# POLYUNPHASED

Version 1.0

Complement to the UNPHASED user guide

By

Alexandre Bureau

And

Jordie Croteau

Université Laval

And

Centre de recherche de l'Institut Universitaire en Santé Mentale de Québec

Québec, Qc, Canada

November 2015

## Introduction

Polyunphased is an extension of Unphased, which is an application for performing genetic association analysis in nuclear families and unrelated subjects. This extension was implemented to perform polytomous analyses by treating jointly two binary outcomes, T1 and T2, or using one polytomous outcome.

Typically, T2 can be the disease and T1 can be a related endophenotype. The reference phenotype in this joint analysis is defined by T1=T2=0, i.e. not affected by the disease nor the endophenotype.

The method is described in :

Bureau, A and Croteau, J. *Polyunphased : an extension to polytomous outcomes of the Unphased package for family-based genetic association analysis* (under review)

The Unphased software and its documentation can be found at: https://sites.google.com/site/fdudbridge/software/unphased-3-1.

## Installation and execution

The C++ source code, executable files for Ubuntu Linux and Windows, and the Java graphical user interface (GUI) are available under the GNU General Public License at github.com/abureau/polyUnphased.

Installation can be done in exactly the same way as Unphased (see Unphased documentation). The only difference is the name of the executable file (polyunphased instead of unphased).

Most functionalities of Unphased are also available within PolyUnphased, with the same command line and GUI options and same input file formats. Polyunphased analysis can be run by additionally using the options –joint or –polytomous (see "options" section below and Polyunphased paper for more details). However, note that the covariate options (from the Covariate menu) are not yet available, and that unrelated individuals are not allowed with polytomous outcomes.

## Input files

Two files are required: one data file and one pedigree file, with the same formats as Unphased. The outcome names must be contained in the data file. No phenotype file should be provided. Below is an example of data file with two affection status entries (A) in the case of two binary outcomes. In the case of one polytomous outcome, only one affection status entry (A) is needed.

A T1
A T2
M SNP1
M SNP2
…

The columns of the ped file following the 5th column have to correspond to the data file description, in the same order.

## Options

*-joint <Pheno1:Pheno2>*
or
*-polytomous <Polytomous_outcome_name>*

Here's an example of command line for this analysis:
*polyunphased -pedfile pedfile.txt -datafile datafile.txt -joint T1:T2*

## Output

Outputs are very similar to Unphased outputs (see the "screen output" section of Unphased manual). There are basically three differences:

1) **Reference trait level**: In the case of two binary outcomes, level 4 corresponds to T1=T2=0, which is the reference. So the output will include haplotype (or allele) effect estimations for each one of the levels 1, 2 and 3, which are defined by:
   Level 1 : T1=2 and T2=1
   Level 2 : T1=1 and T2=2
   Level 3 : T1=2 and T2=2
   In the case of a polytomous outcome, the reference is the highest level (2, 3 or 4). As a consequence, if the outcome is dichotomous, the estimated odds ratio will be the inverse of the one calculated by Unphased.
2) **Trans/Untrans:** The estimated numbers of transmitted and untransmitted haplotypes (or alleles) in offspring (from both parents) are given for each level of the outcome. Contrary to Unphased, no transmissions to unaffected offspring are included in the untransmitted counts, they are instead reported as transmissions to the level corresponding to unaffected). Thus in the case of a dichotomous outcome, the numbers of untransmitted haplotypes differ from those calculated by Unphased because the latter includes transmissions to unaffected and all non-transmissions to affected and unaffected children.
3) **Odds-R:** Estimates of haplotype (or allele) odds ratios are calculated for each non-reference outcome with respect to the reference.