



Published in final edited form as:

*Genet Epidemiol.* 2010 September ; 34(6): 569–574. doi:10.1002/gepi.20513.

## A Bayesian approach to genetic association studies with family-based designs

Melissa G. Naylor<sup>1</sup>, Scott T. Weiss<sup>2</sup>, and Christoph Lange<sup>1,2</sup>

<sup>1</sup>Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA

<sup>2</sup>Channing Laboratory, Brigham and Women's Hospital, Boston, Massachusetts, USA

### Abstract

For genomewide association studies with family-based designs, we propose a Bayesian approach. We show that standard TDT/FBAT statistics can naturally be implemented in a Bayesian framework. We construct a Bayes factor conditional on the offspring phenotype and parental genotype data and then use the data we conditioned on to inform the prior odds for each marker. In the construction of the prior odds, the evidence for association for each single marker is obtained at the population-level by estimating the genetic effect size in the conditional mean model. Since such genetic effect size estimates are statistically independent of the effect size estimation within the families, the actual data set can inform the construction of the prior odds without any statistical penalty. In contrast to Bayesian approaches that have recently been proposed for genomewide association studies, our approach does not require assumptions about the genetic effect size; this makes the proposed method entirely data-driven. The power of the approach was assessed through simulation. We then applied the approach to a genomewide association scan to search for associations between single nucleotide polymorphisms and body mass index in the Childhood Asthma Management Program data.

### Keywords

family-based association tests; Bayes factors; complex traits

## 1 Introduction

Genetic association studies can be dichotomized into those that have population-based designs and those that have family-based designs. Population-based studies test for association between a phenotype and a genotype in a sample of unrelated individuals. These studies are susceptible to population stratification, i.e., systematic differences in allele frequencies between subpopulations in a population not due to a causal association with the phenotype of interest. Family-based studies test for association between a phenotype and a genotype in a sample of related individuals by assessing whether individuals with a given phenotype have a higher transmission ratio than would be expected by chance given their parents' (or other family members') genotypes and Mendel's laws. It is this ability to compare a person's genotype to his expected genotype based on Mendel's laws that makes family-based tests robust to population stratification. Family-based studies can also assess the evidence of association by examining allele frequencies across families in a somewhat

similar fashion to what is done in population-based studies (although this evidence is not inherently robust to population stratification).

The use of the two independent sources of information available in family data is what makes the family-based association test (FBAT) screening method proposed by Steen et al. (2005) so powerful. If we let  $x$  denote the offspring genotypes,  $y$  denote the offspring phenotypes, and  $P$  denote the parents' genotypes, the following relationship holds:

$$P(x, y, P) = P(x|y, P) P(y, P).$$

$P(x|y, P)$  represents the information used in the FBAT statistic; that is, the FBAT statistic conditions on the offspring phenotypes,  $y$ , and parents' genotypes,  $P$ , and only the offspring genotypes,  $x$ , are considered random variables.  $P(y, P)$  represents the information used in Steen et al. (2005)'s screening approach, which first determines which SNPs will have the highest power to be detected if a true association exists in the given study and then applies FBAT only to those SNPs. In determining which SNPs have the highest power, only the offspring phenotypes,  $y$ , and parental genotypes,  $P$  are used. By applying the FBAT to only the SNPs in which a true association would most likely be detectable, the number of comparisons that must be adjusted for is reduced. The statistical validity of the screening method is due to the relationship in the formula above.

We propose a novel approach for family-based genetic association studies that also capitalizes on this relationship. However, instead of using one portion of the family data (i.e.,  $x|y, P$ ) to create a statistical test and the other independent portion (i.e.,  $y, P$ ) to screen SNPs, we take a Bayesian approach: we use one portion to calculate a Bayes factor and the other to calculate the prior odds of the null hypothesis. Our method capitalizes on both the between and within family information in a more intuitive and flexible manner than frequentist approaches. We construct a Bayes factor conditional on parental genotypes and offspring phenotypes and then use the information conditioned on to inform the prior odds of association for each marker. In this way we are able to combine evidence from the over-transmission of alleles within families with evidence based on the allele frequencies across the families into a posterior odds of no association.

## 2 Methods

An intuitive measure of the evidence for genetic association is the posterior odds of no association:  $\frac{P(\text{no association}|\text{data})}{P(\text{association}|\text{data})}$ . We will denote the hypothesis that a given marker is not associated with a phenotype by  $H_0$ . The posterior odds of  $H_0$  can be written as

$$\text{posterior odds of } H_0 = \text{Bayes factor} \times \text{prior odds of } H_0.$$

When there are only two hypotheses under consideration, we can rewrite the above equation as

$$\frac{P(H_0|\text{data})}{P(H_1|\text{data})} = \frac{P(\text{data}|H_0)}{P(\text{data}|H_1)} \times \frac{P(H_0)}{P(H_1)}.$$

We assume throughout that the alternative hypothesis of interest,  $H_1$ , is an additive genetic effect; however the Bayes factor and posterior odds described below can easily be modified to test other hypotheses.

If we now consider family-based association studies with complete genotype data for trios (two parents and an offspring) and a continuous phenotype for the offspring, then the data consists of parental genotypes,  $P$ , offspring genotypes,  $x$ , and offspring phenotypes,  $y$ . In the derivation of our Bayes factor we borrow from the FBAT statistic the concept of conditioning on the parental genotypes,  $P$ , and offspring phenotypes,  $y$ , so that the formula for the posterior odds of  $H_0$  becomes

$$\frac{P(H_0|x, y, P)}{P(H_1|x, y, P)} = \frac{P(x|y, P, H_0)}{P(x|y, P, H_1)} \times \frac{P(H_0|y, P)}{P(H_1|y, P)}.$$

By conditioning on  $y$  and  $P$  when we assess the weight of evidence for over-transmission of alleles in the Bayes factor, we are able to use the between-family information available in  $y$  and  $P$  to gain further evidence for or against association in the conditional prior odds. Two main goals of genomewide association scans can be addressed by a measure of the posterior odds of  $H_0$ , no association: ranking markers to decide which to follow up and determining which findings to report as noteworthy, or significant. In the following three subsections we discuss the conditional Bayes factor, the conditional prior odds, and how to use these measures to infer the noteworthiness of individual markers.

### 2.1 A Bayes factor for $x$ conditional on $y$ and $P$

We assume our data consists of independent trios (two parents and an offspring in each family) and  $H_0$  and  $H_1$  correspond to the hypotheses of no genetic effect and an additive genetic effect, respectively. We assume the  $i$ th individual's phenotype,  $y_i$ , is independently distributed as  $y_i \sim N(\mu + ax_i, \sigma^2)$ , where  $\mu$  is the overall mean of the quantitative trait and  $a$  is the genetic effect. Since we assume an additive genetic effect, we code the individual's genotype,  $x_i$ , as the number of copies of an allele, i.e., 0, 1, or 2. The Bayes factor using the offspring genotypes,  $x$ , conditional on the offspring phenotypes,  $y$ , and parental genotypes,  $P$  is

$$BF_{x|y,P} = \frac{P(x|y,P,H_0)}{P(x|y,P,H_1)} = \left[ \int \int \int \frac{P(y | x, a, \mu, \sigma) P(a, \mu, \sigma | y, P)}{\prod_{i=1}^n \left[ \sum_{\tilde{x}_i} P(y_i | \tilde{x}_i, a, \mu, \sigma) P(\tilde{x}_i | P_i) \right]} da d\mu d\sigma \right]^{-1}$$

where the sum is over all possible offspring genotypes that could have occurred in family  $i$  given the parental genotypes,  $P_i$ . (See Appendix 6.1 for derivation.)

A fully Bayesian approach requires specification of the conditional prior distribution,  $P(a, \mu, \sigma | y, P)$ . However, the multiple integration needed to calculate this Bayes factor may be computationally prohibitive. The integrals can be evaluated at the maximum likelihood estimates derived by regressing  $y$  on  $x$ , so that the Bayes factor is approximated as

$$BF_{x|y,P} = \frac{P(x|y, P, H_0)}{P(x|y, P, H_1)} \approx \frac{\prod_{i=1}^n \left[ \sum_{\tilde{x}_i} P(y_i | \tilde{x}_i, \hat{a}, \hat{\mu}, \hat{\sigma}) P(\tilde{x}_i | P_i) \right]}{P(y | x, \hat{a}, \hat{\mu}, \hat{\sigma})}$$

where  $\hat{a}$ ,  $\hat{\mu}$ , and  $\hat{\sigma}$  are the maximum likelihood estimates.

We also assessed the power of fully specifying the priors based on only the parents' genotypes and offsprings' phenotypes, but found this approach to be less powerful and more computationally intensive. In this case, we derived the conditional prior for  $a$ ,  $\mu$  and  $\sigma$  analytically by fitting the conditional mean model (Lange et al., 2003),

$$y_i \sim N(\mu + aE(x_i|P_i), \sigma^2)$$

and specifying an unconditional prior on  $a$ ,  $\mu$  and  $\sigma$ , e.g.,

$$P(\mu, a, \sigma^2) \propto \frac{1}{\sigma^2}.$$

In the conditional mean model,  $E(x_i|P_i)$  is determined using Mendel's Laws. This allows us to gain additional evidence for or against association from  $y$  and  $P$  without using the true offspring genotypes,  $x$ . Since substituting in the maximum likelihood estimates for  $a$ ,  $\mu$  and  $\sigma$  was more powerful than this approach, we only discuss the former method in the remainder of the manuscript. In the next section we describe how we used the conditional mean model to estimate the conditional prior odds.

## 2.2 Deriving the conditional prior odds

The conditional prior odds are used to weigh the evidence for association using only the offspring phenotypes,  $y$ , and the parental genotypes,  $P$  (i.e., the offspring genotypes,  $x$ , are not used).

$$PO_{y,P} = \frac{P(H_0|y, P)}{P(H_1|y, P)}$$

To get an estimate of the genetic effect size,  $a$ , without using the observed offspring genotypes we use the conditional mean model (Lange et al., 2003):

$$E(y_i | P_i) = \mu + aE(x_i | P_i)$$

The conditional mean model uses Mendel's Laws to determine the expected offspring genotypes given only the parents' genotypes. This enables estimation of the genetic effect size,  $a$ , using only the parents' genotypes and the offsprings' phenotypes. We can then use the estimated effect size,  $\tilde{a}$ , and its standard error,  $\tilde{\sigma}_a$ , to summarize the evidence of association contained in  $y$  and  $P$ . One way to do this is to consider approximating the distribution of the genetic effect size,  $a$ , by  $N(|\tilde{a}|, \tilde{\sigma}_a^2)$ . Then we can think of  $P(a < 0)$  as a proxy for  $P(H_0|y, P)$  and calculate the prior odds as

$$PO_{y,P} = \frac{P(H_0|y, P)}{P(H_1|y, P)} \approx \frac{P(a < 0)}{1 - P(a < 0)}$$

We cannot use  $P(a = 0)$  since this quantity is always equal to zero. Instead, we use the probability  $a$  is less than or equal to zero if  $\tilde{a} > 0$  or the probability  $a$  is greater than or equal to zero if  $\tilde{a} < 0$ ; this is equivalent to taking  $P(a > 0)$  when we assume that  $a$  is normally distributed with mean equal to the absolute value of  $\tilde{a}$ . By taking the absolute value of  $a$ , we

are not penalized if it is wrong due to population stratification. This is an admittedly ad hoc way to derive the prior odds, but it does utilize the information in  $y$  and  $P$ . Furthermore, multiplying the conditional Bayes factor by this prior odds yields greater power in simulation (for an individual test with threshold based on null simulations) than the Bayes factor alone (Supplementary Table VI).

### 2.3 Using the posterior odds to make inference

SNPs can easily be ranked according to the posterior odds. Unlike some recently proposed Bayesian methods (Wakefield, 2008a; Consortium, 2007), ranking SNPs by our Bayes factor alone will not yield the same rankings as the posterior odds. This is because the conditional prior odds are not constant for all markers: we use the parental genotypes and offspring phenotypes to inform the prior odds for each marker individually. With our method, it is more powerful to rank the SNPs by posterior odds since doing so uses more of the information available in the data.

One caveat to including the prior odds is the increased susceptibility to detecting association due to population stratification. The conditional Bayes factor may be affected by population stratification if the maximum likelihood estimates are substituted for  $a$ ,  $\mu$ , and  $\sigma$ , but it primarily weighs the evidence for association in the within-family information. (The transmission of alleles within a family is completely robust to population stratification.) On the other hand, the prior odds rely entirely on population differences in allele frequencies and are therefore as susceptible as any unadjusted population-based method. Because of this, we recommend that SNPs be followed up only if the posterior odds of the null hypothesis is less than the prior odds of the null hypothesis, i.e., only if the conditional Bayes factor is less than one.

Inferring which SNPs to report as noteworthy or significant is a more challenging matter than just ranking them. One issue that must be taken into account is that  $P(H_0|y, P) = 0.5$  which is not likely to be a realistic reflection of one's true beliefs in a genomewide association scan; i.e., it is not likely that at least half of the markers will be associated with a given phenotype. Also, substituting the maximum likelihood estimates for the priors in the Bayes factor is likely to favor the alternative hypothesis since the Bayes factor is evaluated at only the best estimate under the alternative. It is relatively straightforward to control the type one error for a single test: use the sample size, allele frequencies, and variance of the phenotype to simulate the data under the null hypothesis of no association and choose as a threshold the 95th percentile of the simulated posterior odds. Alternatively, one could permute the data to obtain an empirical  $p$  value. Bayesian decision theory has been suggested as a tool for controlling the false discovery rate when using Bayesian methods to assess many markers individually in genomewide association scans (Wakefield, 2008b). Wakefield showed that the expected posterior cost is minimized when the posterior odds is less than the cost of a false nondiscovery,  $C_{FND}$ , divided by the cost of a false discovery,  $C_{FD}$ : i.e., when

$$\frac{P(H_0|data)}{P(H_1|data)} < \frac{C_{FND}}{C_{FD}}.$$

This approach can be easily implemented as long as one is careful to weight the conditional prior odds to truly reflect one's beliefs.

## 3 Results

### 3.1 Simulations

Simulations were performed to assess power. We simulated samples of 100, 500, or 1000 trios, each consisting of two parents and an offspring. All probands and parents were assumed to be accurately genotyped. Parental genotypes were drawn from a binomial distribution with the probability of success equal to the allele frequency. Offspring genotypes were generated based on Mendelian transmission. A continuous trait was generated with a standard normal distribution. Bayes factors were calculated for varying combinations of locus-specific heritability of the trait and allele frequency (0.05, 0.1 or 0.2). For each set of conditions, 10,000 iterations were run. All simulations were performed in R version 2.6.1.

Summary statistics for conditional Bayes factors and posterior odds (i.e., conditional Bayes factor times prior odds) simulated assuming no genetic effect are shown in Supplementary Tables II and III. The null simulations were used to determine a threshold for each allele frequency that preserves a type one error rate of 0.05. Thresholds were calculated as the 95th percentile of the null distribution. We examined the distributions of Bayes factors and posterior odds generated under varying allele frequencies and heritabilities and used the threshold values to determine power (Supplementary Tables IV and V). Both the conditional Bayes factor alone and the resulting posterior odds were more powerful than the FBAT statistic (Figure 1 and Supplementary Table VI). Figure 1 also shows the power when maximum likelihood estimates substituted into the Bayes factor are obtained by regressing  $y$  on  $x$  (as opposed to the conditional mean model in which  $y$  is regressed on  $E(x|P)$ ). This did not yield as great a power as using the conditional mean model maximum likelihood estimates. As one would expect, power increased as the amount of heritability explained by the SNP of interest increased. The increase in power was observed across varying heritabilities and allele frequencies. While in some scenarios the improvement was substantial, in others it was slight.

### 3.2 Application to CAMP

We applied these methods to the Childhood Asthma Management Program dataset (199, 1999). We assessed the evidence for association between roughly 500,000 SNPs and baseline body mass index z-scores (BMIZ) (Kuczmarski et al., 2000). The analysis included 389 genotyped trios (two parents, one offspring). All SNPs were individually tested for association. We assumed an additive genetic model. No further adjustments were used since neither age nor sex was significantly associated with BMIZ. Empirical p values for the posterior odds were obtained by permuting the phenotypes repeatedly. Table I shows the twenty SNPs with Bayes factor less than one that had the highest posterior odds of the alternative hypothesis. Since the Bayes factor is

$$\frac{P(data|H_0)}{P(data|H_1)},$$

a Bayes factor of less than one indicates  $P(data|H_0) < P(data|H_1)$ . By only considering the posterior odds for SNPs that have a Bayes factor of less than one, we avoid calling SNPs significant when there is no evidence of association within families. A low posterior odds of no association with a Bayes factor greater than one is likely to result from population stratification.

## 4 Discussion

The use of Bayes factors has been advocated as a powerful way to gauge the evidence for genetic association in population-based genomewide association scans (Wakefield, 2008a; Consortium, 2007). Bayes factors are suitable for both ranking SNPs and for calibrating inferences to decide which SNPs to consider noteworthy (Wakefield, 2008b). Wakefield (2008a) proposed an asymptotic Bayes factor approach for population-based studies that does not require integration and has suggested possible priors that can be used in conjunction with it. As Wakefield mentions, there are two aspects of Bayes factors that have prevented their more widespread use: the multidimensional integrals required to calculate them can be computationally unattractive for large genomewide association scans and specification of priors leads to much controversy. Our approach makes no assumptions about the distribution of alleles or the genetic effect size. Furthermore, our method circumvents both the need to compute multidimensional integrals and the need to subjectively specify a prior.

The Bayes factor approach to family-based association testing presented here is a flexible and intuitive way to utilize both the within- and between-family information. Because the Bayes factor conditions on the parental genotypes and offspring phenotypes, we can use this portion of the data to inform the prior odds. The Bayes factor uses within family information to weigh the evidence for association at a given SNP. A main advantage of using a Bayesian framework is the flexibility it provides. The likelihood can be chosen to model any number of study designs. Additional information regarding individual SNPs can easily be incorporated into the prior as well. Furthermore, this approach does not require distributional assumptions about allele frequencies.

There are many ways this method can be extended and further characterized. The Bayes factor approach could be adapted for use with extended pedigrees, missing data, or gene-environment interactions. Biological information pertaining to individual markers could easily be used to weight the prior odds. To potentially increase the power of our novel method, information from multiple SNPs could be incorporated, or one might consider doing a joint model of all SNPs to eliminate multiple comparisons.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We especially thank Chris Paciorek for his assistance with this project. This work has been supported by the National Institutes of Health through grants T32 MH17119 and T32 HL07427-26. We thank all subjects for their ongoing participation in the CAMP study. We acknowledge the CAMP investigators and research team, supported by NHLBI, for collecting CAMP Genetic Ancillary Study data. All work on data collected from the CAMP Genetic Ancillary Study was conducted at the Channing Laboratory of the Brigham and Women's Hospital under appropriate CAMP policies and human subject's protections. CAMP is supported by U01 HL075419, U01 HL65899, P01 HL083069, and T32 HL07427 from the National Heart, Lung and Blood Institute, National Institutes of Health.

## 6 Appendix

### 6.1 Deriving the Bayes factor for $x$ conditional on $y$ and $P$

The Bayes factor for the offspring genotypes,  $x$ , conditional on the offspring phenotypes,  $y$ , and parental genotypes,  $P$ , is defined as

$$BF_{x|y,P} = \frac{P(x|y, P, H_0)}{P(x|y, P, H_1)}$$

where  $H_0$  and  $H_1$  correspond to no genetic effect and an additive genetic effect, respectively. We assume the phenotypes,  $y$ , are independently distributed as  $y_i \sim \mathcal{N}(\mu + ax_i, \sigma^2)$ . Integrating out the genetic effect,  $a$ , yields

$$BF_{x|y,P} = \frac{P(x|y, P, H_0, a=0)}{\int P(x|y, P, H_1, a) P(a|y, P, H_1) da}$$

Henceforth it is assumed that we are conditioning on  $H_1$  and  $H_0$  and we drop the notation accordingly. Like the genetic effect size,  $a$ , the mean and variance of  $y$  is unknown, so we integrate over these parameters as well.

$$\begin{aligned} BF_{x|y,P} &= \frac{P(x|y, P, a=0)}{\iint \iint P(x|y, P, a, \mu, \sigma) P(a, \mu, \sigma | y, P) dad\mu d\sigma} \\ &= \frac{P(x|P)}{\iint \iint P(x|y, P, a, \mu, \sigma) P(a, \mu, \sigma | y, P) dad\mu d\sigma} \end{aligned}$$

because offspring genotype does not depend on the phenotype when the effect size is zero. Applying the definition of conditional probability,

$$BF_{x|y,P} = P(x|P) \left[ \iint \iint \frac{P(x,y,P|a,\mu,\sigma)}{P(y,P|a,\mu,\sigma)} P(a, \mu, \sigma | y, P) dad\mu d\sigma \right]^{-1}$$

Since the trios are assumed to be independent, it follows that

$$\begin{aligned} BF_{x|y,P} &= P(x|P) \left[ \iint \iint \frac{P(x,y,P|a,\mu,\sigma)}{\prod_{i=1}^n P(y_i, P_i | a,\mu,\sigma)} P(a, \mu, \sigma | y, P) dad\mu d\sigma \right]^{-1} \\ &= P(x|P) \left[ \iint \iint \frac{P(x,y,P|a,\mu,\sigma)}{\prod_{i=1}^n \left[ \sum_{\tilde{x}_i} P(\tilde{x}_i, y_i, P_i | a,\mu,\sigma) \right]} P(a, \mu, \sigma | y, P) dad\mu d\sigma \right]^{-1}, \end{aligned}$$

where the sum is over all the possible offspring genotypes that could have occurred in family  $i$  given the parental genotypes,  $P_i$ . Then,

$$BF_{x|y,P} = P(x|P) \times \left[ \iint \iint \frac{P(y|x,P,a,\mu,\sigma) P(x|P,a,\mu,\sigma) P(P|a,\mu,\sigma) P(a,\mu,\sigma|y,P)}{\prod_{i=1}^n \left[ \sum_{\tilde{x}_i} P(y_i|\tilde{x}_i, P_i, a,\mu,\sigma) P(\tilde{x}_i|P_i, a,\mu,\sigma) P(P_i|a,\mu,\sigma) \right]} dad\mu d\sigma \right]^{-1}$$

In the absence of phenotype information,  $y$ , offspring genotypes depend only on parental genotypes,  $P$ . Thus,



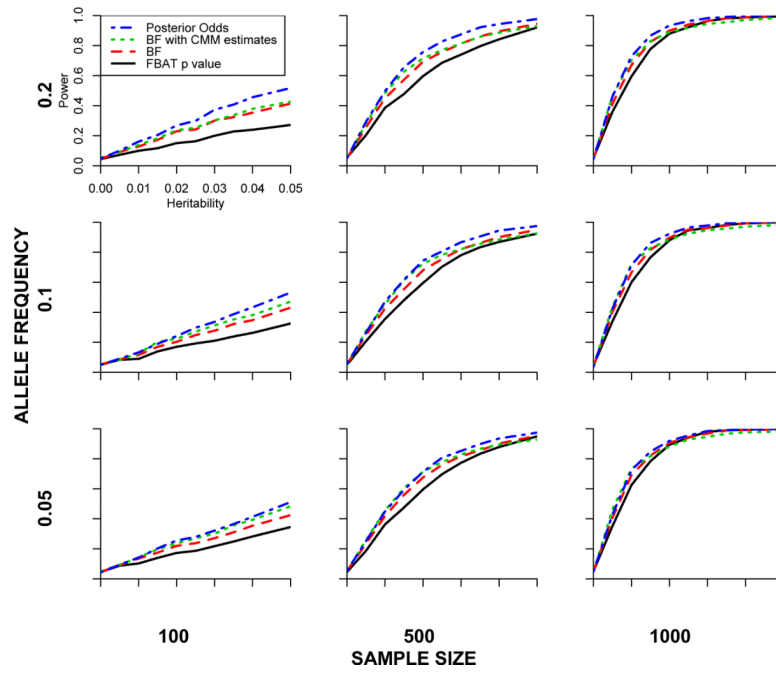
$$BF_{x|y,P} = P(x|P) \times \left[ \int \int \int \frac{P(y|x,P,a,\mu,\sigma)P(x|P)P(P|a,\mu,\sigma)P(a,\mu,\sigma|y,P)}{P(P|a,\mu,\sigma) \prod_{i=1}^n \left[ \sum_{\tilde{x}_i} P(y_i|\tilde{x}_i,P_i,a,\mu,\sigma)P(\tilde{x}_i|P_i,a,\mu,\sigma) \right]} da d\mu d\sigma \right]^{-1}$$

Since we assume the offspring phenotypes,  $y$ , depend on the parental genotypes,  $P$ , only through the offspring genotypes,  $x$ , it follows that

$$BF_{x|y,P} = \frac{P(x|y,P,H_0)}{P(x|y,P,H_1)} = \left[ \int \int \int \frac{P(y|x,a,\mu,\sigma)P(a,\mu,\sigma|y,P)}{\prod_{i=1}^n \left[ \sum_{\tilde{x}_i} P(y_i|\tilde{x}_i,a,\mu,\sigma)P(\tilde{x}_i|P_i) \right]} da d\mu d\sigma \right]^{-1}$$

## References

- The childhood asthma management program (camp): design, rationale, and methods. childhood asthma management program research group. *Control Clin Trials*. 1999; 20:91–120. [PubMed: 10027502]
- Consortium WTCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447:661–678. [PubMed: 17554300]
- Kuczumski RJ, Ogden CL, Grummer-Strawn LM, Flegal KM, Guo SS, Wei R, Mei Z, Curtin LR, Roche AF, Johnson CL. Cdc growth charts: United states. *Adv Data*. 2000; 314:1–27. [PubMed: 11183293]
- Lange C, Lyon H, DeMeo D, Raby B, Silverman EK, Weiss ST. A new powerful non-parametric two-stage approach for testing multiple phenotypes in family-based association studies. *Hum Hered*. 2003; 56:10–17. doi:10.1159/000073728. [PubMed: 14614234]
- Steen KV, McQueen MB, Herbert A, Raby B, Lyon H, Demeo DL, Murphy A, Su J, Datta S, Rosenow C, Christman M, Silverman EK, Laird NM, Weiss ST, Lange C. Genomic screening and replication using the same data set in family-based association testing. *Nat Genet*. 2005; 37:683–691. doi:10.1038/ng1582. [PubMed: 15937480]
- Wakefield J. Bayes factors for genome-wide association studies: comparison with p-values. *Genet Epidemiol*. 2008a doi:10.1002/gepi.20359.
- Wakefield J. Reporting and interpretation in genome-wide association studies. *Int J Epidemiol*. 2008b; 37:641–653. doi:10.1093/ije/dym257. [PubMed: 18270206]



**Figure 1.** Power of the posterior odds (blue) compared with the power of the Bayes factor using maximum likelihood estimates from regressing  $y$  on  $E(x)$  (green), the power of the Bayes factor using maximum likelihood estimates from regressing  $y$  on  $x$  (red), and the power of  $p$  values from the family based association test (black). The posterior odds (blue) is equivalent to the Bayes factor (red) multiplied by the prior odds.

**Table I**

Twenty SNPs with lowest empirical pvalues based on posterior odds.

SNP	Bayes Factor	Posterior Odds empirical pvalue	FBAT pvalue
rs16970570	0.15	0.000003	0.063
rs10428219	0.46	0.000005	0.080
rs7869694	0.14	0.000006	0.062
rs1429862	0.12	0.000009	0.006
rs9374571	0.12	0.000010	0.029
rs3793276	0.11	0.000022	0.022
rs1863159	0.11	0.000031	0.023
rs9307306	0.16	0.000038	0.077
rs10750642	0.10	0.000038	0.024
rs6536309	0.16	0.000054	0.040
rs2195731	0.19	0.000060	0.160
rs6836146	0.20	0.000063	0.154
rs7813810	0.01	0.000072	0.001
rs17075980	0.12	0.000084	0.037
rs9645521	0.11	0.000088	0.019
rs349403	0.12	0.000095	0.037
rs7990570	0.14	0.000096	0.047
rs867477	0.15	0.000097	0.100
rs4720952	0.12	0.000102	0.038
rs242967	0.17	0.000109	0.029