Acta Math. Hungar., **139 (1–2)** (2013), 160–182
DOI: 10.1007/s10474-012-0248-x
First published online August 1, 2012

MULTIFACTORIAL INHERITANCE WITH SELECTION

B. GERENCSÉR[1,*], B. RÁTH[2] and G. TUSNÁDY[1]

[1]Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences,
H-1364 Budapest, Pf. 127, Hungary
e-mails: gerencser.balazs@renyi.mta.hu, tusnady.gabor@renyi.mta.hu

[2]ETH Zürich, Zürich, Switzerland
e-mail: rathb@math.ethz.ch

(Received December 28, 2011; revised March 21, 2012; accepted March 23, 2012)

Abstract. We present an alternative model for multifactorial inheritance. By changing the way the malformation (and selection) is determined from the genetic information, we arrive at a model that can be properly handled in the mathematical sense. This includes the proof of population convergence and computation of conditional malformation probabilities in a closed form. We also present a comparison to similar models and results of fitting our model to Hungarian data.

1. Introduction

The concept of multifactorial inheritance goes back to Francis Galton, a contemporary of Gregor Johann Mendel (see Karlin [4]). Instead of the case investigated by Mendel, where the appearance of a congenital malformation is controlled by a single gene, in multifactorial inheritance the number of genes involved is large or infinite. As a result their effect is concentrated in a virtual quantity, the liability having standard normal distribution. The joint distribution of the liabilities of members of a family is also normal with covariances

$$\operatorname{cov}(X, Y) = \frac{h^2}{2^d},$$

determined by the remove degrees of the relationship where h is the heritability of the malformation and d is the degree of relationship. In the

* Corresponding author.

Key words and phrases: multifactorial inheritance, birth defect, balance of selection and mutation.

Mathematics Subject Classification: primary 62P10, secondary 92D25.

simplest case of $h = d = 1$ the conditional probability that a first order relative of a malformed person has the malformation is roughly $\sqrt{p}$, where p is the population incidence of the malformation. This approximation is due to A. W. F. Edwards [2]. The multifactorial model was tested on Hungarian data by Czeizel and Tusnády [1] which work was criticized by Kari Sankaranarayanan because the effect of selection was neglected. He organized a group to solve the problem and some preliminary results were published by members of the group [6] while Tusnády tested the new model on original data [7]. Unfortunately a question remained unsettled: the stability of the proposed model. Here we offer a partial solution of the problem.

Let X and Y be the liabilities of the parents, then the liability of their child is

$$Z = \frac{X + Y}{2} + U,$$

where U is a normal variable with expectation zero and variance $\frac{1}{2}$. The main observation of Sankaranarayanan was that in the case of selection the bad genes causing the malformation simply flow out from the population like the water from a bathtub. It is the mutation which can supplant the bad genes. The effect in the model may be represented by changing the expectation of U to some positive number to balance the effect of selection. As usual, let $L = Z + V$ be the liability of the investigated child, where V is the environmental effect with appropriate variance, and let us postulate that the appearance of the malformation is equivalent with the event $L > T$, where T is the threshold. (The random variables X, Y, U, V are independent.)

The effect of selection may be represented in the model by a second threshold $S > T$ such that if $L > S$ then there will be no descendant for the person having liability L. The stability of the model means that starting with an arbitrary distribution on parents in course of generations the distribution of the liability goes to a limit which is independent of the original distribution. This is observed for computer simulations but we have no theoretical proof. Instead we turn to the case of finitely many genes. In this case the environmental effect may be represented with a Poisson variable with an appropriate parameter, any bad gene will be given to the child with probability $\frac{1}{2}$, and the effect of mutation is also a Poisson variable. In the general case let $p(L)$ be the probability that a person with liability L has the malformation. (L may be identified in this case with a natural number coming partly from bad genes and partly from quantized environmental effects with the same habit as bad genes.) If $p(L) = 1$ iff $L \geqq T$, the situation is the same as in the continuous case but if $p(L) = 1 - \rho^L$ with some $0 < \rho < 1$, then the question of stability turns to be solvable.

Let us say we are thinning a Poisson variable if we represent it with balls and kill independently the balls with a certain probability. It is a well

known fact that the thinning of a Poisson variable results in a Poisson variable again. Let Z be a Poisson variable with parameter λ and let it be thinned independently into random variables X_1 and Y_1 with probabilities p and q accordingly. Let the random variables X_2 and Y_2 be Poisson with parameters $(1-p)\lambda$ and $(1-q)\lambda$ and independent of the earlier random variables. The joint distribution of

$$X = X_1 + X_2, \quad Y = Y_1 + Y_2$$

is somewhat cumbersome:

$$P(X = x, Y = y) = \sum_{z=0}^{\infty} \mathrm{Pois}\,(z, \lambda) \Big[\sum_{i=0}^{z} \mathrm{Bin}\,(z, p, i)\, \mathrm{Pois}(x - i, (1-p)\lambda) \Big]$$

$$\times \Big[\sum_{j=0}^{z} \mathrm{Bin}\,(z, q, j)\, \mathrm{Pois}(y - j, (1-q)\lambda) \Big].$$

but its generating function is easily found. This observation is the driving force in our calculations on the conditional probabilities for pairs of relatives.

In Section 2 we present the working model, in Section 3 we prove the main theorem, in Section 4 we develop the conditional probabilities for the malformation in the relatives of an affected person. In Section 5 the theory is applied on the Hungarian data, and in Section 6 the conclusions are drawn.

2. The working model

We consider a population with sexual reproduction, selection, synchronous generations on a short time frame in the evolutionary sense. We assume all relevant loci have the same effect in view of the birth defect, so the only thing we keep track of is the number of mutant genes one has. To get the genetic information of offsprings, we need recombination, mutation, and selection.

During recombination we assume crossovers may happen, and there is a low number of mutant genes, that is, each of them is inherited independently with probability $1/2$. If the two parents have x and y mutant genes, the child will receive a random number from the $\mathrm{Binom}\,(x + y, 1/2)$ distribution.

The child is affected by additional mutation, this is represented by adding an independent $\mathrm{Poisson}\,(\mu)$ random variable to the inherited mutant gene count.

Given the number of mutant genes the child has, we have to find out two things: whether he/she is affected by the disorder and whether he/she is fertile (and viable). We assume each mutant gene may cause the disorder to appear or the loss of fertility. There is an ordering of the two symptoms,

a gene causing the loss of fertility also causes the disorder to appear. The probability of a single gene *not* causing the disorder is denoted by Δ, and the probability of *not* inhibiting fertility is ρ. Clearly $\rho > \Delta$. Once again, each gene has a random effect on the individual in the following way:

- with probability Δ it has no effect,
- with probability $\rho - \Delta$ it causes the individual to be affected by the disorder, but has no effect on fertility,
- with probability $1 - \rho$ it causes the individual to be affected by the disorder and lose fertility.

We need to easily refer to the combination of these operations. For a pair of distribution of mutant genes (P_f, P_m), let us denote the female distribution of the next generation by $T_f(P_f, P_m)$. We use the analogous notation for the male counterpart. We vaguely use $T_f^k(P_f, P_m)$ for the female distribution after k generations (although we should use $T_f\big(T_f(P_f, P_m), T_m(P_f, P_m)\big)$ instead of $T_f^2(P_f, P_m)$).

3. Stationary genotype distribution

This section deals with the long-term behavior of the genotype distribution. It is rather clear that if there is no selection which has the role of filtering out the mutant genes, then their number will grow unboundedly. Consequently, to have a chance of stationarity, we need $\rho < 1$. We claim that in this case the distribution of mutant genes in the population stabilizes over time. We assume there is a separate set of parameters for females $(\mu_f, \rho_f, \Delta_f)$ and males $(\mu_m, \rho_m, \Delta_m)$. We do not see biological evidence for μ_f and μ_m to differ but it does no harm to include it in our study, and we get a more general result.

THEOREM 1. *If $\rho_f, \rho_m < 1$ then for any pair P_f, P_m of initial distributions of mutant genes, the distributions of $T_f^k(P_f, P_m)$, $T_m^k(P_f, P_m)$ will converge in distribution to a pair of limiting Poisson distributions with parameters*

$$\lambda_f = \frac{\rho_f\rho_m(\mu_m - \mu_f) + 2\rho_f\mu_f}{2 - \rho_f - \rho_m}, \qquad \lambda_m = \frac{\rho_f\rho_m(\mu_f - \mu_m) + 2\rho_m\mu_m}{2 - \rho_f - \rho_m},$$

for females and males, respectively, when $k \to \infty$.

PROOF. We work with generating functions. We say that $P = (p_i)_{i=0}^{\infty}$ is a probability distribution on $\mathbb{N}$ if $p_i \geqq 0$ and $\sum_{i=0}^{\infty} p_i = 1$. Denote by $\mathcal{P}$ the set of probability distributions on $\mathbb{N}$. For $P \in \mathcal{P}$ and $x \in [0, 1]$ let us define

$$G_P(x) = \sum_{i=0}^{\infty} p_i x^i.$$

The coefficients of the power series form a probability distribution, consequently $G_P(x)$ is analytic on $[0,1]$. The operations used in our model are easy to handle with generating functions. We write out the equations for a daughter, we get the analogous equations for a son by exchanging f and m in the indices.

Convolution of distributions is reflected as multiplication of the generating functions, so adding up parental mutant genes translates to

$$G_{P'}(x) = G_{P_f}(x)G_{P_m}(x).$$

Plugging the value of the variable into a binomial distribution with parameter $1/2$ (also known as "thinning") translates to changing the argument from x to $(1+x)/2$. We get

$$G_{P''}(x) = G_{P'}\left(\frac{1+x}{2}\right).$$

Adding external mutation is another multiplication with the generating function of a Poisson variable with parameter μ_f:

$$G_{P'''}(x) = G_{P''}(x)e^{\mu_f(x-1)}.$$

During selection, we put weights on each p'''_i, then normalize to obtain a probability distribution in the following fashion: the probability of having i mutant genes is p'''_i, and the probability that a female with i mutant genes remains fertile is ρ_f^i, thus a female in the community of fertile females will have i mutant genes with probability $p'''_i\rho_f^i/\sum_{j=0}^{\infty} p'''_j\rho_f^j$. This operation is known as the "exponential tilting" of the distribution P'''. For generating functions, the effect of selection can be computed the following way:

$$G_{P''''}(x) = \sum_{i=0}^{\infty} \frac{p'''_i\rho_f^i}{\sum_{j=0}^{\infty} p'''_j\rho_f^j}x^i = \frac{G_{P'''}(\rho_f x)}{G_{P'''}(\rho_f)}.$$

Composing the three transformations we get

$$G_{T_f(P_f,P_m)}(x) = \frac{G_{P_f}\big((1+\rho_f x)/2\big)\,G_{P_m}\big((1+\rho_f x)/2\big)}{G_{P_f}\big((1+\rho_f)/2\big)\,G_{P_m}\big((1+\rho_f)/2\big)}e^{\mu_f\rho_f(x-1)}. \tag{1}$$

We want to iterate T n times. Naturally we want to avoid writing down all these complicated formulas. In order to see the structure of what we get, let us write down the formula for T^2, but without arguments:

$$G_{T_f^2}(x) = \frac{G_{T_f()}()G_{T_m()}()}{G_{T_f()}()G_{T_m()}()}e^{\cdots} = \frac{\frac{G_{P_f}()G_{P_m}()}{G_{P_f}()G_{P_m}()}e^{\cdots}\frac{G_{P_f}()G_{P_m}()}{G_{P_f}()G_{P_m}()}e^{\cdots}}{\frac{G_{P_f}()G_{P_m}()}{G_{P_f}()G_{P_m}()}e^{\cdots}\frac{G_{P_f}()G_{P_m}()}{G_{P_f}()G_{P_m}()}e^{\cdots}}e^{\cdots}. \tag{2}$$

From (1) we see that the denominator of $G_{T_f(P_f,P_m)}$ is constant in x and the constant is the normalizing factor which guarantees that $G_{P'''}(1) = 1$. Rearranging (2) we end up with a formula that is the product of four $G()/G()$ terms (where the denominator normalizes the numerator and the ratio takes value 1 for $x = 1$) and an exponential term. After n iterations we get that $G_{T_f^n}(x)$ is a product of the functions $\hat{G}_{T_f^n}(x)$ and $E_f^n(x)$, where $\hat{G}_{T_f^n}(x)$ is a product of 2^n terms of the form $G()/G()$ and $E_f^n(x)$ is an exponential term (the generating function of some Poisson random variable).

Let us treat $\hat{G}_{T_f^n}(x)$ and $E_f^n(x)$ separately.

We first show that $\hat{G}_{T_f^n}(x) \to 1$ for all $x \in [0,1]$ as $k \to \infty$.

If we put back the arguments in one of the 2^n terms of $\hat{G}_{T_f^n}(x)$, we see that it is of the form

$$\frac{G\big(B(x)\big)}{G\big(B(1)\big)},$$

where B is an affine function, an n-fold composition of either $x \mapsto (1+\rho_f x)/2$ or $x \mapsto (1+\rho_m x)/2$, and the generating function G is either G_{P_f} or G_{P_m}. The product of all these terms look like

(3)
$$\frac{G\big(B(x)\big)}{G\big(B(1)\big)} \cdot \ldots \cdot \frac{G\big(B(x)\big)}{G\big(B(1)\big)} = \exp\left(\log \frac{G\big(B(x)\big)}{G\big(B(1)\big)} + \cdots + \log \frac{G\big(B(x)\big)}{G\big(B(1)\big)}\right),$$

with G and B changing throughout the formula. Let us make sure the use of logarithms is feasible. It is easy to see that $B(x) > 0$ for $x \geqq 0$. The generating function G is a power series with non-negative (and at least one positive) coefficients, so $G\big(B(x)\big) > 0$ for $x \in [0,1]$. Now we have to estimate the terms of the form $\log\big(G\big(B(x)\big)/G\big(B(1)\big)\big)$. Denote $\rho_* = \max(\rho_f, \rho_m) < 1$. By the mean value theorem, for every $x \in [0,1]$ there is a $\xi \in \big[B(x), B(1)\big] \subseteqq \big[1/2, (1+\rho_*)/2\big]$ such that

$$\log \frac{G\big(B(x)\big)}{G\big(B(1)\big)} = \log G\big(B(x)\big) - \log G\big(B(1)\big) = \big(B(x) - B(1)\big)(\log G)'(\xi).$$

The coefficient of x in $B(x)$ will be at most $(\rho_*/2)^n$. Thus for any $x \in [0,1]$ we get

$$\big|B(x) - B(1)\big| \leqq \left(\frac{\rho_*}{2}\right)^n.$$

The function G is continuously differentiable and bounded away from 0 on the interval $\xi \in \big[B(x), B(1)\big] \subseteqq \big[1/2, (1+\rho_*)/2\big]$, consequently the deriva-

tive of the logarithm can be bounded in absolute value by some C. In the end we get

$$\left|\log \frac{G(B(x))}{G(B(1))}\right| < C\left(\frac{\rho_*}{2}\right)^n.$$

Adding up 2^n of such terms gives the bound

$$\left|\log \frac{G(B(x))}{G(B(1))} + \cdots + \log \frac{G(B(x))}{G(B(1))}\right| \leqq C\rho_*^n.$$

This tends to 0 for all $x \in [0,1]$, thus the product on the left-hand side of (3) converges to 1 as $n \to \infty$. Observe that the exponential term in $E_f^n(x)$ does not depend on the initial distributions P_f, P_m. Thus we have just shown that the only part depending on the initial distributions vanishes. Consequently the convergence and the potential limit does not depend on the initial distributions.

It is now enough to show a pair of distributions satisfying

$$(P_f, P_m) = \left(T_f(P_f, P_m), T_f(P_f, P_m)\right),$$

as the previous reasoning ensures that the trivial convergence of this case implies convergence for any initial generating functions to this fixed point. We search among Poisson distributions because this family is closed for all the transformations we use. The pair (λ_f, λ_m) is invariant exactly when

$$\lambda_f = \left(\frac{\lambda_f + \lambda_m}{2} + \mu_f\right)\rho_f, \qquad \lambda_m = \left(\frac{\lambda_f + \lambda_m}{2} + \mu_m\right)\rho_m.$$

Taking the average of the two equations results in a simple expression for $(\lambda_f + \lambda_m)/2$, plugging it back gives us the parameters stated in the theorem.

To conclude we use the fact that the convergence of a sequence of generating functions to a generating function on $[0,1]$ implies the convergence of the corresponding probability distributions (see e.g. Mukherjea et al. [5]). □

We should note that the proof strongly relies on the specific choice of selection which we can conveniently handle using generating functions. As we mentioned in the introduction, it makes sense to consider different functions determining the risk based on the mutant gene count. However, it is unclear how one should modify the proof to resolve the alternative cases.

4. Theoretical disorder probabilities

From the previous section we learn that it makes sense to assume the population to be in the stationary state. It is easy to check that the num-

ber of mutant genes a newborn has follows a Poisson distribution with the following parameters depending on the gender:

$$\frac{\lambda_f+\lambda_m}{2}+\mu_f=\frac{\lambda_f}{\rho_f}, \qquad \frac{\lambda_f+\lambda_m}{2}+\mu_m=\frac{\lambda_m}{\rho_m}.$$

Consequently his/her probability of being healthy is

$$p_f=\exp\left(\lambda_f\frac{\Delta_f-1}{\rho_f}\right), \qquad p_m=\exp\left(\lambda_m\frac{\Delta_m-1}{\rho_m}\right).$$

Similarly, the probability of being fertile is

$$\tilde{p}_f=\exp\left(\lambda_f\frac{\rho_f-1}{\rho_f}\right), \qquad \tilde{p}_m=\exp\left(\lambda_m\frac{\rho_m-1}{\rho_m}\right).$$

However, if we look at a family tree at once, we see a complex multidimensional joint distribution. We want to answer simple questions like "What is the (conditional) probability of an aunt of a malformed child being affected?".

We claim that we can get a closed form expression on any reasonable conditional probabilities like above. The resulting formulas often become enormous, but there is a way to derive them with reasonable effort.

We would like a general iterative computational scheme that can be used for most cases. The idea is to draw a graph of the family tree, transform it to simpler graphs while building the formula for the probability.

We include the possible dependence on the gender of the patient. Therefore the parameters we have are

$$\mu_f,\mu_m,\rho_f,\rho_m,\Delta_f,\Delta_m.$$

The parameters of the stationary distributions are

$$\lambda_f=\frac{\rho_f\rho_m(\mu_m-\mu_f)+2\rho_f\mu_f}{2-\rho_f-\rho_m}, \qquad \lambda_m=\frac{\rho_f\rho_m(\mu_f-\mu_m)+2\rho_m\mu_m}{2-\rho_f-\rho_m}.$$

From now on to reduce the number of formulas, we use $x,y,\dots$ for one gender or another, thus μ_x or λ_y is the parameter corresponding to the appropriate gender. In addition we use x' for the gender different from x.

4.1. Representing graphs. First, let us visualize the situation. We may draw a family tree with some additional information.

We use Fig. 1 as an example. Suppose $x=m$, $y=m$, $z=f$ for a moment. The circles in the graph represent members or couples of the family.

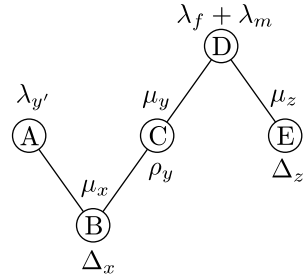


Fig. 1: Healthy boy and aunt (or similar)

In this case B is the male patient we start at, A is the mother, C is the father. D represents the paternal grandparents together. We do not separate them as we use only the joint genetic information of them. The last member E is an aunt.

The genetic information moves in the following way. Each line represents a parental relation, so each gene is inherited downwards independently with probability $1/2$. The values above the circles show where additional mutant genes enter the system. We always mean a Poisson random variable with the parameter being the value indicated. These are obviously μ_u for most people, and λ_u or $\lambda_f + \lambda_m$ for the people or couples we start with.

The event we want to investigate is coded in the values below the circles. They show a per-gene probability for mutant genes that the actual person complies with the event. In the figure above we have Δ_u in two positions which means we want the patient and the aunt (or uncle) to be healthy. The ρ_y under C is an implied restriction, as we need the father (or mother) to be fertile for the graph to be valid. Some places have no value indicated, we have no restriction there, we may also write 1 to these places.

This way we can only express events requiring some to be healthy, some to be fertile, but these are the ones that are easy to directly compute. By basic inclusion-exclusion formulas we can also handle events about some being affected or infertile. To compute conditional probabilities we simply need to divide two of such probabilities.

Now let us get into computational details to work through our plan.

4.2. Processing graphs. We can handle the simplest graph possible:

Fig. 2: Basic graph

The probability of the event described by this basic graph is

$$\sum_{i=0}^{\infty} \frac{\eta^i}{i!} e^{-\eta} \alpha^i = \exp\left(\eta(\alpha - 1)\right).$$

We introduce a few graph operations so we can transform complex graphs into simpler ones. Observe that if a final descendant receives mutant genes from multiple sources, they pose independent threats, so we can split the graph as pictured below.

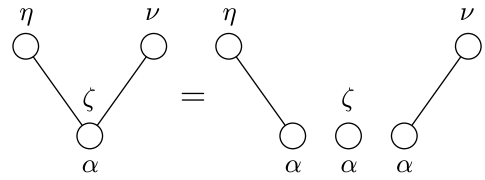


Fig. 3: Splitting a graph

The other operation we use is to merge a child to the parent. Consider the following setting:

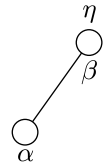


Fig. 4: Parent and child

We condition on the number of mutant genes the parent has, suppose it is c. Then the distribution of mutant genes the child inherits follows a $\mathrm{Binom}\,(c, 1/2)$ distribution. So the probability that the child behaves according to the event is

$$\sum_{i=0}^{c} \binom{c}{i} \left(\frac{1}{2}\right)^c \alpha^i = \left(\frac{1+\alpha}{2}\right)^c.$$

This is an exponential term in c, so we do not change the overall probability of the event if we omit the child but multiply the risk factor of the parent by $(\alpha + 1)/2$.

It is easy to see that any acyclic family tree can be reduced to contain only a few copies of the simplest one-node graph.

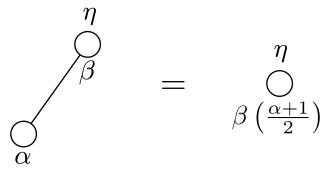


Fig. 5: Merging a child

4.3. Siblings. Let us start with the simplest case, computing conditional probabilities for first order relatives. We want to find out the conditional probability of a sibling of a malformed child being affected. Fig. 6 shows the graph for the sibling.

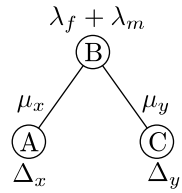


Fig. 6: Healthy patient and sibling

Let us use the notation scheme $p_{\bar{A}C}$, this stands for the probability of A being affected by the risk and C not (and we do not count on others). This means the conditional probability q_S we need is

$$q_S = \frac{p_{\bar{A}\bar{C}}}{p_{\bar{A}}}.$$

Using inclusion-exclusion formulas we have

$$p_{\bar{A}\bar{C}} = 1 - p_A - p_C + p_{AC}, \quad p_{\bar{A}} = 1 - p_A.$$

The method in the previous section allows us to compute these probabilities. When computing p_A, we replace the risk of C by 1. The graph decomposition is shown in Fig. 7. By symmetry we can calculate p_C analogously. We show the graph decomposition for computing p_{AC} in Fig. 8.

We do not aim for the simplest expressions, we rather leave it in a form that is easier to check.

$$p_A = \exp\left(\left(\mu_x + \frac{\lambda_f + \lambda_m}{2}\right)(\Delta_x - 1)\right),$$

$$p_C = \exp\left(\left(\mu_y + \frac{\lambda_f + \lambda_m}{2}\right)(\Delta_y - 1)\right),$$

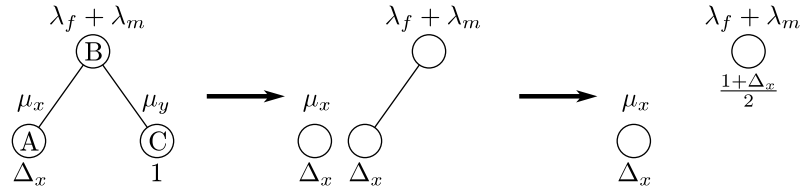


Fig. 7: Graph decomposition to compute p_A

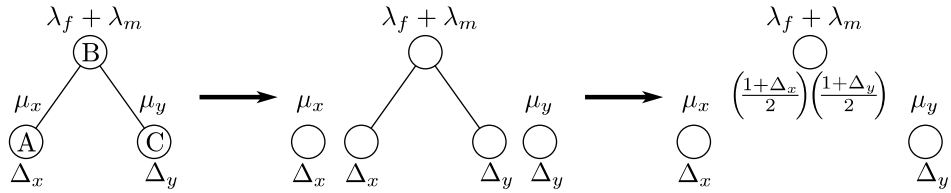


Fig. 8: Graph decomposition to compute p_{AC}

$$p_{AC} = \exp\left(\mu_x(\Delta_x - 1) + \mu_y(\Delta_y - 1) + \frac{\lambda_m + \lambda_f}{4}\big((\Delta_x + 1)(\Delta_y + 1) - 4\big)\right).$$

In case of complete selection and symmetric gender roles, i.e.

$$\Delta_m = \Delta_f = \rho_m = \rho_f = \rho, \quad \lambda_f = \lambda_m = \lambda, \quad \text{and} \quad \mu_m = \mu_f = \mu,$$

the conditional probability q_S is

$$q_S = 2 - \frac{1 - e^{-t}}{1 - e^{-\mu}},$$

where

$$t = 2\mu\big(1 - \frac{1}{4}\rho(1 - \rho)\big),$$

and $\rho = \frac{\lambda}{\lambda+\mu}$. Surprisingly q_s depends on ρ through the term $\rho(1-\rho)$. In this case the population prevalence simplifies to

$$p_{\bar{A}} = 1 - \exp\big((\lambda + \mu)(\rho - 1)\big) = 1 - \exp\big(\lambda - (\lambda + \mu)\big) = 1 - \exp(-\mu).$$

thus ρ is a free parameter and q_S is a symmetric function of ρ regarding the swap $\tilde{\rho} = 1 - \rho$. We are curious whether there is a direct explanation for this symmetry. When μ is small and $\rho = \frac{1}{2}$, then $\lambda = \mu$ and a bad gene is rare. An affected child gets a bad gene fifty-fifty either from mutation or from one of his/her parents. In the second case the sibling gets the bad gene from the affected parent with half probability and the bad gene is expressed again with probability half. Accordingly q_S is close to $\frac{1}{8}$. We shall refer to this parametrization as *standard model.*

4.4. Parent. Next we calculate the conditional probability for a parent being affected, which is also fairly simple. See Fig. 9 for the describing graph. The only novelty is the Δ_y/ρ_y risk of the parent. It is easy to see that this is the risk of not being affected by the disorder conditioned on being fertile.

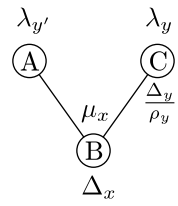


Fig. 9: Healthy patient and parent

$$q_P = \frac{p_{\bar{B}\bar{C}}}{p_{\bar{B}}} = \frac{1 - p_B - p_C + p_{BC}}{1 - p_B},$$

$$p_B = \exp\left(\left(\mu_x + \frac{\lambda_f + \lambda_m}{2}\right)(\Delta_x - 1)\right),$$

$$p_C = \exp\left(\lambda_y\left(\frac{\Delta_y}{\rho_y} - 1\right)\right),$$

$$p_{BC} = \exp\left(\left(\mu_x + \frac{\lambda_{y'}}{2}\right)(\Delta_x - 1) + \lambda_y\left(\frac{\Delta_y}{\rho_y}\left(\frac{\Delta_x + 1}{2}\right) - 1\right)\right).$$

In the standard model, when $\rho_f = \rho_m = \Delta_f = \Delta_m = 1/2$ and $\mu_f = \mu_m$ is small, we get $q_P = 0$. This is rather clear because this special case implies complete selection.

4.5. Grandparent. Let us move on to higher order relatives, starting with grandparents. Fig. 10 shows the actual graph to be processed. The conditional probability can be expressed as

$$q_G = \frac{p_{\bar{B}C\bar{D}}}{p_{\bar{B}C}} = \frac{p_C - p_{BC} - p_{CD} + p_{BCD}}{p_C - p_{BC}},$$

$$p_C = \exp\left(\left(\mu_y + \frac{\lambda_f + \lambda_m}{2}\right)(\rho_y - 1)\right),$$

$$p_{BC} = \exp\left(\left(\mu_x + \frac{\lambda_{y'}}{2}\right)(\Delta_x - 1) + \left(\mu_y + \frac{\lambda_f + \lambda_m}{2}\right)\left(\rho_y\frac{\Delta_x + 1}{2} - 1\right)\right),$$

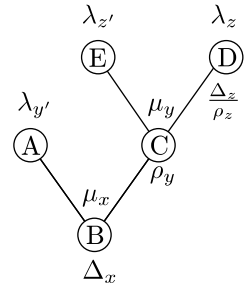


Fig. 10: Healthy patient and grandparent

$$p_{CD} = \exp\left(\left(\mu_y + \frac{\lambda_{z'}}{2}\right)(\rho_y - 1) + \lambda_z\left(\frac{\Delta_z}{\rho_z}\left(\frac{\rho_y + 1}{2}\right) - 1\right)\right),$$

$$p_{BCD} = \exp\left(\left(\mu_x + \frac{\lambda_{y'}}{2}\right)(\Delta_x - 1) + \left(\mu_y + \frac{\lambda_{z'}}{2}\right)\left(\rho_y\frac{\Delta_x + 1}{2} - 1\right)\right.$$

$$\left. + \lambda_z\left(\frac{\Delta_z}{\rho_z}\left(\frac{\rho_y\frac{\Delta_x+1}{2} + 1}{2}\right) - 1\right)\right).$$

In the standard model we get $q_G = 0$ as we expect because of the complete selection.

4.6. Aunt and uncle. Let us turn to investigating aunts and uncles. We use Fig. 1 for the calculation. The conditional probability can be expressed as

$$q_A = \frac{p_{\bar{B}C\bar{E}}}{p_{\bar{B}C}} = \frac{p_C - p_{BC} - p_{CE} + p_{BCE}}{p_C - p_{BC}}.$$

We can compute the occurring probabilities as before. Without going into details, we get

$$p_C = \exp\left(\left(\mu_y + \frac{\lambda_f + \lambda_m}{2}\right)(\rho_y - 1)\right),$$

$$p_{BC} = \exp\left(\left(\frac{\lambda_{y'}}{2} + \mu_x\right)(\Delta_x - 1) + \left(\mu_y + \frac{\lambda_f + \lambda_m}{2}\right)\left(\rho_y\frac{\Delta_x + 1}{2} - 1\right)\right),$$

$$p_{CE} = \exp\left(\mu_y(\rho_y - 1) + \mu_z(\Delta_z - 1)\right.$$

$$\left. + (\lambda_f + \lambda_m)\left(\frac{(\rho_y + 1)(\Delta_z + 1)}{4} - 1\right)\right),$$

$$p_{BCE} = \exp\left(\mu_z(\Delta_z - 1) + \left(\mu_x + \frac{\lambda_{y'}}{2}\right)(\Delta_x - 1) + \mu_y\left(\rho_y \frac{\Delta_x + 1}{2} - 1\right)\right.$$

$$\left. + \frac{\lambda_f + \lambda_m}{4}\left(\left(\rho_y \frac{\Delta_x + 1}{2} + 1\right)(\Delta_z + 1) - 4\right)\right).$$

Plugging these back gives us the conditional probability we were looking for.

In the standard model the number of halving factors is 5:

– the affected child might get the bad gene by mutation,

– or by the parent out of link to aunt-uncle,

– the parent in the link to aunt-uncle might get the bad gene by mutation,

– the grandparents need not to pass it to another child,

– who needs not to express the malformation.

We get $q_A = 1/32$ as well by using the expressions above for the standard model.

4.7. Cousin. To compute the analogous conditional probability for cousins, we will use Fig. 11 below.

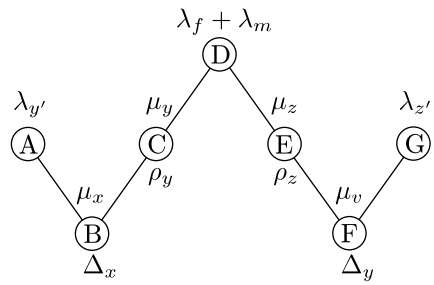


Fig. 11: Healthy patient and cousin

Using the same method, we want to compute

$$q_C = \frac{p_{\bar{B}CE\bar{F}}}{p_{\bar{B}CE}} = \frac{p_{CE} - p_{BCE} - p_{CEF} + p_{BCEF}}{p_{CE} - p_{BCE}} = 1 - \frac{p_{CEF} - p_{BCEF}}{p_{CE} - p_{BCE}}.$$

For the individual probabilities in this setting we get

$$p_{CE} = \exp\left(\mu_y(\rho_y - 1) + \mu_z(\rho_z - 1) + \frac{\lambda_f + \lambda_m}{4}\big((\rho_y + 1)(\rho_z + 1) - 4\big)\right),$$

$$p_{BCE} = \exp\left(\left(\mu_x + \frac{\lambda_{y'}}{2}\right)(\Delta_x - 1) + \mu_z(\rho_z - 1) + \mu_y\left(\rho_y\frac{\Delta_x + 1}{2} - 1\right)\right.$$
$$\left. + \frac{\lambda_m + \lambda_f}{4}\left(\left(\rho_y\frac{\Delta_x + 1}{2} + 1\right)(\rho_z + 1) - 4\right)\right),$$

$$p_{CEF} = \exp\left(\left(\mu_v + \frac{\lambda_{z'}}{2}\right)(\Delta_v - 1) + \mu_y(\rho_y - 1) + \mu_z\left(\rho_z\frac{\Delta_v + 1}{2} - 1\right)\right.$$
$$\left. + \frac{\lambda_m + \lambda_f}{4}\left(\left(\rho_z\frac{\Delta_v + 1}{2} + 1\right)(\rho_y + 1) - 4\right)\right),$$

$$p_{BCEF} = \exp\left(\left(\mu_x + \frac{\lambda_{y'}}{2}\right)(\Delta_x - 1) + \left(\mu_v + \frac{\lambda_{z'}}{2}\right)(\Delta_v - 1)\right.$$
$$+ \mu_y\left(\rho_y\frac{\Delta_x + 1}{2} - 1\right) + \mu_z\left(\rho_z\frac{\Delta_v + 1}{2} - 1\right)$$
$$\left. + \frac{\lambda_f + \lambda_m}{4}\left(\left(\rho_y\frac{\Delta_x + 1}{2} + 1\right)\left(\rho_z\frac{\Delta_v + 1}{2} + 1\right) - 4\right)\right).$$

These are rather cumbersome formulas, but in the standard model, we get $q_C = 1/124$. At first this is a bit surprising, because by counting the number of halving factors as before, we get $1/2^7 = 1/128$. We should note that checking a cousin for the disorder implies he is already born, that is, his parents are fertile. Conditioning on this accounts for a division by $31/32$ which brings us to the correct value.

5. Validation of the model

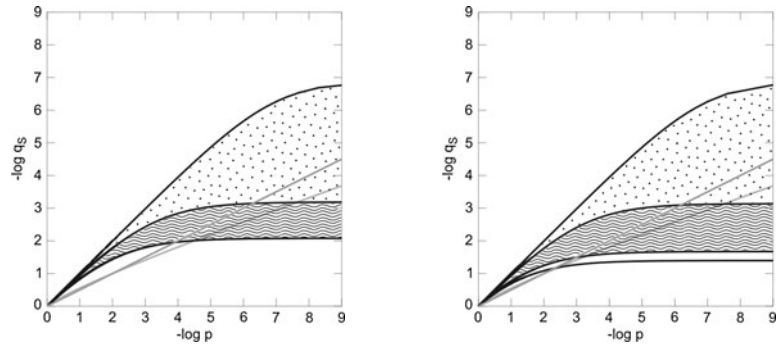
It is an important milestone to have a model which we can handle, we still have to check how well does it follow biological principles and how does it fit the population. Let us recall the notations introduced in Section 4:

$$p = P(\text{subject is affected}), \quad q_S = P(\text{sibling is affected} \mid \text{subject is affected}).$$

The initial requirement for a model with inheritance is to have high conditional probabilities for first order relatives, in other words $q_S \gg p$. To test

this, we will try to choose the parameters to increase q_S as much as possible within the given constraints. Another guideline we use is a fundamental approximation on multifactorial disorders given by the Edwards formula [2] which states that $q_S \approx \sqrt{p}$.

We do not want to go into theoretical details, let us just present Fig. 12 showing the relation between $\log p$ and $\log q_S$ for $\mu \in [0.005, 1]$ and $\rho \in [0.5, 1)$. On the left side, we assume complete selection, that is, $\rho = \Delta$, on the right side we consider a partial selection with $\rho = (1 + \Delta)/2$.



(a) Complete selection (b) Partial selection

Fig. 12: Model probabilities and the Edwards formula

The upper diagonal line shows where the Edwards formula is precisely satisfied, the lower one corresponds to probabilities of the Gaussian model used by Czeizel and Tusnády in [1]. We prefer parameters where the disorder is mainly inherited, that is, $\lambda \gg \mu$. Thus we split the domain the model sweeps through into three regions, the values we can reach while $\lambda \geqq 10\mu$, or just $10\mu > \lambda \geqq \mu$, or only $\mu \geqq \lambda$ (top to bottom). Although the model does not satisfy the formula in general, we may choose the parameters to do so.

Here is another way of comparing with the Gaussian model. Let the population frequency of a malformation be 0.00071 for males and 0.00317 for females, suppose there is no selection. The following table shows the conditional probabilities of the malformation in the first, second and third degree relatives in the Gaussian model. The rows correspond to different genders of the malformed child, columns represent the degree of relationship and the gender of corresponding member of the family.

It is a remarkable property of the multifactorial threshold model that a relative with the gender of larger frequency of a malformed child with the gender of smaller frequency has the maximal conditional probability. The reason for the property is that the malformed child with smaller frequency has larger liability shifting the liability of his family upwards. The relatives

	I M	I F	II M	II F	III M	III F
M	0.0393	0.1149	0.0076	0.0276	0.0024	0.0195
F	0.0232	0.0749	0.0055	0.0204	0.0014	0.0087

Table 1: Conditional probabilities in the Gaussian model without selection

with gender of larger frequency is evaluated with a smaller threshold which results in the mentioned property. The following table gives the conditional probabilities for the case with complete selection for the Gaussian model.

	I M	I F	II M	II F	III M	III F
M	0.0365	0.1025	0.0085	0.0255	0.0032	0.0113
F	0.0242	0.0739	0.0063	0.0234	0.0020	0.0088

Table 2: Conditional probabilities in the Gaussian model with complete selection

We compare these values with those coming from the Poisson model. We assume $\mu_f = \mu_m$, and use the remaining degree of freedom to get the highest conditional probabilities as mentioned in the beginning of this section. Having no selection means $\rho_f = \rho_m = 1$ but in this case we cannot apply Theorem 1. We rather choose $\rho_f = \rho_m = 1 - \varepsilon$ for some small $\varepsilon > 0$ to allow only negligible selection, but stay within the conditions of Theorem 1.

	I M	I F	II M	II F	III M	III F
M	0.1124	0.5015	0.0566	0.2523	0.1475	0.1722
F	0.1123	0.5012	0.0565	0.2521	0.1473	0.1721

Table 3: Conditional probabilities in the Poisson model with negligible selection

With complete selection:

	I M	I F	II M	II F	III M	III F
M	0.0452	0.2017	0.0135	0.0603	0.0342	0.0418
F	0.0452	0.2015	0.0135	0.0602	0.0342	0.0418

Table 4: Conditional probabilities in the Poisson model with complete selection

The reassuring fact we see is that we can set the conditional probabilities even higher than in the Gaussian model while leaving population probabilities unchanged.

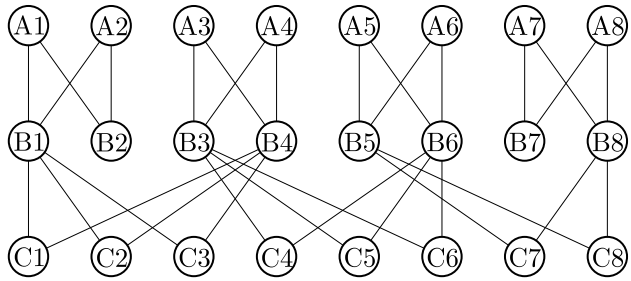


Fig. 13: Model family

Next, we perform a Monte Carlo simulation on a model family given in Fig. 13. We fix that A2, A4, A6, A8, B6 are women, A1, A3, A5, A7, B3 are men. The following numbers in Table 5 are probabilities conditioned on C5 having the malformation. We generated a large number of families starting from A1–A8 and only selected those where C5 was born and had the malformation. This explains the zeros in the first lines as they are all parents and consequently they are healthy. This does not hold for B1 as we allow him/her to be infertile thus C1 might not be born.

Gender of relative	Gender of index	A1	A2	A3	A4	A5	A6	A7	A8
M	M	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
M	F	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
F	M	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
F	F	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
		B1	B2	B3	B4	B5	B6	B7	B8
M	M	0.00100	0.00071	0.00000	0.01853	0.00752	0.00000	0.00077	0.00077
M	F	0.00079	0.00064	0.00000	0.01890	0.00887	0.00000	0.00069	0.00068
F	M	0.00295	0.00305	0.00000	0.08952	0.03948	0.00000	0.00288	0.00293
F	F	0.00334	0.00290	0.00000	0.08458	0.03906	0.00000	0.00306	0.00310
		C1	C2	C3	C4	C5	C6	C7	C8
M	M	0.00701	0.00684	0.00813	0.04489	1.00000	0.04284	0.00221	0.00290
M	F	0.00706	0.00657	0.00717	0.04408	0.00000	0.04392	0.00310	0.00312
F	M	0.03072	0.03130	0.03076	0.18805	0.00000	0.19658	0.01346	0.01445
F	F	0.02944	0.02999	0.02988	0.19319	1.00000	0.19074	0.01485	0.01411

Table 5: Conditional probabilities in the Poisson model with complete selection

The gender of the affected child has seemingly no effect beyond randomness. One explanation for this phenomena is that in case of rare malformations the only effect that the affected child might cause is that he/she has a bad gene which is independent of gender differences. Using this setup also allows us to numerically compute more elaborate conditional and joint probabilities.

table no	page	type of ICCA	Sex	pop freq 1000	No	Father			Mother			Brother			Sister		
						m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$
20	71	ASB	Boy	2.22	134	134	0	*0.0*	134	0	*0.0*	102	1	*2.3*	86	2	*2.8*
			Girl	3.59	309	309	0	*0.0*	309	0	*0.0*	210	5	*4.0*	177	4	*5.3*
30	96	CL(P)	Boy	1.33	369	304	3	*3.6*	304	5	*6.9*	143	16	*5.7*	121	0	*3.9*
			Girl	0.77	200	161	3	*2.4*	166	7	*4.9*	80	1	*4.4*	89	4	*4.1*
35	114	CHPS	Boy	2.18	112	112	1	*0.5*	112	2	*1.4*	48	2	*2.0*	38	2	*1.1*
			Girl	0.69	36	36	0	*0.5*	36	1	*1.4*	15	3	*1.3*	10	0	*0.8*
43	134	VSD	Boy	1.72	180	180	1	*0.4*	180	*2*	*1.8*	121	0	*2.6*	109	2	*2.1*
			Girl	1.41	197	197	0	*0.5*	197	3	*2.3*	96	2	*2.2*	81	1	*1.6*
56	166	CDH-BB	Boy	11.96	422	422	7	*6.3*	422	20	*32.2*	125	20	*22.6*	126	20	*32.9*
			Girl	39	1345	1345	9	*6.4*	1345	46	*32.7*	421	29	*34.3*	398	79	*72.4*
57	170	CDH-CB	Boy	8.13	75	75	1	*0.4*	75	2	*7.3*	22	2	*4.8*	21	7	*7.0*
			Girl	50.57	304	304	0	*0.4*	304	13	*7.3*	89	6	*6.8*	75	14	*17.0*
69	195	STEV	Boy	1.65	118	118	4	*2.7*	118	3	*1.5*	61	4	*2.6*	60	1	*2.2*
			Girl	0.82	56	56	1	*2.5*	56	0	*1.3*	29	2	*2.0*	30	3	*1.9*

		Paternal uncle			Paternal aunt			Maternal uncle			Maternal aunt			Paternal cousins Boy			Paternal cousins Girl			Maternal cousins Boy			Maternal cousins Girl		
		m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$
ASB	Boy	306	0	*4.5*	260	1	*6.6*	290	1	*4.3*	302	0	*7.6*	387	0	*5.7*	447	0	*11.3*	283	0	*4.2*	307	1	*7.8*
	Girl	505	1	*7.4*	512	0	*12.9*	497	3	*7.3*	534	0	*13.5*	850	2	*12.6*	785	2	*19.9*	606	3	*9.0*	621	3	*15.8*
CL(P)	Boy	283	3	*3.6*	307	3	*2.3*	383	0	*5.5*	366	2	*3.3*	362	1	*4.6*	374	0	*2.8*	365	1	*4.7*	387	2	*2.9*
	Girl	163	1	*2.1*	198	1	*1.5*	172	3	*2.7*	200	2	*2.0*	174	0	*2.2*	219	1	*1.6*	156	2	*2.0*	166	0	*1.3*
CHPS	Boy	153	0	*3.1*	146	0	*1.0*	180	1	*3.6*	166	1	*1.2*	130	2	*2.6*	136	0	*0.9*	154	1	*3.0*	144	1	*1.0*
	Girl	42	0	*0.9*	33	0	*0.3*	42	0	*0.9*	39	0	*0.4*	31	0	*0.6*	28	0	*0.2*	45	1	*0.9*	31	0	*0.2*
VSD	Boy	272	3	*4.3*	225	5	*3.0*	214	3	*3.4*	214	0	*2.9*	261	1	*4.1*	248	2	*3.3*	192	0	*3.0*	181	1	*2.4*
	Girl	208	3	*3.3*	168	1	*2.2*	188	4	*3.0*	209	3	*2.8*	239	0	*3.8*	168	1	*2.3*	175	0	*2.8*	174	0	*2.3*
CDH-BB	Boy	675	2	*29.7*	609	11	*92.7*	561	7	*29.4*	600	8	*95.3*	602	4	*29.3*	666	58	*103.8*	544	7	*26.6*	544	35	*84.9*
	Girl	1676	7	*71.5*	1473	25	*222.6*	1530	5	*69.4*	1607	55	*246.2*	1643	50	*79.7*	1637	128	*255.0*	1333	52	*64.8*	1282	172	*199.8*
CDH-CB	Boy	60	0	*2.4*	45	0	*9.0*	60	0	*3.1*	45	0	*9.4*	42	1	*2.0*	44	6	*9.0*	42	1	*2.0*	45	6	*9.2*
	Girl	273	2	*10.5*	263	6	*52.2*	273	1	*11.3*	263	6	*52.8*	197	4	*9.2*	203	12	*41.5*	197	4	*9.2*	204	12	*41.7*
STEV	Boy	140	0	*2.3*	161	2	*1.5*	141	0	*2.2*	151	1	*1.3*	131	1	*2.0*	130	0	*1.1*	148	2	*2.3*	121	0	*1.0*
	Girl	86	2	*1.5*	70	0	*0.7*	69	0	*1.1*	84	0	*0.8*	72	0	*1.1*	67	0	*0.6*	55	2	*0.8*	55	2	*0.5*

Table 6: Number of relatives (m), number of affected relatives in Hungarian data (M), expected number of affected relatives for the Poisson model ($\tilde{M}$)

Another way to qualify the power of the Poisson model is to check its goodness-of-fit on the Hungarian data. In Table 6 we show the Poisson model fitted to 7 different data sets. The population data were gathered and published by Czeizel and Tusnády [1].

In Table 7 we present the goodness-of-fit values for the same data. We calculate the weighted average of the divergences for each relative. From another viewpoint, this is the normalized log-likelihood loss when changing real frequencies to the predicted probabilities.

disorder	GOF for all relatives	GOF for first order relatives
ASB	0.012189	0.000615
CLP	0.005341	0.008989
CHPS	0.007234	0.007099
VSD	0.005122	0.003212
CDH-BB	0.031767	0.002309
CDH-CB	0.050819	0.007456
STEV	0.007865	0.007432

Table 7: Goodness-of-fit of the Poisson model to Hungarian data

Finally let us present the parameter values for the best fit in Table 8.

disorder	μ_m	μ_f	ρ_m	ρ_f	Δ_m	Δ_f	λ_m	λ_f
ASB	0.015	0.026	0.018	0.010	0.018	0.010	0.00027	0.00026
CLP	0.012	0.0075	0.019	0.143	5.0e-14	0.085	0.00024	0.0012
CHPS	0.020	0.006	0.069	0.078	0.061	0.00052	0.0015	0.00052
VSD	0.016	0.013	0.0040	0.023	1.7e-17	1.3e-17	6.2e-5	0.00031
CDH-BB	0.036	0.175	0.028	0.142	3.4e-32	0.105	0.0014	0.027
CDH-CB	0.030	0.237	0.010	0.137	6.5e-16	0.102	0.00050	0.035
STEV	0.015	0.0073	0.091	0.048	0.047	1.2e-14	0.0015	0.00039

Table 8: Parameters of the Poisson model for Hungarian data

We present in Table 6 the observed occurrences M together with their expected numbers $\tilde{M}$ because these pairs offer the most plausible insight into goodness-of-fit. As it is transparent the majority of $\tilde{M}$-s are close to M with the exception of third degree relatives.

Accordingly, the first kind errors given in Table 7 are encouragingly small with the only exception of CDH-CB for all relatives. This body of Hungarian family data is less reliable because the lack of sound agreement of diagnosis of congenital dislocation of hip for different generations in XXth century countryside in Hungary.

Honestly we were quite shocked by the extremely small frequencies in Table 8. Only after understanding the dynamics of the standard model can

we state that the Poisson model offers rather acceptable fit for Hungarian data. The only question remained is the relation of Poisson model with extremely low presence of bad genes in population with the Mendelian model with dominance where the probability of expression of the malformation is around 1/2.

6. Conclusion

Let us denote by b_k the probability that a fertile person has k bad genes. Then the distribution of bad genes in the next generation has the form

$$c_k = \sum_{i=1} \sum_{j=1} H(i,j,k) b_i b_j,$$

where the kernel H is determined by biology. As it was shown in Tusnády [7] with an example, bilinear transformations of this form may be chaotic even if all the elements of H are positive. In the paper Hatvani et al. [3], the case of continuous time is investigated. It is shown that the stability of a positive bilinear operator is not ensured by the positivity of the kernel, but no example was found having chaotic attractor.

The form of selection investigated in the present paper is fortunate and ensures stability. The goodness-of-fit to population data is acceptable, the only problems are the extraordinarily small values for the parameter λ. This means that the number of bad genes is usually zero, and the appearance of a single bad gene causes the malformation or selection. Still, the low λ does not necessarily mean that the number of genes involved is small. As we mentioned in the introduction, we qualify our solution partial. It is a first acceptable solution for the problem resulting in a sound and practically applicable model. Still, the stability of the models with threshold remains open.

In a certain way the Poisson setup is richer than the Gaussian one as the expression of the malformation is randomized. The situation of this model is close to dominant Mendelian inheritance with restricted expression. If the probability of the expression depends on the gender then the situation is rather complex. In the standard model the conditional probabilities resemble the formulas of Gaussian correlations. However, when allowing gender differences in the parameters the Poisson model becomes richer: conditional probabilities (of a relative being affected when the child is affected) show stronger gender dependence in the Poisson model than in the Gaussian one. Now we are facing the question, whether the Poisson model incorporated with environmental effects offer a substantially better goodness-of-fit than the Gaussian one.

Acknowledgements. We would like to express our thanks to Villő Csiszár for her helpful and inspiring comments. We would also like to thank György Michaletzky for his support on performing the numerical investigations of the models.

We are grateful to the anonymous referee for his/her thorough and detailed review which we believe helped us to significantly improve the clarity and the quality of this paper.

References

[1] E. Czeizel and G. Tusnády, *Aetiological Studies of Isolated Common Congenital Abnormalities in Hungary*, Akadémiai Kiadó (Budapest, 1984).
[2] J. Edwards, The genetic basis of common disease, *Am. J. Med.*, **34** (1963), 627–638.
[3] L. Hatvani, F. Toókos and G. Tusnády, A mutation-selection-recombination model in population genetics, *Dyn. Syst. Appl.*, **18** (2009), 335–362.
[4] S. Karlin, Models of multifactorial inheritance: I, multivariate formulations and basic convergence results* 1, *Theor. Popul. Biol.*, **15** (1979), 308–355.
[5] A. Mukherjea, M. Rao and S. Suen, A note on moment generating functions, *Stat. Probab. Lett.*, **76** (2006), 1185–1189.
[6] K. Sankaranarayanan, N. Yasuda, R. Chakraborty, G. Tusnády and A. Czeizel, Ionizing radiation and genetic risks. V. Multifactorial diseases: A review of epidemiological and genetic aspects of congenital abnormalities in man and of models on maintenance of quantitative traits in populations, *Mutat. Res., Rev. Genet. Toxicol.*, **317** (1994), 1–23.
[7] G. Tusnády, Mutation and selection, *Magy. Tud.*, **7** (1997), 792–805 (in Hungarian).