

# PBAT: ANALYTICAL POWER CALCULATIONS FOR FAMILY-BASED ASSOCIATION TESTS

CHRISTOPH LANGE AND NAN M. LAIRD  
MARCH 5, 2002

## 1. INTRODUCTION

*PBAT* is an interactive software package for the design of genetic family-based association studies. It provides analytical power calculations for family-based association tests (Laird et al (2001)). The computations are based on the approach by Lange & Laird (2001a,b). The power can be computed for a variety of scenarios:

- dichotomous and continuous traits
- missing parental information
- multiple offspring per family
- combinations of different family-types
- different genetic models
- different ascertainment conditions for the first and second proband
- marker and disease locus are not identical
- combination of different family-types and different ascertainment conditions

Each power calculation can be verified by a Monte-Carlo simulation. Further, all output displayed on the screen is written to a log-file called "ptdtlog.txt". The designs used for the power calculations can also be saved to a file.

*PBAT* is available for Windows platforms and Unix. Versions for Linux and MAC are in preparation.

FIGURE 1. Welcome window

```
*****
*
*   * * *   * * *   *   *****
*   * *   * * *   * * *   *
*   * * *   * * *   * * *   *
*   *   *   *   *   *   *   *
*   *   *   *   *   *   *   *
*   *   *   * * * *   *   *
*
*   Christoph Lange, Dec 2001, v0.1
*   Department of Biostatistics
*   Harvard School of Public Health
*
*****
Press RETURN to clear screen
```

## 2. GETTING STARTED: CHOICE BETWEEN DICHOTOMOUS AND CONTINUOUS TRAITS

After *PBAT* is started a welcome window will appear on the screen and remain there for a few seconds (Figure 1). The welcome window will be followed by a menu that allows the user to choose the trait type (Figure 2). The options are dichotomous traits and continuous traits. When continuous traits are selected the power will be computed under the assumption that the traits are normally distributed.

FIGURE 2. Choice between dichotomous and continuous traits

```
Power calculations for family-based association test
=====
<1>  binary trait & multiple family-types
<2>  continuous trait & multiple family-types

<0>  Exit
```

## 3. POWER CALCULATIONS FOR DICHOTOMOUS TRAITS

When power calculations for dichotomous traits have been selected the window shown in Figure 3 will appear on the screen. It allows the user to specify

- (1) Design: Family-types  
The user can specify the family-types, the ascertainment condition and the number of families sampled.
- (2) Genetic model:  
All parameters of the underlying genetic model are specified, eg frequency of the disease gene and the marker gene, the disease penetrance functions, etc
- (3) Statistical parameters:  
In this menu the significance level and the offset used in the test to weight affected and unaffected offspring can be defined
- (3) Power calculation:  
Based on the parameter choices the power is computed either by numerical integration, by approximation or by Monte-Carlo simulation.

The menu shown in Figure 3 is the central window for power calculations for dichotomous traits. The program will return to this menu after each submenu

point is finished. For example, when the power has been computed for a specified design and the user wants to see how the design affects the power one can go back to menu point 1, change the design and then choose directly menu point 4 to compute the power for the new design. The other parameters defined in menu point 2 and 3 need not to be specified again.

FIGURE 3.

```

Power calculations for family-based association test (for binary traits)
=====
(1)  Design: Family-types, sample sizes and ascertainment condition
(2)  Genetic model
(3)  Statistical parameters
(4)  Power computation

(0)  Exit
    
```

**3.1. Design: Family-types, sample sizes and ascertainment condition.**

Here we describe how to define the design of the study. When no design has been specified yet and submenu 1 in Figure (3) is selected the window shown in Figure (4) will be displayed. The user can then specify a design by adding family types, save the current design or load a previous design. Let’s define a design now (Figure 5). First, the user has to specify the total number of offspring per family, the number of missing parents, number of families sampled from the family-type and whether the additional offspring have been phenotyped. The value shown at the left from ”?” in each line is the current value of the parameter. If one wants to keep this value, the RETURN-key has to be pressed. Otherwise the new value has to be entered.

Here all offspring other than the first proband are called ”additional” offspring (including the second proband). After that the ascertainment condition for the first and second proband have to be specified:

- ”0=”: the proband is unaffected
- ”1=”: the proband is affected
- ”2=”; NA = no ascertainment condition, ie the proband’s affection status is given by the underlying genetic model

The specified design is then displayed in the next window (Figure 6). The user has now the option to add further family types to the design by pressing the ”1”-key or to continue with any other key. Note that the number of different family types has no influence on the computation time of the power.

FIGURE 4

```

Family design:
=====

(1) Change family design
(2) Load family design
(3) Save family design
or Press any other key to clear screen

```

FIGURE 5

```

Family design:
=====

Change family-type:
=====

Number of offspring per family:  0 ? 3
Number of missing parents(0,1,2): 0 ? 2
Number of families:              0 ? 200
Additional offspring phenotyped?
1=yes 0=no                       1 ? 1

Ascertainment condition for proband 1:
unaffected=0, affected=1, NA=2    0 ? 1
Ascertainment condition for proband 2:
unaffected=0, affected=1, NA=2    0 ? 2

```

3.2. **Genetic model.** The genetic model used for the power calculation can be specified by (Figure 7):

- the allele frequency of the disease gene and the penetrances for three genotypes
- the mode of inheritance, the allele frequency of the disease gene and the attributable fraction.

Note that allele "A" denotes the disease gene. For example, when we define a genetic model based on the penetrance values and allele frequency, the population prevalence of the disease and the attributable fraction of the disease are also computed by PBAT (Figure 7).

Further, the user has to decide whether the disease locus and the marker locus are identical (Figure 7). When the marker locus and the disease locus are different, the allele frequency of the marker gene has to be specified (Figure 8). The linkage disequilibrium between the marker and the disease locus is defined by the probability of the disease allele conditional on the marker allele (Figure 8). The potential range of this probability computed based on the marker allele frequency and the disease allele frequency is also shown on the screen (Figure 8).

FIGURE 6. Design

```
Family design:
=====
(*) Ascert! proband 1: affected   proband 2: NA
    Additional offspring phenotyped
    # offspring: 3   # missing parents: 2   # families: 200

(1) Change family design
(2) Load family design
(3) Save family design
or Press any other key to clear screen
```

FIGURE 7. Specification of the genetic model:

```

Genetic model:
=====
Model: Add=0 Multi=1 Recessive=2 Dom=3:           0
Allele frequency of the disease gene:             0
Population prevalence of the disease:             0.3
Genetic attributable fraction of the gene:        0.1
Penetrance for AA:                               0.8
Penetrance for AB:                               0.5
Penetrance for BB:                               0.3

<0> Return to main menu
<1> Specify genetic model based on MOI, p, K & AF
<2> Specify genetic model based on penetrance values & allele frequency

```

FIGURE 8. Specification of the genetic model: Disease locus and marker locus are not identical

```

Genetic model:
=====
Allele frequency of the disease gene: 0 ? .2
Penetrance for AA:                   0.8 ?
Penetrance for AB:                   0.5 ?
Penetrance for BB:                   0.3 ?

Population prevalence of the disease:  0.384
Genetic attributable fraction of the gene: 0.219

Disease locus = marker locus? Yes=1 No=0  0 ? 0
Allele frequency of the marker gene:     0 ? .2

P(Disease allele A | Marker allele A) in [0;1]:
1 ? .9

```

**3.3. Statistical parameters.** In this sub-menu the user has to define the statistical parameters, ie the significance level and the offset (Figure 9). The offset describes is weighing of affected and unaffected in the test statistic. An offset choice of 0 means that only affected offspring are included in the computation of the FBAT statistic, while unaffected are ignored. When the offset is 1 the test statistic is computed based on unaffected offspring, but not on affected. A good rule of thumb is to chose to the offset close to the disease prevalence (Whittaker & Lewis (1998)).

FIGURE 9. Statistical parameters: significance level and offset

```
Statistical parameters:
=====
Significance level: 0.01 ? .05
Offset             : 0.3 ? .3

Press RETURN to clear screen
-
```

**3.4. Power computation.** In this submenu the power of the FBAT statistic is computed based on the parameters defined by the user in submenus 1-3. The user has the choice between (Figure 10)

- "1": MCMC-integration: not yet implemented
- "2": Numerical integration:  
The analytical power of FBAT is computed based on numerical integration with a precision of 0.01. This routine can take between 1-60 seconds.
- "3": Approximation  
This routine computes the analytical power of FBATs based on a second-order Taylor expansion. The precision of this method is good for sample sizes of at least 100 families. This method is the fastest approach (computation time: 1 second or less).
- "4": Simulation  
The power of FBAT is assessed by simulation experiments. The power is estimated based on 100,000 Monte-Carlo simulation experiments. Such a Monte-Carlo simulation can take a few minutes.

Figure (11) shows the result of the power computations. First, the family design is given, i.e. which family-types have been used, which ascertainment condition has been applied, how many families of this type have been sampled, etc. Then all genetic and statistical parameters are displayed. Finally, the power and the computation time are given. All results are written to the log-file "ptdt-log.txt".

FIGURE 10. Power computation: Available methods

```
Computation method :
=====
2=numerical integration, 3=approximation, 4=simulation:
```

FIGURE 11. Power computation: Results

```
Family design:
=====
(*) Ascertainment: proband 1: affected   proband 2: NA
    Additional offspring phenotyped
    # offspring: 3   # missing parents: 2   # families: 200

other parameters:
=====
Allele frequency of the disease gene:           0.2
Population prevalence of the disease:           0.384
Genetic attributable fraction of the gene:       0.219
Penetrance for AA:                              0.8
Penetrance for AB:                              0.5
Penetrance for BB:                              0.3
Offset:                                          0.3
Significance level:                             0.05
Allele frequency of the marker gene:            0.2
P(Disease allele A | Marker allele A):          0.9
D' :                                             0.875

Power:                0.951
Computation time: 0

Press RETURN to clear screen
```

## 4. POWER CALCULATIONS FOR CONTINUOUS TRAITS

When power calculations for continuous traits have been selected the window shown in Figure 12 will appear on the screen. It allows the user to specify

- (1) Design: Family-types  
In this menu the user can specify the family-types (number of offspring per family, number of missing parents, the number of families sampled ...) and the ascertainment conditions for the first and second proband.
- (2) Genetic model:  
All parameters of the underlying genetic model are specified, eg frequency of the disease gene and the marker gene, mode of inheritance (additive, dominance, ...) and heritability.
- (3) Statistical parameters:  
In this menu the significance level and the offset used in the test to weight affected and unaffected offspring have to be defined
- (4) Power calculation:  
Based on the parameter choices the power is computed either by numerical integration, by approximation or by Monte-Carlo simulation.

The menu shown in Figure 12 is the central window for power calculations for continuous traits. The program will return to this menu after each submenu point is finished. For example, when the power has been computed for a specified design and the user wants to see how the design affects the power one can go back to menu point 1, change the design and then choose directly menu point 4 to compute the power for the new design. The other parameters defined in menu point 2 and 3 have not to be specified again.

FIGURE 12. Power calculations for continuous traits

```
Power calculations for family-based association test (for continuous traits)
=====
(1)  Design: Family-types & sample sizes
(2)  Genetic model
(3)  Statistical parameters
(4)  Power computation

<0>  Exit
```

4.1. **Design: Family-types and sample sizes.** Here we describe how to define the design and the ascertainment conditions of the study. When no design has been specified yet and submenu 1 in Figure (12) is selected the window shown in Figure (13) will be displayed. The user can then specify the family-types and the ascertainment conditions. Further, one can save the current design or load a previous design.

- Family-types: Figure (14) shows the definition of a family-type. The user has to enter the total number of offspring per family, the number of missing parents, the number of families sampled and whether the additional offspring (here every offspring other than the first proband) are phenotyped. After that all so far defined ascertainment condition are shown. The user can selected one of maximal 9 ascertainment condition. Ascertainment condition 1 (population sample) is predefined and can not be changed by the user.
- Ascertainment conditions: Figure (15) shows the ascertainment condition which are already defined. Ascertainment condition 1 (equivalent to a total population sample) is predefined and can not be changed by the user. Figure (16) shows an example for a definition of an ascertainment condition. The sampling conditions for the phenotypes of the first and second proband are specified by the corresponding probabilities of the phenotypic distributions of the traits. For example, Figure (16) shows an ascertainment condition where the trait of the first proband must be in the 10th. percentile of phenotypic distribution, while the trait of the second proband must be in the first percentile of phenotypic distribution.

FIGURE 13. Power calculation for continuous traits: Family-design

```
Family design:
=====

<1> Change family design
<2> Change ascertainment conditions
<3> Save family design
<4> Load family design
<0> Clear the screen
```

FIGURE 14. Power calculation for continuous traits: Family-design

```

Family design:
=====

Add family-types:
=====
Number of offspring per family:  0 ? 3
Number of missing parents(0,1,2): 0 ? 2
Number of families:              0 ? 200
Additional offspring phenotyped?
1=yes 0=no                       1 ? 1

Ascertainment conditions:
=====
Ascertain cond 1: proband 1 in [ 0: 1]: proband 2 in [ 0: 1]

Select ascertainment condition  0 ? 1

```

FIGURE 15. Power calculation for continuous traits: Family-design cont

```

Ascertainment conditions:
=====
Ascertain cond 1: proband 1 in [ 0: 1]: proband 2 in [ 0: 1]

Select condition 2-9 (Press 0 to exit):

```

4.2. **Genetic model.** The genetic model used for the power calculation is specified by the mode of inheritance (additive model, recessive model or dominant models), the allele frequency of the disease gene and the heritability (Figure 17). Further, the user has to choose whether the disease locus and the marker locus are identical (Figure 17). When the marker locus and the disease locus are different, the allele

FIGURE 16. Power calculation for continuous traits: Family-design cont

```
Specification of ascertainment condition 2
```

```
Proband 1: lower limit: 0 ? .9
Proband 1: upper limit: 1 ? 1
```

```
Proband 2: lower limit: 0 ? 0
Proband 2: upper limit: 1 ? .1_
```

frequency of the marker gene has to be specified (Figure 17). The linkage disequilibrium between the marker and the disease locus is defined by the probability of the disease allele conditional on the marker allele (Figure 17). The potential range of this probability computed based on the marker allele frequency and the disease allele frequency is also shown on the screen (Figure 17).

**4.3. Statistical parameters.** In this sub-menu the user has to define the statistical parameters, ie the ascertainment condition, the significance level and the offset (Figure 18).

- Offset
 

The offset weights the offspring based on their traits. Based on genetic model and the ascertainment condition PBAT computes the phenotypic mean of the population. While the total population mean seems to be always a good choice for the offset regardless of the ascertainment condition, offset choices close to the sample mean should be avoided, when an strong ascertainment condition is used (eg first percentile of phenotypic distribution is sampled).
- Significance level of the FBAT

FIGURE 17. Specification of the genetic model

```
Genetic model:
=====
Model: Add=0, Recess=1, Dom=2          0 ?
Allele frequency of the disease gene:  0.1 ? .2
Heritability:                          0.1 ? .2

Disease locus = marker locus? Yes=1 No=0 0 ?
Allele frequency of the marker gene:     0 ? .15

P(Disease allele A| Marker allele A) in [0;1]:
1 ? .9_
```

FIGURE 18. Statistical parameters

```
Statistical parameters:
=====
Total population mean:  0.334
Offset:                0 ?
Significance level:    0.05 ? .01_
```

**4.4. Power computation.** In this submenu the power of the FBAT statistic is computed based on the parameters defined by the user in submenus 1-3. The user has the choice between (Figure 19)

- "1": MCMC-integration: not yet implemented
- "2": Numerical integration:  
The analytical power of FBAT is computed based on numerical integration with a precision of 0.01. This routine can take between 1-60 seconds.
- "3": Approximation  
This routine computes the analytical power of FBATs based on a second-order Taylor expansion. The precision of this method is good for sample sizes of at least 100 families. However, this approach should be avoided when families with more than 3 offspring are analyzed.
- "4": Simulation  
The power of FBAT is assessed by simulation experiments. The power is estimated based on 100,000 Monte-Carlo simulation experiments. Such a Monte-Carlo simulation can take a few minutes.

Figure (20) shows the result of the power computations. First, the family design is given, i.e. which family-types have been used, how many families of this type have been sampled, ... . Then all genetic and statistical parameters are displayed. Finally, the power and the computation time are given. All results are written to the log-file "ptdt-log.txt".

FIGURE 19. Power calculations for continuous traits

```

Computation method :
=====
2=numerical integration, 3=approximation, 4=simulation: _

```

FIGURE 20. Power calculations for continuous traits

```

Additional offspring phenotyped
# offspring: 3 # missing parents: 2 # families: 200

other parameters:
=====
Model: Add=0, Recess=1, Dom=2           0
Allele frequency of the disease gene:   0.2
Heritability:                           0.2
Total population mean:                   0.334
Offset:                                  0
Significance level:                       0.01
Allele frequency of the marker gene:     0.15
P(Disease allele A| Marker allele A):    0.9
D' :                                       0.875

Power:           0.985
Computation time: 11

Press RETURN to clear screen

```

## REFERENCES

1. N.M. Laird, S. Horvath, and X. Xu, *Implementing a unified approach to family based tests of associaton*, Genetic Epi **19**(Suppl 1) (2000), no. S36-S42.
2. C. Lange and N.M. Laird, *Analytical sample size and power calculations for a general class of family-based association tests: Dichotomous traits*, submitted (2001).
3. ———, *On a general class of conditional tests for family-based association studies in genetics: the asymptotic distribution, the conditional power and optimality considerations*, submitted (2001).
4. J.C. Whittaker and C.M. Lewis, *Power comparisons of the transmission/disequilibrium test and sibtransmission/disequilibrium-test statistics*, American Journal of Human Genetics **65** (1998), no. 578-580.