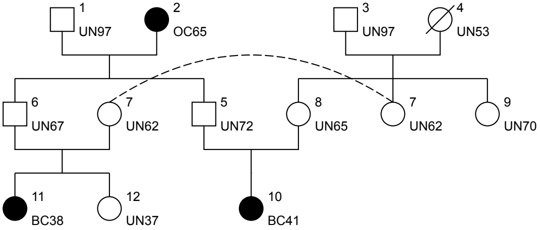
Computing competing risks based on family history in genetic diseases with variable age at onset

Alexandra Lefebvre, Olivier Bouaziz, Grégory Nuel
Curie Institut - LPMA, UPMC - University Paris 11, France

SAM 2017
University of Liverpool

Introduction : Context

- The breast cancer
  - 1st cancer in women. 55,000 women in the UK each year
  - Complex disease due to an accumulation of mutations (ie. : BRCA 1/2 and/or PALB2 and/or RAD51 and/or, etc.)
  - 10 to 15% cases : inherited mutation.
- The genetic counseling



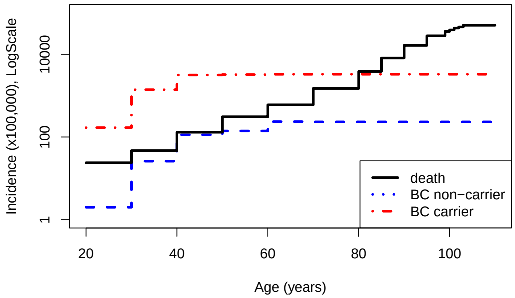
Genetic testing
$\mathbb{P}(\text{genetic predisposition} \mid \mathrm{FH})$
Individual risk of the disease $\mid \mathrm{FH}$
$\Big\}$ → recommendations

Introduction : existing models

- Empirical : The GAIL model (logistic regression)
- Mendelian : Claus-Easton, BRCAPro, BOADICEA, etc.

The Claus-Easton model [Claus et al., 1991, Easton et al., 1993]

- Autosomal, biallelic, dominant, estimated allele frequency f=0.33%
- The hazard functions per genotype (piecewise constant) :



Objectives : Impl. sum/product algorithm + competing risk of death

The model : the likelihood and the genotypes

$$\mathbb{P}(X, \mathrm{FH}) = \underbrace{\prod_i \mathbb{P}\left(X_i | X_{\mathrm{pat}_i}, X_{\mathrm{mat}_i}\right)}_{\text{genotypes}} \times \underbrace{\prod_i \mathbb{P}\left(\mathrm{PH}_i | X_i\right)}_{\text{phenotypes}}, \quad \mathrm{FH} = \{\mathrm{PH}_i\}_i$$

Mode of inheritance : 1 autosomal biallelic gene, f = 0.33%

Founders (**Hardy-Weinberg**) :
$$\begin{cases} \mathbb{P}(X_i = 00) = (1-f)^2 \\ \mathbb{P}(X_i = 10) = \mathbb{P}(X_i = 01) = f(1-f) \\ \mathbb{P}(X_i = 11) = f^2 \end{cases}$$

Offsprings (**Mendel**):
$$\begin{cases} \mathbb{P}(X_i = 00) = (1-\Theta(X_{\mathrm{pat}})) \times (1-\Theta(X_{\mathrm{mat}})) \\ \mathbb{P}(X_i = 10) = \Theta(X_{\mathrm{pat}}) \times (1-\Theta(X_{\mathrm{mat}})) \\ \mathbb{P}(X_i = 01) = (1-\Theta(X_{\mathrm{pat}})) \times \Theta(X_{\mathrm{mat}}) \\ \mathbb{P}(X_i = 11) = \Theta(X_{\mathrm{pat}}) \times \Theta(X_{\mathrm{mat}}) \end{cases}$$

with $\Theta(00) = 0, \quad \Theta(10) = \Theta(01) = 0.5, \quad \Theta(11) = 1$

The model : the phenotypes

With T_i, the age at disease onset for individual i

Survival data : $\mathrm{PH}_i = \begin{cases} \{T_i > \tau_i\} & \text{if } i \text{ is censored at age } \tau_i \\ \{T_i = \tau_i\} & \text{if } i \text{ is affected at age } \tau_i \end{cases}$

Dominant model of disease :

$$\lambda(t|X_i) = \begin{cases} \lambda_0(t) & \text{if } X = 00 \\ \lambda_1(t) & \text{if } X \neq 00 \text{ i.e. } \in \{10, 01, 11\} \end{cases}$$

- For a censored individual at age τ_i

$$\mathbb{P}(\mathrm{PH}_i|X_i) = \mathbb{P}(T_i > \tau_i|X_i) = \begin{cases} S_0(\tau_i) & \text{for non-carriers} \\ S_1(\tau_i) & \text{for carriers} \end{cases}$$

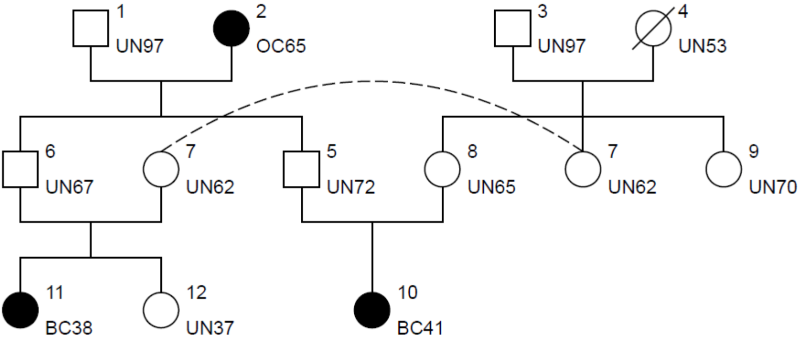
- For an affected individual at age τ_i

$$\mathbb{P}(\mathrm{PH}_i|X_i) = \mathbb{P}(T_i = \tau_i|X_i) = \begin{cases} S_0(\tau_i)\lambda_0(\tau_i) & \text{for non-carriers} \\ S_1(\tau_i)\lambda_1(\tau_i) & \text{for carriers} \end{cases}$$

Model: The bayesian network[1] ; sum-product algorithm

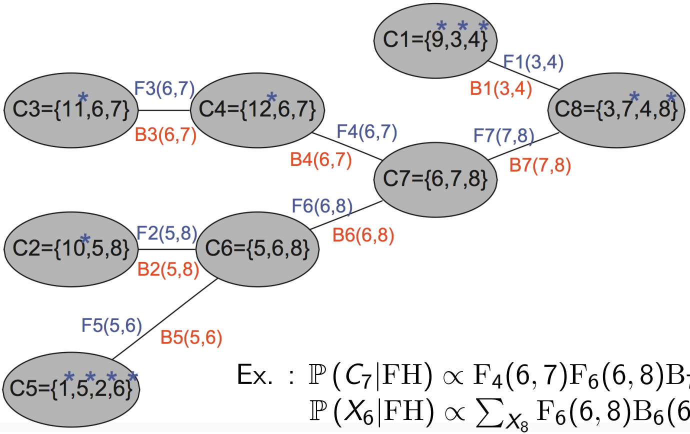
$$\mathbb{P}(X, \mathrm{FH}) = \prod_i \underbrace{\mathbb{P}\left(X_i | X_{\mathrm{pat}_i}, X_{\mathrm{mat}_i}\right) \mathbb{P}\left(\mathrm{FH}_i | X_i\right)}_{\mathrm{K}_i(X_i | X_{\mathrm{pat}_i}, X_{\mathrm{mat}_i})} \rightarrow \mathbb{P}(\mathrm{FH}) = \sum_X \prod_i \mathrm{K}_i(X_i | X_{\mathrm{pat}_i}, X_{\mathrm{mat}_i})$$

With $X \in \{00, 10, 01, 11\}^n \rightarrow 4^n$configurations



[1] [Koller and Friedman, 2009]

Belief propagation



Ex. : $\mathbb{P}\left(C_7|\mathrm{FH}\right) \propto \mathrm{F}_4(6,7)\mathrm{F}_6(6,8)\mathrm{B}_7(7,8)$

$\mathbb{P}\left(X_6|\mathrm{FH}\right) \propto \sum_{X_8} \mathrm{F}_6(6,8)\mathrm{B}_6(6,8)$

Bayesian network

Complexity $\mathcal{O}(4^n) \to \mathcal{O}(n \times 4^k)$, k:tree-width (ex: k=3 if no loop)

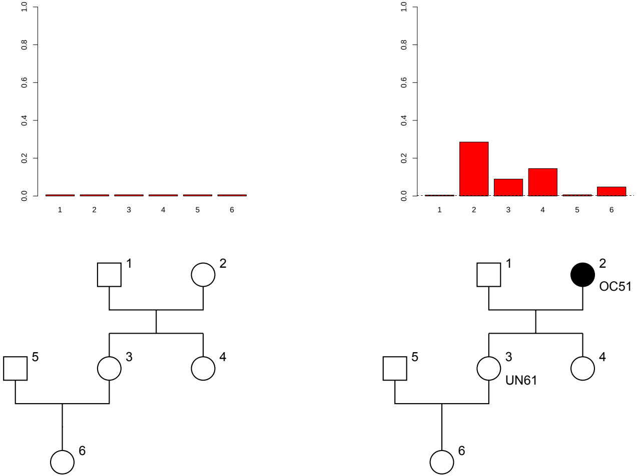
F & B computed once for any later marginal or joint distribution needed

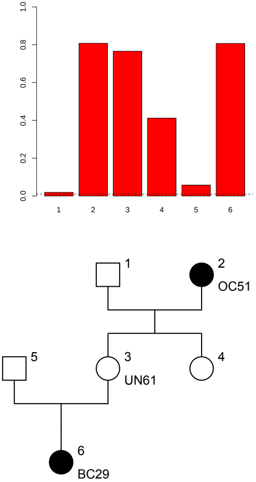
Model : Disease risk prediction for an unaffected individual

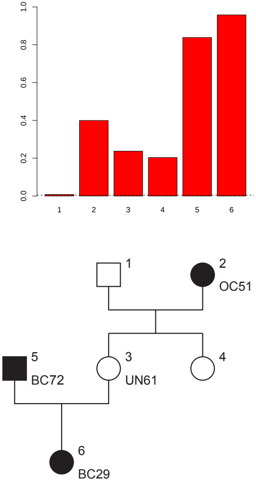
$\pi(\tau) = \mathbb{P}(\text{carrier}|\mathrm{FH})$

- Breast cancer specific, no competing risk
  $S(t|\mathrm{FH}) = \sum_{X_i} \mathbb{P}(T > t, X_i|\mathrm{FH}) = \pi(\tau)\frac{S_1(t)}{S_1(\tau)} + (1-\pi(\tau))\frac{S_0(t)}{S_0(\tau)}$
  $\pi(t|\mathrm{FH}) = \dfrac{\pi(\tau)S_1(t)}{S(t|\mathrm{FH})S_1(\tau)}$
  $\boldsymbol{\lambda_{disease}(t|\mathrm{FH})} = \pi(t|\mathrm{FH})\lambda_1(t) + (1-\pi(t|\mathrm{FH}))\lambda_0(t)$
- With competing risk of death : $T^* = \min(T_{\text{disease}}, T_{\text{death}})$
  $\lambda_{\text{both}}(t|\mathrm{FH}) = \boldsymbol{\lambda_{disease}(t|\mathrm{FH})} + \lambda_{\text{death}}(t)$: hazard function of T^*
  $$\begin{aligned}\mathbb{P}(T \le t|\mathrm{FH}) &= \int_\tau^t S_{\text{both}}(u)\lambda_{\text{disease}}(u)du \\ &= \int_\tau^t \exp\left(-\int_\tau^u \lambda_{\text{both}}(v)dv\right)\lambda_{\text{disease}}(u)du\end{aligned}$$
  $\rightarrow$ discretized λ $\rightarrow$ closed form formulas for pch

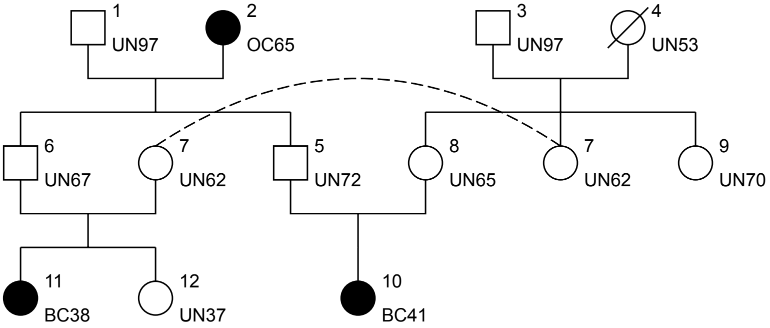
Results : Carrier risk, posterior marginal carrier distribution



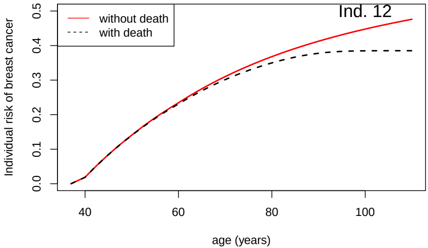

0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
1
2
OC51
5
3
UN61
4
6
BC29


0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
1
2
OC51
5
BC72
3
UN61
4
6
BC29

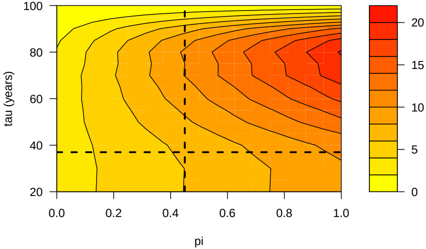
Results, Disease risk, With competing death



Results, Disease risk, With competing death



Ind. risk with vs without competing risk of death



Difference in % of S(100) with vs without competing risk

Conclusion

- The model
  - Adaptable (any genetic disease and model of disease)
  - Fast (bayesian network - sum-product algorithm - belief propagation)
  - Takes into account the competing risk of death

- What's next?
  - Parameters estimations
  - Complex distributions (number of carriers in the family) with generating functions of probabilities (polynomials) $\rightarrow$ familial risk
  - Multi-state and frailty models

Claus, E. B., Risch, N., and Thompson, W. D. (1991).
Genetic analysis of breast cancer in the cancer and steroid hormone study.
American journal of human genetics, 48(2):232.

Easton, D., Bishop, D., Ford, D., and Crockford, G. (1993).
Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. the breast cancer linkage consortium.
American journal of human genetics, 52(4):678.

Koller, D. and Friedman, N. (2009).
Probabilistic graphical models: principles and techniques.
MIT press.

Thank you for your attention