Pedigree Test Harness

# Motivation

For the disease risk calculator program, pedigrees are specified in the well known ped/linkage file format. Erroneous pedigrees can easily be specified using this format, e.g. one can specify a pedigree were someone is their own ancestor. The disease risk calculator program performs various checks to validate the pedigree is sensible. This document describes testing done to ensure the correctness of the disease risk calculator program’s pedigree validation.

# Introduction

A set of ped file test cases are created. These may be valid or invalid specifications of pedigrees. A program known as a test harness applies the pedigree validation to each of these ped files. The test harness output allows one to check whether the pedigree validation is working as expected. The test harness program is called

PedigreeTestHarness.R

This is an R program, designed to be run from the command line.

It takes as its input the name of a ped file.

It reads this ped file to extract the pedigree information.

It runs pedigree validation on the pedigree information.

If the pedigree is validated a pedigree diagram is generated.

If any problems are detected, the program reports the problem and stops.

The test ped files are held in directory – ValidatePedigree

This contains two sub-directories

* good – contains ped files that should run without any errors being reported
* bad – contains problem ped files, a meaningful error should be reported

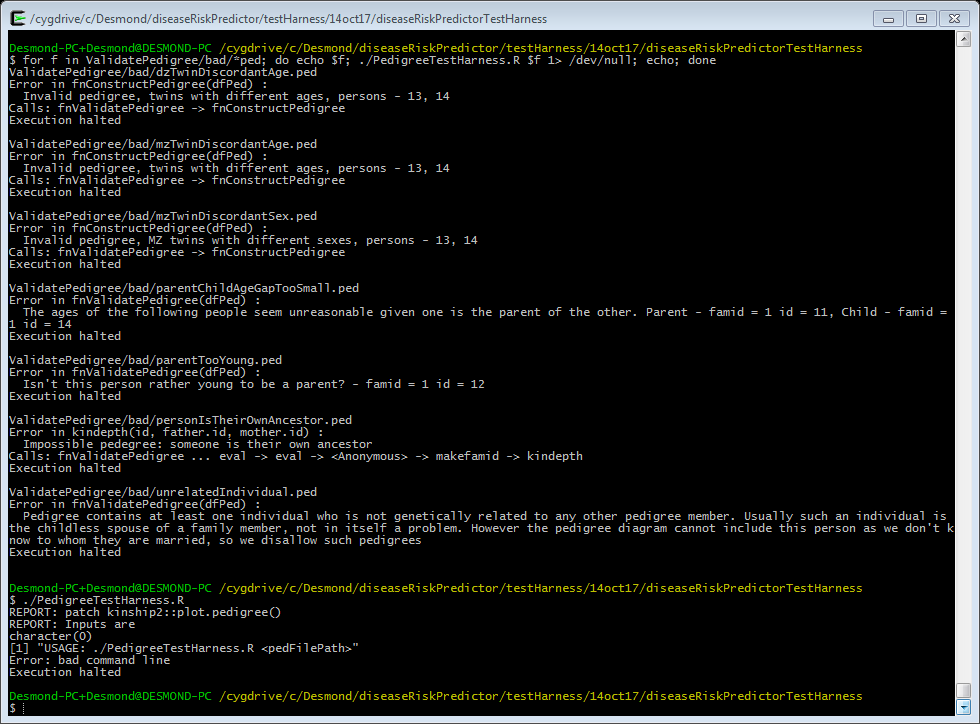
# Check the Test Harness is Working

Open a command line prompt.

Go to the directory containing the test harness program

Enter the following at the command line prompt

./PedigreeTestHarness.R

The program should report a USAGE statement telling you what constitutes a valid command line for the program. This is what that looked like when run on my machine from a Cygwin prompt 

If you do not get a USAGE statement, then the program probably did not run. The following requirements must be met for the program to run

* the file PedigreeTestHarness.R needs execute permissions
* R must be on the PATH

# Testing a ped file

First we are going to test a valid ped file. This file contains pedigree information defining a three child nuclear family.

Enter the following at the command prompt

./PedigreeTestHarness.R ValidatePedigree/good/nuclearFamily.ped

You should see output like the following

$ ./PedigreeTestHarness.R ValidatePedigree/good/nuclearFamily.ped

REPORT: Inputs are

[1] "ValidatePedigree/good/nuclearFamily.ped"

REPORT: Read pedigree from file - ValidatePedigree/good/nuclearFamily.ped

REPORT: Pedigree is

famid id fatherid motherid sex affected

1 1 11 0 0 1 0

2 1 12 0 0 2 1

3 1 13 11 12 1 0

4 1 14 11 12 2 1

5 1 15 11 12 1 1

REPORT: run fnValidatePedigree()

...

REPORT: run fnConstructPedigree()

...

REPORT: writing plots to file - PedigreeTestHarness.R.plots.pdf

REPORT: run plot()

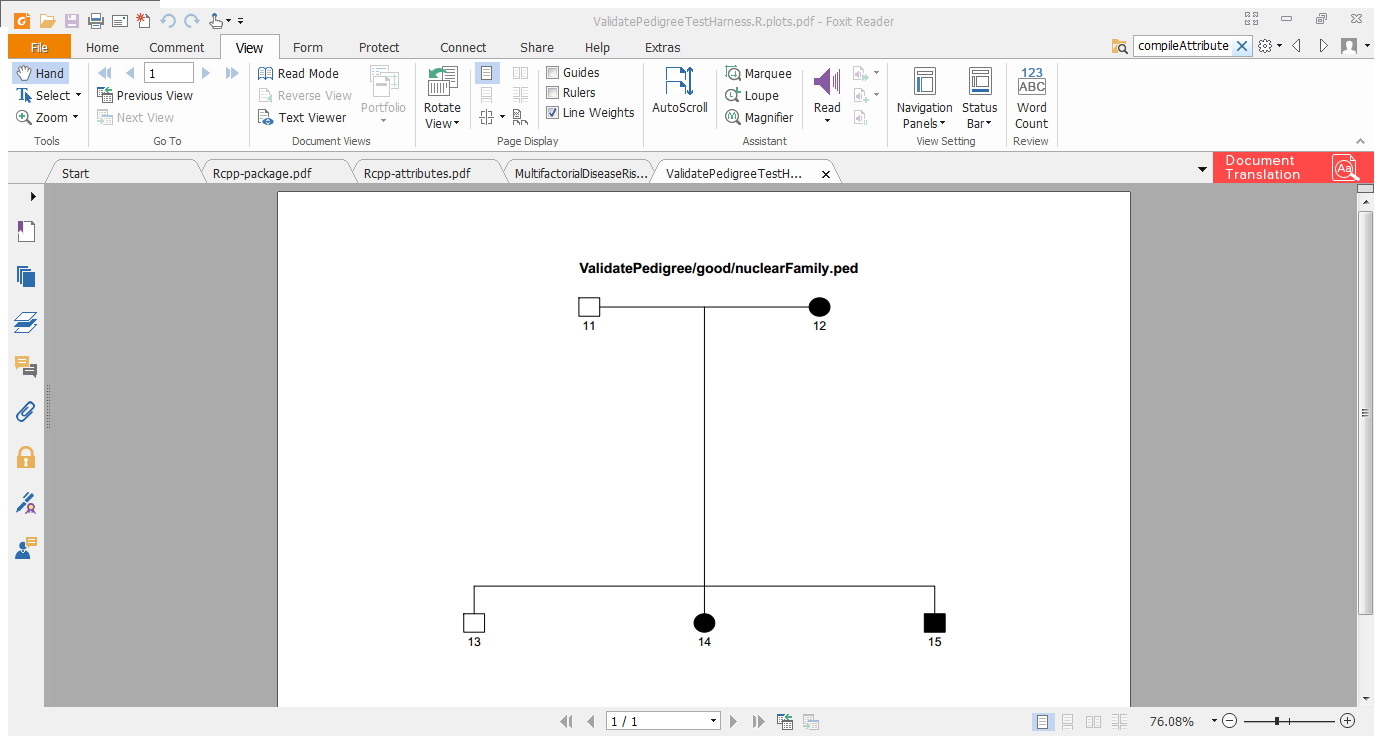
...

REPORT: Completed Successfully

The final line says the test harness program completed successfully.

A few lines earlier is given the name of the file into which the pedigree diagram was drawn.

Open this pdf file. You should see a pedigree diagram like the following



**NB:** Be sure to close this pdf file after you’ve looked at it. If you don’t, then the next time the test harness is invoked it will not be able to write a new diagram into the file. It won’t have permission to do this and will report an error (even if the ped file being tested is valid).

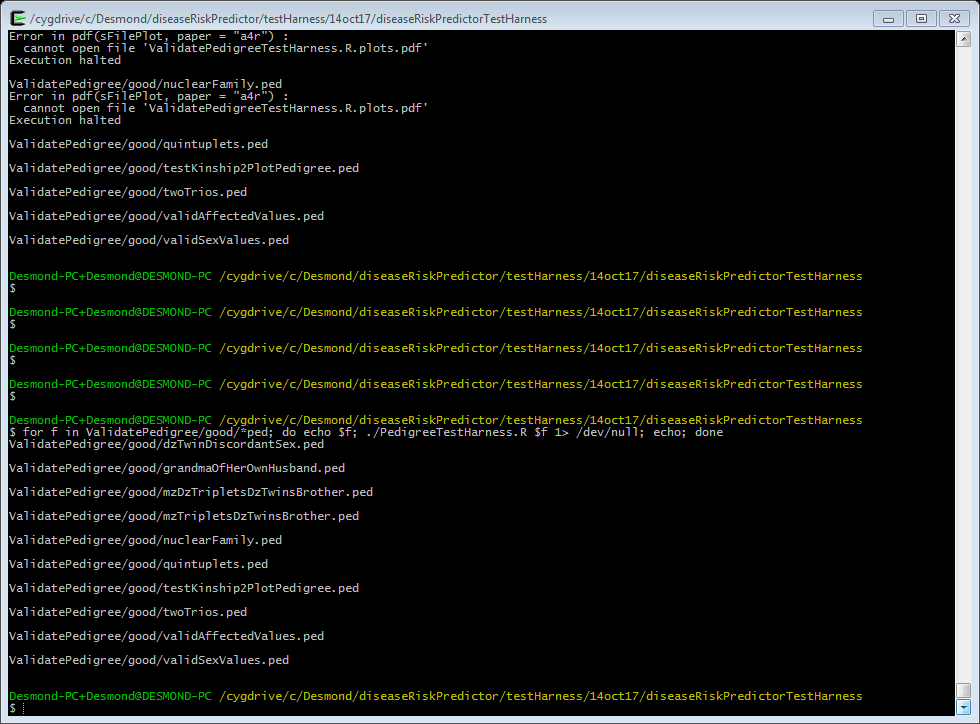
# Testing the good ped files

Testing each ped file individually is a bit tedious; let’s automate that. We will apply the following procedure to each ped file in the ValidatePedigree\good directory

* report file name
* invoke the test harness program on the ped file
* discard all standard (i.e. non-error) output

Enter the following command at the command prompt.

for f in ValidatePedigree/good/\*ped; do echo $f; ./PedigreeTestHarness.R $f 1> /dev/null; echo; done

You should see output like 

You should see reported the name of each ped file tested, and nothing else. All these ped files should be valid. If anything else is reported there is a problem.

The following command is adapted from the above one. It renames the pedigree diagram file created to an appropriately named one, allowing you to see the pedigree diagrams generated for each test case.

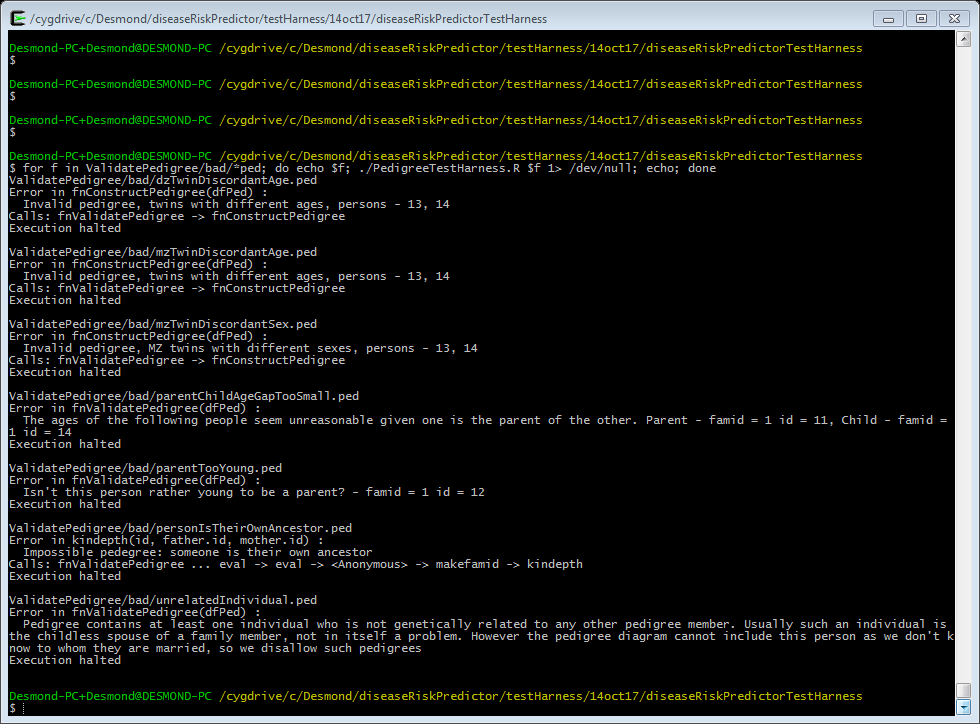
for f in ValidatePedigree/good/\*ped; do echo $f; ./PedigreeTestHarness.R $f 1> /dev/null; bf=`basename $f`; mv PedigreeTestHarness.R.plots.pdf ${bf}.pdf; echo; done

# Testing the bad ped files

We can do the same for the bad ped files. These are ped files constructed so as to be invalid in some way.

Enter the following command at the command prompt.

for f in ValidatePedigree/bad/\*ped; do echo $f; ./PedigreeTestHarness.R $f 1> /dev/null; echo; done

You should see output like 

The test harness should report an error for each ped file. Check that the error reported makes sense given the name of the ped file.