N Year Risk Prediction

# Motivation

This document describes the method used by MultifactorialDiseaseRiskCalculator.R (and its online version), for predicting unaffected disease pedigree members’ *n* year risk.

# What is *n* Year Risk?

The *n* year risk for a disease for a currently unaffected person is the risk that they will become affected with the disease within the next *n* years. This statistic is of interest for screening programs for e.g. breast cancer and prostate cancer. It may be used to determine how frequently to screen. It may also be used to justify targeting high risk individuals for more frequent screening.

We further define it to be

* NA - for those whose affection status is unknown
* 0 - for affected individuals

# Modelling of *n* Year Risk

## Singleton Example

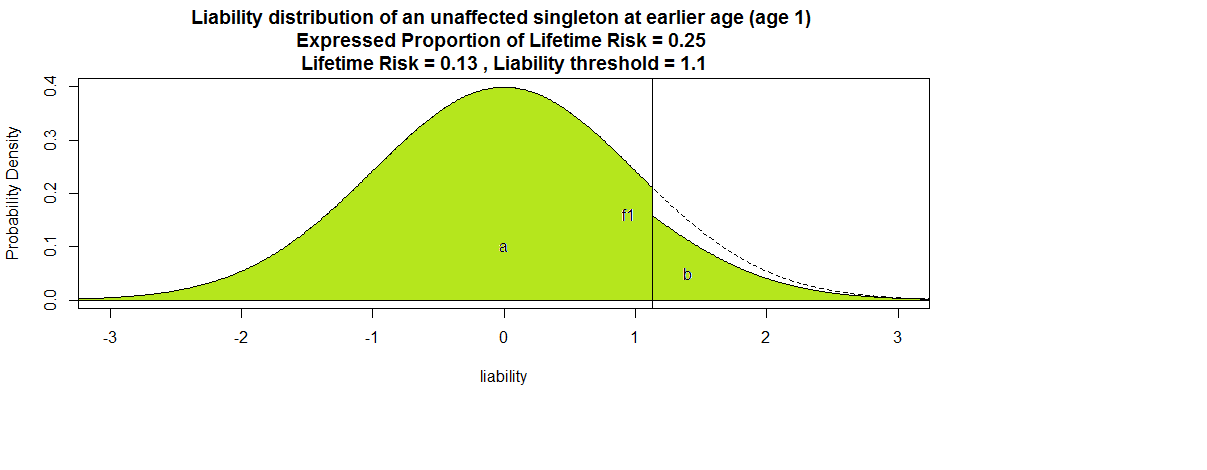
Assume a multifactorial disease has a lifetime risk of 13%, and that the disease is only expressed in later life. By age 1, a quarter of those destined to become affected by the disorder will have done so. By age 2, *n* years later, two thirds of those destined to become affected have become affected.

Figure 1 shows the liability distribution for a currently unaffected singleton aged age 1. If the person’s liability is in area *a* then they will never develop the disorder, if in area *b* they will eventually develop the disorder.

Figure 1

Liability distribution of an unaffected singleton at age 1 (green area).

* a = area of distribution where liability < critical threshold
* b = area of distribution where liability >= critical threshold
* = 1 - expressed proportion of lifetime risk at age 1

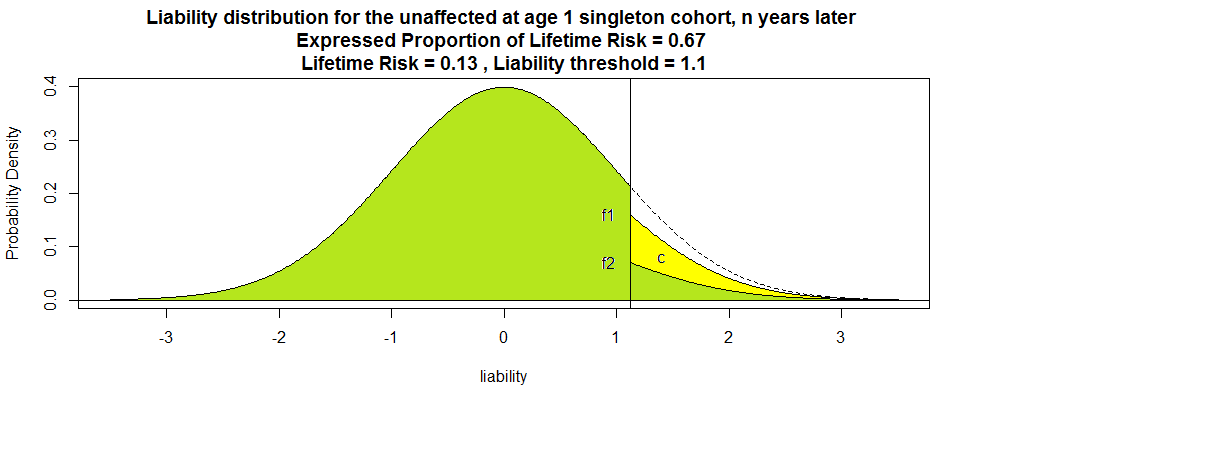


If we look at the stratum of all such people, *n* years later (at age 2), then many of them will have become affected, as depicted in Figure 2.

Figure 2

Liability distribution for the unaffected at age 1 singleton cohort, *n* years later

* Green area – liability distribution of those who remain unaffected
* Yellow area - liability distribution of those who have become affected
* c = area of yellow distribution
* = 1 - expressed proportion of lifetime risk at age 1
* = 1 - expressed proportion of lifetime risk at age 2



## Formula

The *n* Year risk is the proportion of the age1 liability distribution that has become yellow in Figure 2.

## N Year Risk in Pedigrees

The above example explains n year risk prediction in a singleton. N year risk prediction for disease pedigree members is very similar. After conditioning on risk factors and family history for disease, the liability distribution for unaffected pedigree members is likely to be mean shifted away from zero and to have a variance of less than one. Having found a disease pedigree unaffected’s distribution, modelling of *n* year risk for proceeds in the same way as for singletons.

# Method for *n* Year Risk Prediction

The risk prediction program is able to predict individuals’ *n* year risks because all the parameters of the *n* year risk formula can be either calculated or estimated. Parameter can be interpolated from the disease’s expressed proportion of lifetime risk curve appropriate for the person, and a person’s age. Parameter can be calculated similarly after making the person *n* years older. The program estimates the disease pedigree's posterior liability (posterior to all risk related info) by obtaining (via Gibbs sampling) a large number of draws (say 10,000) from this distribution. For each unaffected pedigree member, *a* and *b* can be estimated by the proportion of these draws that are below or above the critical threshold for the individual.

# Technical

The figures were generated via script - plotDiagramsForNYearRisk.R