Published in final edited form as:

Genet Epidemiol. 2011 September ; 35(6): 568–571. doi:10.1002/gepi.20592.

When Is Absence of Evidence, Evidence of Absence? Use of Equivalence-Based Analyses in Genetic Epidemiology and a Conclusion for the *KIF1B* rs10492972*C Allelic Association in Multiple Sclerosis

Pierre-Antoine Gourraud¹ and The International Multiple Sclerosis Genetics Consortium (IMSGC)²

¹Department of Neurology, School of Medicine, University of California, San Francisco, San Francisco, CA

²International Multiple Sclerosis Genetics Consortium, Harvard Medical School, Goldenson Building, Boston, MA

Abstract

Statistical equivalence methods have been in development since the late 1980s in order to provide an appropriate statistical methodology to address nondifferences in biological experiments. This is analogous to genetic association studies in which a polymorphism “is not associated” with a trait. We applied the equivalence method to genetic data to confirm that an association between the *KIF1B* (kinesin family member 1B) rs10492972 allele and multiple sclerosis (MS), reported in *Nature Genetics* in 2008, is not present in eight datasets of cases and controls, nor in three independent datasets of the International Multiple Sclerosis Genetic Consortium. When the datasets are considered together, a nonsuperiority test excludes the rs10492972*C allele as a major “risk” allele for MS with a high degree of confidence ($p = 1.18 \times 10^{-4}$). We propose that equivalence methods are more appropriate for stating that a polymorphism does not contribute to disease susceptibility. If an equivalence test applied to genetic datasets fails to reveal an association based on standard methods, it demonstrates that there is no genetic association—i.e., the absence of evidence is evidence of absence. When reporting genetic association based on a cohort of a limited size, caution is needed regardless of how attractive the underlying biological rationale is. The data gathered for *KIF1B* in MS also underscore the need for very large sample sizes with the appropriate equivalence statistical methods in order to exclude reported false-positive results.

Keywords

negative results; bioequivalence; multiple sclerosis; *KIF1B*

INTRODUCTION

An association between a variation at a *KIF1B* (kinesin family member 1B) locus (rs10492972) and multiple sclerosis (MS) was initially reported in a genome-wide association study (GWAS) that included 45 MS cases and 195 controls from a genetically

Correspondence to: Pierre-Antoine Gourraud, Department of Neurology, School of Medicine, University of California, San Francisco, 513 Parnassus Ave. S-258 San Francisco, CA 94143-0435. Phone: +1 415-476-2508; Fax: +1 415-476-5229; pierreantoine.gourraud@ucsf.edu.

isolated Dutch population, and reported an odds ratio of 1.35 ($p = 2.5 \times 10^{-10}$) for a pooled dataset that included samples from the isolated Dutch population and unrelated Swedish and Canadian populations [Aulchenko, et al. 2008]. The kinesin motor protein Kif1b has previously been implicated in the axonal transport of mitochondria and synaptic vesicles, and a recent study using a zebrafish model suggested that Kif1b is required for the localization of myelin basic protein mRNA to processes of myelinating oligodendrocytes [Lyons, et al. 2009]. Unfortunately, subsequent efforts in an Italian primary progressive MS dataset [Martinelli-Boneschi, et al. 2010] and in a large multicenter study by Ban et al. on behalf of the International Multiple Sclerosis Genetics Consortium (IMSGC) [IMSGC, et al. 2010] found no evidence for association, suggesting that the *KIF1B* association represents a false positive. Population stratification, lack of power, genotyping errors, gene–environment population-specific interactions, and prevalence–incidence bias were suggested as possible sources for the discrepancy.

As shown in the field of bioequivalence clinical trials, the use of classical difference testing may not be optimal for assessing an absence of association when dealing with the modest effects that characterize genetic risk factors in complex diseases [Altman and Bland 1995; Schuirmann 1987]. While difference testing reveals the probability of observing an association by chance, equivalency testing yields the probability of observing a lack of association by chance. Equivalence testing is often implemented as two one-sided tests: one test that gives the probability of observing a lack of association if the actual association is positive and another test that gives the probability of observing a lack of association if the actual association is negative [Schuirmann 1987].

This methodology is well accepted in classical epidemiology. Because a nonsignificant *difference* test cannot be interpreted as acceptance of the null hypothesis, equivalence tests have repeatedly proven of great utility in clinical trials as a means to show no difference between alternative treatment modalities both for quantitative and qualitative outcome measurements. The relevance of the statistical method certainly extends to the field of genetic epidemiology because interpretation of nonsignificant results in underpowered genetic association studies remains problematic. We thus applied an equivalence-based method in a nonsuperiority mode to the IMSGC data [IMSGC, et al. 2010] to test the nonassociation hypothesis between rs10492972*C and MS susceptibility in a statistically more appropriate manner.

METHODS AND RESULTS

The IMSGC data was re-analyzed using the nonsuperiority statistical test as implemented in computing equivalence asymptotic interval [Barker, et al. 2001]. The hypothesis that the frequency of rs10492972*C carrier is greater by a specified percentage compared with the frequency in controls is the null hypothesis of the test (H_0)

$$H_0: \widehat{P}_{\text{cases}} > (\widehat{P}_{\text{controls}} + \Delta\%); \Delta = 1\%$$

$$H_1: \widehat{P}_{\text{cases}} < (\widehat{P}_{\text{controls}} + \Delta\%)$$

Where $\widehat{P}_{\text{cases}}$ and $\widehat{P}_{\text{controls}}$ are the carrier frequency of the C allele of rs10492972 in case and control samples, respectively; n_{Cases} and n_{Controls} are the sample sizes in cases and controls samples, respectively.

$$\frac{\widehat{P}_{\text{cases}} - (\widehat{P}_{\text{controls}} + \Delta\%) }{\sqrt{\frac{\widehat{P}_{\text{cases}} \times (1 - \widehat{P}_{\text{cases}})}{n_{\text{cases}}} + \frac{\widehat{P}_{\text{controls}} \times (1 - \widehat{P}_{\text{controls}})}{n_{\text{controls}}}}} \rightarrow N(0, 1)$$

EQUATION (1)

The specified amount Δ may be arbitrary, or as we did here, it may be taken from the lower boundary of the difference estimated in the initial report of the genetic association. The frequencies reported by [Aulchenko, et al. 2008] were (controls vs. cases): 37.44% [30.6, 44.23] ($n=195$) vs. 66.67% [52.9, 80.44] ($n=45$) for the isolated Dutch population; 46.01% [41.3, 50.74] ($n=426$) vs. 56.94% [52.6, 61.32] ($n=490$) for the Dutch samples; and 48.85% [45.7, 51.95] ($n=997$) vs. 56.05% [52.7, 59.44] ($n=826$) for the Swedish samples [Aulchenko, et al. 2008]. Thus, a 1% excess in cases represents a fairly stringent cut-off.

The corresponding nonsuperiority p -values are presented Table 1. Overall, the analysis of the case-control rs10492972 IMSGC datasets supports the absence of association. Some population sample sets provided borderline significance, suggesting that if any association exists, the increased frequency of rs10492972 in the cases is lower than 1%. No heterogeneity can be detected in either IMSGC cases ($p=0.247$) or IMSGC controls ($p=0.754$). When the samples are considered together, the use of a nonsuperiority test excludes rs10492972*C allele as a major risk allele for MS with confidence ($p=1.18 \times 10^{-4}$).

Nonsuperiority methods are also applicable to transmission disequilibrium test (TDT) analyses, which determine linkage in the presence of association. The one-sided test of the equivalence method is slightly different because the observed transmission rate ($\widehat{T}$) is compared with a fixed transmission rate (50%). The hypothesis that the transmission rate for rs10492972*C carriers is greater than 50% by a specified delta becomes the null hypothesis of the test (H_0).

$$H_0: \widehat{T} > T_0^{\text{sup}}; T_0^{\text{sup}} = 0.5 + \Delta\%; \Delta = 2.5\%$$

$$H_1: \widehat{T} < T_0^{\text{sup}}$$

Where $\widehat{T}$ is the observed transmission rate rs10492972*C allele, T_0^{sup} is the nonsuperiority limit of the transmission rate, and n_T is the number of transmissions observed.

$$\frac{\widehat{T} - T_0^{\text{sup}}}{\sqrt{\frac{T_0^{\text{sup}} \times (1 - T_0^{\text{sup}})}{n_T}}} \rightarrow N(0, 1)$$

EQUATION (2)

In Table 2, the nonsuperiority p -values for the equivalence TDT are reported. When using meta-analysis of transmission rates, no heterogeneity can be detected across the IMSGC datasets ($p=0.689$). The three combined datasets show that if any association exists, the transmission distortion of the rs10492972*C allele is lower than 2.5%, providing further support for the absence of association with MS.

DISCUSSION

Altogether, these equivalence-based statistics confirm that the absence of evidence for association between rs10492972*C and MS susceptibility seen in the IMSGC datasets is compelling evidence for the absence of such an association. To explain this controversial result, several interpretations have been proposed by Aulchenko et al. (2008) and IMSGC and Aulchenko et al. In their reply. The data from the isolated Dutch population, which was initially used to propose the association, showed decreased frequency of the risk allele in controls; however, recruitment bias may have occurred. The decreased frequency was confirmed when Aulchenko et al. (2008) and IMSGC (2010) control frequencies were compared in a joint meta-analysis (random-effect $p = 1 \times 10^{-3}$; fixed-effect $p = 1.19 \times 10^{-8}$). In addition, genotyping errors have been proposed as an additional possible explanation for the inconsistency in the allelic frequencies [Aulchenko, et al. 2008]. As noted by Ban et al. [IMSGC, et al. 2010], the small sample of the initial study in the Netherlands (45 cases and 195 controls) was suggestive of a false-positive association, which would not have passed a genome-wide significance threshold. The possibility that this variant is relevant in the Dutch, Swedish, and Canadian samples (the populations initially studied) but not elsewhere in the IMSGC samples seems unlikely. Our own meta-analysis of the available data did not detect population stratification for rs10492972*C in either cases ($p = 0.247$) or controls ($p = 0.754$) (data not shown).

The analysis demonstrates the utility and advantages of equivalence-based methods when assessing negative associations (the so-called “negative results”). Whereas one-sided tests are required for the absence of evidence for previously reported associations, a “two one-sided test” procedure (TOST equivalence test) is needed to demonstrate nonsuperiority and noninferiority at the same time. As shown in equivalence tests for binomial variables [Barker, et al. 2001], several applications of the equivalence tests may help to refine their use in genetic association studies. Equivalence tests alternatively evaluate the local statistical power of studies that use classical difference tests. However, when samples size are low ($n < 1000$), neither nonsignificant difference tests nor significant (with a large-delta hypothesis) equivalence tests are conclusive. Somewhat arbitrary power computations can be avoided by evaluating the noninferiority–nonsuperiority interval of a genetic association. The equivalence methods are complementary to classical difference testing because they evaluate the extent to which biomarkers can be excluded.

When reporting a genetic association based on a cohort of a limited size for the first time, caution is needed regardless of how attractive the underlying biological rationale is. The data gathered for *KIF1B* in MS also underscore the need for very large sample sizes to exclude false-positive results with the appropriate equivalence statistical methods.

Acknowledgments

This work was supported by a grant from the National Institute of Neurological Disorders and Stroke (NS049477). Details of the International Multiple Sclerosis Genetics Consortium (IMSGC) can be found at <http://www.imsgc.org>. The IMSGC is supported by both NIH and NMSS grants.

REFERENCES

- Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ*. 1995; 311(7003):485. [PubMed: 7647644]
- Aulchenko YS, Hoppenbrouwers IA, Ramagopalan SV, Broer L, Jafari N, Hillert J, Link J, Lundstrom W, Greiner E, Dessa Sadovnick A, et al. Genetic variation in the *KIF1B* locus influences susceptibility to multiple sclerosis. *Nat Genet*. 2008; 40(12):1402–3. [PubMed: 18997785]

- Barker L, Rolka H, Rolka D, Brown C. Equivalence Testing for Binomial Random Variables. *The American Statistician*. 2001; 55(4):279–287.
- IMSGC IMSCG. Booth DR, Heard RN, Stewart GJ, Cox M, Scott RJ, Lechner-Scott J, Goris A, Dobosi R, Dubois B, et al. Lack of support for association between the KIF1B rs10492972[C] variant and multiple sclerosis. *Nat Genet*. 2010; 42(6):469–70. author reply 470–1. [PubMed: 20502484]
- Lyons DA, Naylor SG, Scholze A, Talbot WS. Kif1b is essential for mRNA localization in oligodendrocytes and development of myelinated axons. *Nat Genet*. 2009; 41(7):854–8. [PubMed: 19503091]
- Martinelli-Boneschi F, Esposito F, Scalabrini D, Fenoglio C, Rodegher ME, Brambilla P, Colombo B, Ghezzi A, Capra R, Collimedaglia L, et al. Lack of replication of KIF1B gene in an Italian primary progressive multiple sclerosis cohort. *Eur J Neurol*. 2010; 17(5):740–5. [PubMed: 20067515]
- Schuirman DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm*. 1987; 15(6):657–80. [PubMed: 3450848]

TABLE 1

Nonsuperiority tests in the IMSGC case-control datasets

Sample ^a	No. of rs10492972*C carriers (%)		Delta of C-allele carrier frequency (%)	Nonsuperiority <i>p</i> -value
	Cases	Controls		H0: C frequency, cases > controls + 1%
Australia	153 (52.29%)	115 (58.26%)	-5.97	0.05
Belgium	766 (53%)	913 (51.81%)	1.20	0.55
Finland	793 (55.86%)	984 (57.72%)	-1.86	0.04
Italy	814 (51.35%)	618 (53.24%)	-1.88	0.06
Norway	696 (51.72%)	1173 (54.05%)	-2.33	0.02
Sweden	1222 (53.36%)	731 (52.39%)	0.96	0.49
United Kingdom	1369 (52.59%)	1516 (54.16%)	-1.56	0.03
United States	2440 (53.28%)	1768 (54.02%)	-0.74	0.06
Combined	8253 (53.06%)	7818 (54.11%)	-1.05	1.18E-04

^a As found in Ban et al. in IMSGC datasets [IMSGC, et al. 2010]. For each of the samples a nonsuperiority *p*-value is reported that corresponds to the statistical significance of the null hypothesis that the frequency of the rs10492972 *C allele is greater in cases than in controls and differs by at least 1%. The frequencies reported by Aulchenko et al. (2008) were (controls vs. cases): 37.44% [30.6, 44.23] (*n* = 195) vs. 66.67% [52.9, 80.44] (*n* = 45) for an isolated Dutch population; 46.01% [41.3, 50.74] (*n*=426) vs. 56.94% [52.6, 61.32] (*n* = 490) for Dutch samples; 48.85% [45.7, 51.95] (*n* = 997) vs. 56.05% [52.7, 59.44] (*n* = 826) for Swedish samples. Meta-analysis of the frequencies suggest heterogeneity in cases (*p* = 0.01) for data sets reported by Aulchenko et al. [2008]. No heterogeneity can be detected in either cases (*p* = 0.247) or controls (*p* = 0.754) in IMSGC data sets. Aulchenko et al. [2008] and IMSGC control datasets differed significantly (random-effect *p* = 1×10^{-3} , fixed-effect *p* = 1.97×10^{-8}). In the combined sample analysis, the difference in frequency between cases and controls is smaller than 1%, with a statistical significance measured by the nonsuperiority *p*-value of 1.18×10^{-4} .

TABLE 2

Equivalence-based statistics in the three IMSGC datasets

Sample ^a	rs10492972*C transmission to proband		Observed transmission rate % (OR)	Nonsuperiority p-value
	Transmitted	Untransmitted		H0: Transmitted > 52.5%
Australia	140	147	48.78 (0.95)	0.1037
France	249	256	49.31 (0.97)	0.0756
United Kingdom	483	461	51.17 (1.05)	0.206
All	872	864	50.23 (1.01)	0.0293

^a As found in Ban et al. in IMSGC datasets [IMSGC, et al. 2010]. The observed transmission rates are used to assess the statistical significance of the null hypothesis corresponding to an over-transmission of at least 52.5%. A 2.5% over-transmission of the rs10492972*C allele corresponds to an odds ratio of 1.105 [$OR^{sup} = (T_0^{sup}) / (1 - T_0^{sup})$], which can be considered a lower boundary estimate from Aulchenko et al. [2008] data. The standard nonsuperiority one-sided test equivalence methods were performed using STATA (Equip) and confirmed in SAS.