# Fitting

# a classical twin model using SOLAR

David Duffy

*Queensland Institute of Medical Research*
*Brisbane, Australia*

# Overview of SOLAR

- Getting twin data into SOLAR using Sib-pair
- Running variance components twin models in SOLAR
- Running MFT twin models in SOLAR
- Comparison with other programs

SOLAR is quite a mature package, though new functions are still being added. One can obtain more up-to-date development versions from the authors.

SOLAR’s philosophy is to join together a number of other programs (FISHER, MLINK, specialized executables) via a unified Tcl-based front end.

The IBD estimation approach used by SOLAR is applicable to large pedigrees, and although an approximation, agrees well with other methods.

# Getting twin data into SOLAR

Like FISHER (on which the main variance components engine is based) and MENDEL, SOLAR is one of the few commonly used genetics packages that knows about monozygotic twins (others include MERLIN).

MZ twins are indicated by a shared MZTWIN ID variable in the comma-delimited (CSV) pedigree file. Each pair is given a unique ID – there cannot be more than 999 twin pairs in pedigree file (An error message does not always appear).

The family shared environment is indicated by a similar indicator variable, the HHID (household ID). Each household has a unique HHID value. If you wish, HHID can actually cross families (if one had two families in the same house, for example).

```
famid,id,fa,mo,sex,mztwin,hhid
00002,0000203„,m„1
00002,0000204„,f„1
00002,0000201,0000203,0000204,m,1,1
00002,0000202,0000203,0000204,m,1,1
00010,0001003„,m„2
00010,0001004„,f„2
00010,0001001,0001003,0001004,m„2
00010,0001002,0001003,0001004,f„2
…
10950,1095003„,m„2011
10950,1095004„,f„2011
10950,1095001,1095003,1095004,f,998,2011
10950,1095002,1095003,1095004,f,998,2011
…
21705,2170503„,m„3010
21705,2170504„,f„3010
21705,2170501,2170503,2170504,m„3010
21705,2170502,2170503,2170504,f„3010
```

# The SOLAR phenotype data file

Phenotype data is in a separate CSV file from the pedigree data.

```
famid,id,asm12_ca,haf12_ca,age_80,sx
00002,0000203„„1.0000
00002,0000204„„0.0000
00002,0000201,2.0000,2.0000,36.0000,1.0000
00002,0000202,1.0000,2.0000,36.0000,1.0000
00010,0001003„„1.0000
00010,0001004„„0.0000
00010,0001001,1.0000,1.0000,25.0000,1.0000
00010,0001002,1.0000,1.0000,25.0000,0.0000
00022,0002203„„1.0000
00022,0002204„„0.0000
…
21705,2170503„„1.0000
21705,2170504„„0.0000
21705,2170501,1.0000,2.0000,25.0000,1.0000
21705,2170502,1.0000,1.0000,25.0000,0.0000
```

# Running a variance components twins analysis in SOLAR

```
davidD@moonboom:~/Genetics/Korea$ solar

SOLAR version 2.1.4 (Official), last updated on December 30, 2004
Copyright (c) 1995-2004 Southwest Foundation for Biomedical Research
Enter help for help, exit to exit, doc to browse documentation.
```

Firing up the linux version of SOLAR.

# Running a variance components twins analysis in SOLAR

```
solar> load pedigree asthmaex.ped
Unloading current pedigree data …
Loading pedigree data from the file asthmaex.ped …
solar> load phenotypes asthmaex.dat
asthmaex.dat: famid id asm12_ca haf12_ca age_80 sx
solar> trait asm12_ca
solar> house
solar> polygenic
```

The “trait” command declares the variable to be analysed,  A multivariate analysis is asked for by including more than one variable.  If the trait is found to have two levels, the MFT is fitted.

The “house” command tells SOLAR to prepare a “C” family environment covariance matrix. The “polygenic” command runs an E, AE, CE, and ACE model.

# Running a variance components twins analysis in SOLAR 2

```
*****************************************************************
*  Maximize sporadic model                                      *
*****************************************************************
    *** Loglikelihood of sporadic model is -2383.165778
*****************************************************************
*  Maximize polygenic model                                     *
*****************************************************************
    *** Loglikelihood of polygenic model is -2265.442698
    *** H2r in polygenic model is 0.7459631
*****************************************************************
*  Maximize household model                                     *
*****************************************************************
    *** Loglikelihood of household model is -2274.437471
    *** C2 in household model is 0.6112551
```

```
***********************************************************************
*  Maximize household polygenic model                                 *
***********************************************************************
    *** Loglikelihood of household polygenic model is -2261.683314
    *** H2r in household polygenic model is 0.4606102

    *** C2 in household polygenic model is 0.2725513

    *** Determining significance of C2
    *** Comparing household polygenic and polygenic models
    *** chi = 7.5188, deg = 1, p = 0.0030530

    *** Determining significance of H2r
    *** Comparing household polygenic and household models
    *** chi = 25.5083, deg = 1, p = 0.0000002
```

```
******************************************************************
*                     Summary of Results                         *
******************************************************************
  Pedigree:    asthmaex.ped
  Phenotypes:  asthmaex.dat
  Trait:       asm12_ca              Individuals:  6020
                   H2r is 0.4606102  p = 0.0000002  (Significant)
         H2r Std. Error:  0.0419633
                    C2 is 0.2725513  p = 0.0030530  (Significant)
          C2 Std. Error:  0.0401069

  No pedigrees shared households with other pedigrees
  Output files and models are in directory asm12_ca/
  Summary results are in asm12_ca/polygenic.out
  Loglikelihoods and chi’s are in asm12_ca/polygenic.logs.out
  Best model is named housepoly and null0 (currently loaded)
  Final models are named housepoly, house, poly, spor
```

# Adding in covariates/fixed effects

Adding in fixed effects (to what in Mx is called the "means model") is simply:

```
solar> covariate age_80 sx
solar> polygenic
```

# Comparison with other programs

The only other program that can easily fit the multifactorial threshold model (MFT) for twins is Mx. Programs such as PAP can also fit the MFT, but don’t know about monozygotic twins.

Solar automatically created the model and the necessary matrices, while these all have to be created by the user in Mx.

Solar cannot currently do a multitrait MFT analysis, but can do a bivariate continuous trait analysis. A linkage analysis can be performed by just adding two more commands. Mx is completely flexible, providing you don’t run out of matrices.