for the deviations from the expected values across cells of the table. Here, the comparison is made between the observed values of a particular genotype with its expected values. For example, the observed value for M/M genotype in cases ($O_1 = n_1$) is compared with its expected value given the total number of cases (n_{cases}) and the total number of M/M genotypes, and hence, $E_1 = n_{\text{cases}} n_{M/M} / N$. Now, the test statistic is,

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

The statistic in Equation (1) follows χ^2 -distribution with 2 d.f., where, $n = 1, 2, 3, 4, 5, 6$ presents the number of cells in the table (**Table 1**). The test statistic defines (Equation (1)) the summation over all the six cells of the table, where, O_i represents the observed cell counts for each of the six cells: n_1 , n_2 , n_3 , n_4 , n_5 , n_6 . Here, under the null hypothesis of no association, the test statistic compares the observed number of M/M genotypes in cases with the corresponding expected assuming that the relative allele (or genotype frequencies) to be the same in case and control groups for the M/M genotype [4] [20].

Usually GWAS emphasis on the associations between single SNP and a trait viz. major human diseases. The association study includes 2×3 contingency table (**Table 1**) tests by the simple χ^2 -test (Equation (1)) [13] for each of the SNP for a gene. A gene is reported as associated with a particular phenotype if any of the SNPs of the gene is significant in comparison of the test p -values with the GWAS threshold, $p = 5 \times 10^{-08}$ [9].

3.2. Testing for the Ordered Genotypes

Pooling the genotype counts: For dominant and recessive models

The concept of ordering genotypes based on the disease risk is not considered in the above χ^2 -test (Section 3.1). The disease risk for an individual is defined from the genotype or allele at a specific marker. In the above χ^2 -test, the independence between the binary phenotype and the individual genotypes was assumed. But, in practice, the risk of developing a particular disease for each person is not the same as different person have different genotype combination, where the frequency of the disease responsible allele would not be the similar.

This order or trend of genotypes can be included in the association tests of contingency tables by specifying the disease penetrance with respect to a penetrance model. Rearrangement or pooling of the genotype counts is one way to consider this order in association studies [19].

The full genotype table for a general genetic model provides the unordered genotype counts for a single SNP (Table 1). Let us demonstrate how to include the concept of penetrance by rearranging the counts of Table 1, which specifies a genetic model at prior, where, “ m ” is the disease responsible allele.

If the hypothesis is, carrying any number of copies of allele “ m ” increases the disease risk, then the assumed model is dominant. This implies that one or two copies of disease responsible allele are required to increase the risk of an individual. Hence, the counts for the M/m and m/m genotypes are to be pooled in Table 1, and thus produce a 2×2 table of genotype counts for the dominant model (Table 2).

Table 2. For the dominant genetic model (M/M versus both M/m and m/m combined).

	M/M	$M/m + m/m$
Cases	n_1	$n_2 + n_3$
Controls	n_4	$n_5 + n_6$

On the other hand, if the hypothesis is, carrying two copies of disease responsible allele “ m ” increases the disease risk for an individual, then the assumed model is recessive. So, the counts for the M/M and M/m genotypes are to be pooled in Table 1, and thus produce a 2×2 table of genotype counts for the recessive model (Table 3).

Table 3. For the recessive genetic model (m/m versus both M/M and M/m combined).

	$M/M + M/m$	m/m
Cases	$n_1 + n_2$	n_3
Controls	$n_4 + n_5$	n_6

A χ^2 -test with one (1) d.f. is used for these 2×2 tables (Table 2 and Table 3) of case-control allele counts, which is the widely used allelic association test. This 1 d.f. allelic test is more powerful than the genotypic test with 2 d.f. (Section

3.1), under certain conditions of the penetrance parameter [19] [20].

Association testing based on the allelic counts gives an alternative method in case-control association studies. This approach splits the genotypes and hence compares the total number of M and m alleles in cases and controls, regardless of the genotypes from which these alleles are being counted [19] [20].

Using score vectors: For dominant, recessive and additive models

The trend of disease risk can be quantified using the penetrance models of additive, dominant and recessive, according to number of increase of disease responsible alleles [19] [20] [21]. The association of trends is tested by the Cochran-Armitage trend test. This trend test leads the χ^2 -test to the comparatively narrower alternatives [19] [22] [23]. Here, the trend or ordering of genotype is implemented by specifying a score vector for each of the models. The scores are to be chosen to construct this trend test. Let us represent **Table 1** as in the form of **Table 4**, for the notational convenience of defining the Cochran-Armitage trend test statistic with the scores for each of the genotypes.

Table 4. Scores assigned for each of the genotypes in a contingency table for any SNP.

	Genotypes			Total
	M/M	M/m	m/m	
Cases	n_{10}	n_{11}	n_{12}	N_1
Controls	n_{00}	n_{01}	n_{02}	N_0
Total	N_{+0}	N_{+1}	N_{+2}	N
Scores	x_0	x_1	x_2	

For testing association of the contingency table (**Table 4**), the Armitage's trend test statistic is [22],

$$\chi_A^2 = \frac{N \left(N \sum n_{1j} x_j - N_1 \sum N_{+j} x_j \right)^2}{N_1 N_0 \left\{ N \sum N_{+j} x_j^2 - \left(\sum N_{+j} x_j \right)^2 \right\}} \quad (2)$$

The statistic in Equation (2) is approximately distributed as χ^2 with 1 d.f., and the validity of the test will not be affected regardless the choice of the scoring system.

For defining the values of the score vector for different models (additive, dominant and recessive), the disease penetrance associated with a given genotype are to be considered by defining a probabilistic function [11] [17] [18]. This function provides an approach to define an appropriate model from the relationship between a particular SNP, and the disease risk by considering the trend or order of the genotype [12] [14] [15] [16].

It is known that the risk of developing disease for an individual increases for having any number of copies or two copies of disease responsible allele, for the case of dominant or recessive inheritance pattern. For additive genetic model, the risk have a clear linear trend with the increased number of disease responsi-

ble allele. For example, the risk for the M/m genotypes is approximately half that for the m/m .

Hence, for a penetrance parameter γ ($\gamma > 1$), the risk of an individual carrying a particular genotype (or number of the disease responsible allele) for the three genetic models can be rephrased. For an γ -fold increase in individual's disease risk one or two copies of "m" allele are required for a dominant model, and two copies of "m" allele are required for a recessive model. An additive model assumes γ -fold increase for the genotype M/m and 2γ -fold for the genotype m/m [10] [19] [20] [21].

The three common choices for the scoring system with the reference of the model definition as given above in terms of the penetrance parameter (γ) are: Additive score: $x_0 = 0$, $x_1 = 1$, $x_2 = 2$; Dominant score: $x_0 = 0$, $x_1 = 1$, $x_2 = 1$ and recessive score: $x_0 = 0$, $x_1 = 0$, $x_2 = 1$.

4. Genotype Data Preparation

First, the SNP genotype data for single SNP was generated via computer simulation in R programming language for 3,000 individuals. Individuals were then assigned at random to the cases and controls with the equal probabilities of cases and controls: (0.5, 0.5) (Data 1). Then, another data for each of the gene was generated through the simulation in R for the same number of individuals with the random assignment of the equal probabilities to the cases and controls as for Data 1 (Data 2).

5. Application to Genotype Data: For Single SNP

5.1. For Unordered Genotypes

The following 2×3 contingency table (**Genotype Table 1**) is presenting genotype counts for a randomly selected gene with single SNP, constructed from Data 1 (Section 4).

Genotype Table 1. For single SNP ("m" is disease responsible allele).

	<i>M/M</i>	<i>M/m</i>	<i>m/m</i>
Cases	578	278	606
Controls	774	314	450

Testing association for the null hypothesis of no association using the 2 d.f. χ^2 -test (Section 3.1) for **Genotype Table 1**, produces the p -value = 5.769914×10^{-12} . Hence, this SNP is significant as compared with the GWAS threshold, $p = 5 \times 10^{-8}$ [9].

5.2. For Ordered Genotypes

Using pooling for dominant and recessive models

In order to include the ordering of the genotypes in association tests for the

genetic models dominant and recessive, respectively, the counts of the columns of **Genotype Table 1** are rearranged (**Genotype Table 2** and **Genotype Table 3**) according to the definition given in Section 3.2.

Genotype Table 2. For the dominant genetic model.

	<i>M/M</i>	<i>M/m + m/m</i>
Cases	578	884
Controls	774	764

Genotype Table 3. For the recessive genetic model.

	<i>M/M + M/m</i>	<i>m/m</i>
Cases	856	606
Controls	1088	450

The 1 d.f. χ^2 -tests are applied for testing association of the above two tables (**Genotype Table 2** and **Genotype Table 3**), and the recorded p -values are summarized in **Table 5**.

Setting score vectors of penetrance models

Assigning score vectors according to the mode of inheritance is the alternative way to include the trend in association testing (**Genotype Table 4**). The Cochran-Armitage trend test is applied here for testing genetic association (Section 3.2).

Genotype Table 4. Scoring genotypes.

	<i>M/M</i>	<i>M/m</i>	<i>m/m</i>
Cases	578	278	606
Controls	774	314	450
Scores (Additive)	0	1	2
Scores (Dominant)	0	1	1
Scores (Recessive)	0	0	1

The p -values obtained from these tests are summarized in **Table 5** along with the p -values from the above tests of pooling genotypes.

Table 5. The p -values from the pooling and scoring tests of trend for single SNP.

Models	Tests	
	Using pooled genotypes	Using score vectors
Scores (Dominant)	3.625972×10^{-09}	2.900783×10^{-09}
Scores (Recessive)	3.646926×10^{-12}	2.779008×10^{-12}
Scores (Additive)		1.598713×10^{-12}
	$p\text{-value} (\chi^2_2) = 5.769914 \times 10^{-12}$	

The SNP is significant for all the cases shown in **Table 5** as compared with the GWAS threshold of $p = 5 \times 10^{-08}$. But, different p -values are producing for all the cases. As trends are considering in both methods such as pooling genotypes and the score vector cases, the dominant model tests are producing almost similar p -values. The scenarios are the same for the recessive model. The p -value from the test for the unordered genotypes (χ^2_2) is the most different from the model tests for trend. Overall, the p -values from the recessive model tests are smaller for the scoring case, and the Cochran-Armitage trend test for the additive model is giving smallest p -value among all the tests.

5.3. Features for the Genes with Multiple SNPs

To extract the gene wise features of the two ways (pooled; scoring), the dominant and recessive models were applied for the genotype data (Data 2). Three genes having multiple SNPs were selected randomly from Data 2 (Section 4). There are 3, 5 and 1,000 SNPs in the selected 3 genes, GENE1, GENE2 and GENE3, respectively. Individual SNP tests were performed for each of the three genes using 1 d.f. tests by considering both of the above mentioned methods that is pooling the genotypes and assigning the score vectors for the genotypes according to the definition of the dominant and recessive models, respectively. The p -values for GENE1 and GENE2 are shown in **Table 6**.

Table 6. The p -values for the two genes with 3 and 5 SNPs from the pooling and scoring tests of trend.

Models	GENE1		GENE2	
	Tests		Tests	
	Using pooled genotypes	Using score vectors	Using pooled genotypes	Using score vectors
Dominant	3.766438×10^{-11}	2.936793×10^{-11}	2.907425×10^{-07}	2.391704×10^{-07}
	1.919598×10^{-06}	1.599422×10^{-06}	5.427484×10^{-10}	4.294702×10^{-10}
	6.438253×10^{-08}	5.242511×10^{-08}	7.449102×10^{-04}	6.512162×10^{-04}
			3.792536×10^{-09}	3.032072×10^{-09}
			2.135144×10^{-07}	1.751733×10^{-07}
		2.376526×10^{-11}	1.830878×10^{-11}	1.008278×10^{-09}
Recessive	3.033727×10^{-07}	2.476262×10^{-07}	7.357196×10^{-14}	5.489098×10^{-14}
	7.558394×10^{-08}	6.103088×10^{-08}	2.458697×10^{-05}	2.073627×10^{-05}
			4.959689×10^{-11}	3.840915×10^{-11}
			3.488981×10^{-09}	2.763850×10^{-09}

From **Table 6**, it is observed that the obtained p -values from both of the methods are almost similar for the dominant model for the two genes having multiple SNPs. This scenario is also same for the recessive model. Generally, the

p -values from the recessive model tests are smaller for the scored vector case for both of the two genes.

Gene wise features for the methods were also investigated for the GENE3 that has huge numbers of SNPs (1000). The p -values from this investigation are presented in **Figure 1**. Here, the p -values are plotted in the negative-log-transformed scale ($-\log_{10}(p)$). The p -values obtained from the two methods are plotted in the two panels of **Figure 1**, where, **Figure 1(a)** for the dominant model and **Figure 1(b)** for the recessive model.

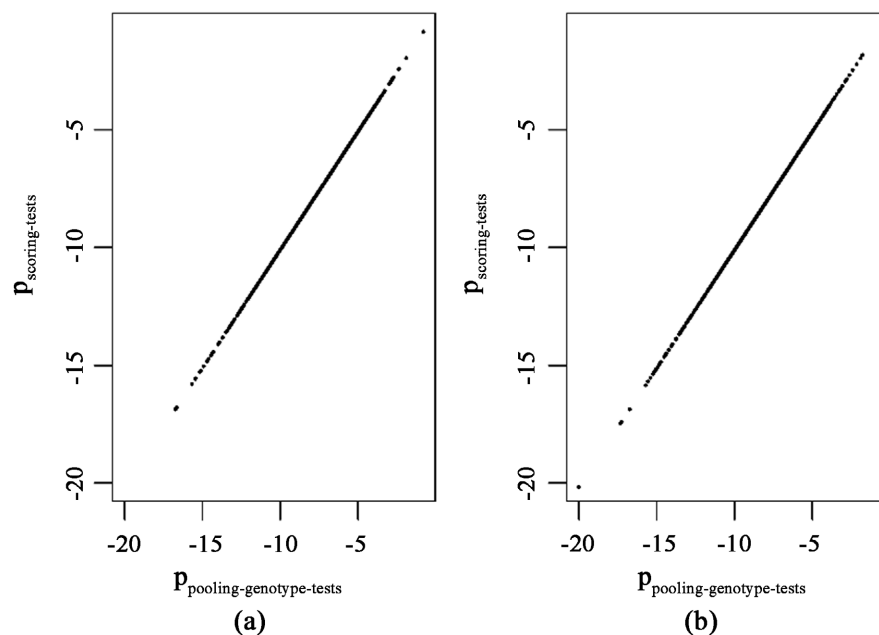


Figure 1. The plot of the p -values obtained from the tests using pooled genotypes along with the tests using score vectors for GENE3 having 1000 SNPs. (a) Dominant model. (b) Recessive model.

From the **Figure 1**, it is observed that each of the two panels is showing almost straight lines. Hence, the p -values obtained from two methods for each model are nearly similar having almost perfect positive relation. This feature is the same as observed for the GENE1, GENE2 (**Table 6**), and also for the single SNP cases (**Table 5**).

6. Conclusion

This paper presents the possible ways of considering genotype ordering in contingency table tests of genetic association by applying trend test. Though, this research used simulated genotype data, but, the methods could also be applied for the real genotype data. As the basic structure of the simulated and the real data are the same, so, the directions or pattern of the obtained results would be the same in both cases. Both the mathematical and practical demonstrations are provided here. Polling of the genotype counts and assigning the score of the genotypes of a contingency table are two possible ways to consider the trend or

order of genotypes according to the mode of inheritance (additive, dominant, recessive). The dominant and recessive model tests can be performed in either way that is by pooling genotypes and using scoring of the genotypes. The additive model could be tested by the method of scoring genotypes. The results show that the two ways are producing almost similar p -values for the dominant and recessive cases regardless of the number of SNPs.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Beck, T., Shorter, T. and Brookes, A.J. (2020) GWAS Central: A Comprehensive Resource for the Discovery and Comparison of Genotype and Phenotype Data from Genome-Wide Association Studies. *Nucleic Acids Research*, **48**, D933-D940. <https://doi.org/10.1093/nar/gkz895>
- [2] Svishcheva, G.R., Belonogova, N.M., Zorkoltseva, I.V., Kirichenko, A.V. and Axenovich, T.I. (2019) Gene-Based Association Tests Using GWAS Summary Statistics. *Bioinformatics*, **35**, 3701-3708. <https://doi.org/10.1093/bioinformatics/btz172>
- [3] Cao, X., Wang, X., Zhang, S., and Sha, Q. (2022) Gene-Based Association Tests Using GWAS Summary Statistics and Incorporating eQTL. *Scientific Reports*, **12**, Article No. 3553. <https://doi.org/10.1038/s41598-022-07465-0>
- [4] Wang, Y., Li, Y., Hao, M., Liu, X., Zhang, M., Wang, J., Xiong, M., Shugart, Y.Y., and Jin, L. (2019) Robust Reference Powered Association Test of Genome-Wide Association Studies. *Frontiers in Genetics*, **10**, Article No. 319. <https://doi.org/10.3389/fgene.2019.00319>
- [5] Uffelmann, E., Huang, Q.Q., Munung, N.S., de Vries, J., Okada, Y., Martin, A.R., Martin, H.C., Lappalainen, T. and Posthuma, D. (2021) Genome-Wide Association Studies. *Nature Reviews Methods Primers*, **1**, Article No. 59. <https://doi.org/10.1038/s43586-021-00056-9>
- [6] Boua, P.R., Brandenburg, J.T., Choudhury, A., Sorgho, H., Nonterah, E.A., Agongo, G., Asiki, G., Micklesfield, L., Choma, S., Gómez-Olivé, F.X., Hazelhurst, S., Tinto, H., Crowther, N.J., Mathew, C.G., Ramsay, M., AWI-Gen Study and the H3Africa Consortium (2022) Genetic Associations with Carotid Intima-Media Thickness Link to Atherosclerosis with Sex-Specific Effects in Sub-Saharan Africans. *Nature Communications*, **13**, Article No. 855. <https://doi.org/10.1038/s41467-022-28276-x>
- [7] Tängdén, T., Gustafsson, S., Rao, A.S. and Ingelsson, E. (2022) A Genome-Wide Association Study in a Large Community-Based Cohort Identifies Multiple Loci Associated with Susceptibility to Bacterial and Viral Infections. *Scientific Reports*, **12**, Article No. 2582. <https://doi.org/10.1038/s41598-022-05838-z>
- [8] Loos, R.J.F. (2020) 15 Years of Genome-Wide Association Studies and No Signs of Slowing down. *Nature Communications*, **11**, Article No. 5900. <https://doi.org/10.1038/s41467-020-19653-5>
- [9] Patron, J., Serra-Cayuela, A., Han, B., Li, C. and Wishart, D.S. (2019) Assessing the Performance of Genome-Wide Association Studies for Predicting Disease Risk. *PLOS ONE*, **14**, Article ID: e0220215. <https://doi.org/10.1371/journal.pone.0220215>
- [10] Bush, W.S. and Moore, J.H. (2012) Chapter 11: Genome-Wide Association Studies.

- PLOS Computational Biology*, **8**, Article ID: e1002822.
<https://doi.org/10.1371/journal.pcbi.1002822>
- [11] Setu, T.J. and Basak, T. (2021) An Introduction to Basic Statistical Models in Genetics. *Open Journal of Statistics*, **11**, 1017-1025.
<https://doi.org/10.4236/ojs.2021.116060>
- [12] Basak, T. and Roy, N. (2022) A Preliminary Outline of the Statistical Inference Process in Genetic Association Studies. *Open Journal of Statistics*, **12**, 200-209.
<https://doi.org/10.4236/ojs.2022.122014>
- [13] Plackett, R.L. (1983) Karl Pearson and the Chi-Squared Test. *International Statistical Review*, **51**, 59-72. <https://doi.org/10.2307/1402731>
<https://www.jstor.org/stable/1402731>
- [14] Moore, J.H., Hahn, L.W., Ritchie, M.D., Thornton, T.A. and White, B.C. (2004) Routine Discovery of Complex Genetic Models using Genetic Algorithms. *Applied Soft Computing*, **4**, 79-86. <https://doi.org/10.1016/j.asoc.2003.08.003>
- [15] Cooper, D.N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. and Kehrer-Sawatzk, H. (2013) Where Genotype Is Not Predictive of Phenotype: Towards an Understanding of the Molecular Basis of Reduced Penetrance in Human Inherited Disease. *Human Genetics*, **132**, 1077-1130.
<https://doi.org/10.1007/s00439-013-1331-2>
- [16] Ford, D., Easton, D.F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D.T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M.D., Struwing, J., Arason, A., Scherneck, S., Peto, J., Rebbeck, T.R., Tonin, P., Neuhausen, S., Barkardottir, R., Eyfjord, J., Lynch, H., Ponder, B.A.J., Gayther, S.A., Birch, J.M., Lindblom, A., Stoppa-Lyonnet, D., Bignon, Y., Borg, A., Hamann, U., Haites, N., Scott, R.J., Maugard, C.M., Vasen, H., Seitz, S., Cannon-Albright, L.A., Schofield, A., Zelada-Hedman, M. and The Breast Cancer Linkage Consortium (1998) Genetic Heterogeneity and Penetrance Analysis of the BRCA1 and BRCA2 Genes in Breast Cancer Families. *American Journal of Human Genetics*, **62**, 676-689.
<https://doi.org/10.1086/301749>
- [17] Ziegler, A. and König, I.R. (2010) A Statistical Approach to Genetic Epidemiology: Concepts and Applications. 2nd Edition, Wiley-VCH, Weinheim.
<https://doi.org/10.1002/9783527633654>
- [18] Gong, G., Hannon, N. and Whittemore, A.S. (2010) Estimating Gene Penetrance from Family Data. *Genetic Epidemiology*, **34**, 373-381.
<https://doi.org/10.1002/gepi.20493>
- [19] Clarke, G.M., Anderson, C.A., Pettersson, F.H., Cardon, L.R., Morris, A.P. and Zondervan, K.T. (2011) Basic Statistical Analysis in Genetic Case-Control Studies. *Nature Protocols*, **6**, 121-133. <https://doi.org/10.1038/nprot.2010.182>
- [20] Lewis, C.M. (2002) Genetic Association Studies: Design, Analysis and Interpretation. *Briefings in Bioinformatics*, **3**, 146-153. <https://doi.org/10.1093/bib/3.2.146>
- [21] Fang, Y., Wang, Y. and Sha, N. (2009) Armitage's Trend Test for Genome-Wide Association Analysis: One-Sided or Two-Sided? *BMC Proceedings*, **3**, Article No. Article No. S37. <https://doi.org/10.1186/1753-6561-3-S7-S37>
- [22] Armitage, P. (1955) Tests for Linear Trends in Proportions and Frequencies. *Biometrics*, **11**, 375-386. <https://doi.org/10.2307/3001775>
<https://www.jstor.org/stable/3001775>
- [23] Cochran, W.G. (1954) Some Methods for Strengthening the Common χ^2 Test. *Biometrics*, **10**, 417-451. <https://doi.org/10.2307/3001616>
<https://www.jstor.org/stable/3001616>