CanRisk Data File Format

2019-06-12

Description

The CanRisk file contains a target individual's risk factors (optional) in the header followed by pedigree data (mandatory) as a series of pedigree data records, one for each family member. The following sections cover the File Header, Pedigree Data and Multiple Pedigree Data.

File Header

The header can optionally include the following breast and/or ovarian cancer risk factors for the target individual:

|  |  |  |  |
| --- | --- | --- | --- |
| **VARIABLE NAME** |  | **RISK FACTOR DESCRIPTION** | **ALLOWED VALUES** |
| Menarche | † | Age at menarche | NA=unspecified, integer=age at menarche |
| Parity | †‡ | Parity | NA=unspecified, integer=number of children |
| First\_live\_birth | † | Age at first live birth | NA=unspecified, integer=age at first live birth |
| OC\_use | †‡ | Current use of oral contraception | NA=unspecified, N=Never, F:Years=Former, C:Years=Current; F and C are followed by integer=number of years taken e.g. F:4 means former use of 4 years |
| MHT\_use | †‡ | Current use of menopause hormone therapy | NA=unspecified, N=Never, F=Former, E=Current E-type, C=Current C-type |
| BMI | † | Body mass index | NA=unspecified, real number=body mass index |
| Alcohol | † | Daily alcohol intake in grams per day | NA=unspecified, real number=daily alcohol intake in grams |
| Menopause | † | Age at menopause | NA=unspecified, integer=age at menopause |
| BIRADS | † | Mammographic density as measured by BI-RADS | NA=unspecified, integer=BI-RADS classification (1 2 3 4) |
| Height | †‡ | Height in cm | NA=unspecified, real number=height in cm |
| TL | ‡ | Tubal ligation procedure | NA=unspecified, N=No, Y=Yes |
| Endo | ‡ | Endometriosis | NA=unspecified, N=No, Y=Yes |
| PRS\_BC | † | Polygenic Risk Score (Breast Cancer) | e.g. alpha=real number, beta=real number |
| PRS\_OC | ‡ | Polygenic Risk Score (Ovarian Cancer) | e.g. alpha=real number, beta=real number |

† Breast Cancer Risk Factor ‡ Ovarian Cancer Risk Factor

The PRS\_BC and PRS\_OC values are:

* alpha - the square root of the proportion of the overall BOADICEA polygenic variance explained by the PRS. A real number between 0 and 1.
* beta – the standard normal PRS.

All header lines begin with ‘##’. Any missing risk factor variables are taken as unspecified. The risk factors are given as one per line, the variable names can be any case and in any order in the header. The following example header shows the first mandatory header record **##CanRisk 1.0** followed by some of the optional risk factors for the target in the pedigree. The last line in the header is the second mandatory pedigree data column header record beginning **##FamID Name…**.

##CanRisk 1.0  
##Menarche=13  
##Parity=1  
##First\_live\_birth=24  
##OC\_use=C:2  
##BMI=27.1  
##height=170  
##alcohol=5.1  
##PRS\_BC=alpha=0.45, beta=1.8  
##FamID Name Target IndivID FathID MothID Sex MZtwin Dead Age Yob BC1 ….......

Pedigree Data

The CanRisk pedigree data format is a simple TAB-delimited text format. CanRisk pedigree data files consist of the two mandatory header records followed by a series of pedigree data records, one for each family member. The pedigree data records include 26 parameters (data columns) separated by one or more TAB (or whitespace) characters.

Parameters 1-26 on the pedigree data records are defined as follows:

**FamID**  Family/pedigree ID, character string (maximum 13 characters)  
**Name** First name/ID of the family member, character string (maximum 8 characters)  
**Target** The family member for whom the BOADICEA risk calculation is made, 1 = target for BOADICEA risk calculation, 0 = other family members.  
**IndivID** Unique ID of the family member, character string (maximum 7 characters)  
**FathID** Unique ID of their father, 0 = no father, or character string (maximum 7 characters. Each family member must have either: (1) no parents specified (e.g. see family member '103' in the accompanying sample pedigree '1. Example without risk factors'), or (2) both parents specified (e.g. see family member '201' in the accompanying sample pedigree '1. Example without risk factors').  
**MothID** Unique ID of their mother, 0 = unspecified, or character string (maximum 7 characters)  
**Sex** M or F  
**MZtwin**  Identical twins, 0 = not an identical twin, Use one of these characters to identify MZ twins: 1 2 3 4 5 6 7 8 9 A (e.g. see '301' and '302' above)  
**Dead** The current status of the family member, 0 = alive, 1 = dead  
**Age** Age at last follow up, 0 = unspecified, integer = age at last follow up  
**Yob** Year of birth, 0 = unspecified, or integer (consistent with Age if the person is alive)  
**BC1** Age at first breast cancer diagnosis, 0 = unaffected, integer = age at diagnosis, AU = unknown age at diagnosis (affected unknown)  
**BC2** Age at second (contralateral) breast cancer diagnosis, 0 = unaffected, integer = age at diagnosis, AU = unknown age at diagnosis (affected unknown)  
**OC** Age at ovarian cancer diagnosis, 0 = unaffected, integer = age at diagnosis, AU = unknown age at diagnosis (affected unknown)  
**PRO** Age at prostate cancer diagnosis 0 = unaffected, integer = age at diagnosis, AU = unknown age at diagnosis (affected unknown)  
**PAN** Age at pancreatic cancer diagnosis 0 = unaffected, integer = age at diagnosis, AU = unknown age at diagnosis (affected unknown)  
**Ashkn** Ashkenazi status, 0 = not Ashkenazi, 1 = Ashkenazi  
**BRCA1** BRCA1 genetic test type:result; type 0=untested, S=mutation search, T=direct gene test; result 0=untested, P=positive, N=negative  
**BRCA2** BRCA2 genetic test type:result; type 0=untested, S=mutation search, T=direct gene test; result 0=untested, P=positive, N=negative  
**PALB2** PALB2 genetic test type:result; type 0=untested, S=mutation search, T=direct gene test; result 0=untested, P=positive, N=negative  
**ATM** ATM genetic test type:result; type 0=untested, S=mutation search, T=direct gene test; result 0=untested, P=positive, N=negative  
**CHEK2** CHEK2 genetic test type:result; type 0=untested, S=mutation search, T=direct gene test; result 0=untested, P=positive, N=negative  
**RAD51D** RAD51D genetic test type:result; type 0=untested, S=mutation search, T=direct gene test; result 0=untested, P=positive, N=negative  
**RAD51C** RAD51C genetic test type:result; type 0=untested, S=mutation search, T=direct gene test; result 0=untested, P=positive, N=negative  
**BRIP1** BRIP1 genetic test type:result; type 0=untested, S=mutation search, T=direct gene test; result 0=untested, P=positive, N=negative  
**ER:PR:HER2:CK14:CK56** Colon separated Estrogen receptor, Progestrogen receptor, Human epidermal growth factor receptor 2, Cytokeratin 14, Cytokeratin 56 status, 0 = unspecified, N = negative, P = positive