Optimal Screening for Genetic Diseases

Eric Nævdal

Department of Economics
& HERO
University of Oslo

UNIVERSITY OF OSLO
HEALTH ECONOMICS RESEARCH PROGRAMME

Working paper 2008: 2

HERO

Optimal Screening for Genetic Diseases

Eric Nævdal*

Department of Economics & HERO

Health Economics Research Programme at the University of Oslo

HERO 2008

* Postboks 1095 Blindern, NO-0317 Oslo, Norway. E-mail:eric.navdal@econ.uio.no. I am indebted to Tor Iversen and Franz Wirl for valuable comments.

Health Economics Research Programme at the University of Oslo.
Financial support from The Research Council of Norway is acknowledged.
ISSN 1501-9071 (print version.), ISSN 1890-1735 (online), ISBN 82-7756-184-9

Optimal Screening for Genetic Diseases

Abstract

Screening for genetic diseases is performed in many regions and/or ethnic groups where there is a high prevalence of possibly malign genes. The propagation of such genes can be considered a dynamic externality. Given that many of these diseases are untreatable and give rise to truly tragic outcomes they are a source for societal concern and the screening process should perhaps be regulated. The present paper incorporates a standard model of genetic propagation into an economic model of dynamic management. The paper derives cost benefit rules for optimal screening. The highly non-linear nature of genetic dynamics gives rise to perhaps surprising results which include discontinuous controls and threshold effects. One insight is that a screening program, if at all in place at all at some point in time, should screen all individuals.

1. Introduction

Modern medicine has made great strides in curing and treating diseases caused by pathogens such as viruses and bacteria. This cannot be said about illnesses caused by our own genetic programming. This is unfortunate as many of the genetically induced conditions have devastating consequences for the afflicted individuals and their families. Diseases such as Sickle Cell Anaemia, Cystic Fibrosis and Tay Sachs are examples of diseases for which there is no cure and limited scope for therapy. Further, the economic and human costs of caring for these individuals are often considerable. What modern science is good at is the process of identifying the carriers of the gene through screening processes. Genetic tests are now commercially available and aggressively marketed in many parts of the world. However, this identification process also comes at a cost. The purpose of this paper is to establish economic criteria for how such screenings should be conducted.

There is only a limited literature on the subject and it treats the subject as an insurance issue. Typically an individual decides on whether or not to get tested, and can then choose between insurance contracts contingent on the test result. The fundamental problem here is whether there exists an equilibrium in a market for insurance when information on genetic risk is endogenous, Doherty and Thistle (1996), Hoel *et al* (2006). This approach overlooks two basic features of genetic diseases. First, the transmission of disease inducing genes represents an externality in the economic sense. If parents transmit such genes, it is truly an unintended

consequence that is passed on to future generations, possibly through several generations before manifesting itself. Second, the transmission of genes from generation to generation is a dynamic process. These facts taken together imply that screening for genetic diseases is a long-term process where some degree of public regulatory management may be warranted. Here we incorporate these two aspects into a model and examine the decision on screening as a public health issue. As is shown below, results indicate that for serious genetic disorders, insurance considerations may be of secondary importance.

Considerable effort is spent at screening for genetic diseases. In USA, almost all children are screened for Phenylketonuria in order that an extremely rigorous diet may be initiated to avoid or manage symptoms. Israel has programs that screen for Tay-Sachs disease as well as other genetic diseases.[1] The Israeli programs have been very successful in the sense that the number of children born with Tay-Sachs has been dramatically reduced. In many cases, particular genetic diseases are more common in specific ethnic or social groups. Such groups are often screened separately. When analyzing this topic one enters an ethical minefield. If a couple plans to have children and they have a probability that the children will carry some terrible disease, the simple fact is that there are no options without pain. They can choose/be

[1] Symptoms for Phenylketonuria include severe physical and mental retardation. In almost all cases, the disease presents around the age of six months. Symptoms for Tay-Sachs include blindness, deafness and atrophying muscles. There is no cure or effective therapy. Death usually occurs before the age of five.

ordered not to have children. This option implies that a couple in love is barred from the possibility of having healthy children together.[2] This is in itself a tragedy, which may end a loving relationship. On the other hand, if this couple has a baby, the child may face a very short life span traumatized by pain and retardation. Pre-natal testing for genetic diseases can be done with routine screening procedures such as amniocentesis and chorionic villus sampling. Such testing raises its own set of ethical issues as many parents who learn of that they are pregnant with a child with a serious disease will choose to have the foetus aborted. The ethics of genetic screening is hotly debated issue and rightly so as it deals with fundamental issues of life, death and health.[3] This paper does not take sides in this debate, but takes the utilitarian view that if pre-natal genetic screening already exists and is performed on a substantial scale around the world, the debate can be improved if it is informed by the policy prescriptions of economic analysis.[4]

The rest of the paper is organised as follows. Section 2 gives a short primer in the ecology and mathematics of genetic diseases. Section 3 presents the economic model

---

[2] The probability of having healthy children, given that there is a risk of genetic disease varies with the type of genetic disorder. For the disorders covered by the model in the present paper, the probability that a child will be healthy is 75%.

[3] The Israeli screening programs led to considerable debate among religious scholars, see Rosner (1998) for an overview of the arguments.

[4] Another source of unease with the present topic is that the question of heredity, disease, biology and screening has been a favourite topic for a rare bunch of unpleasant people. Some of these people have been very clever, but unpleasant nevertheless. The famous statistician Sir Francis Galton and Nazi ideologue Alfred Rosenberg come to mind.

and derives the results. Section 4 summarizes the results and indicates possible future research.

2. Genetic Diseases and their Propagation

In order to set the stage for the economic analysis we give an elementary description of the genetic mechanisms involved. A more thorough description may be found in any textbook on quantitative genetics such as Falconer and Mackay (1996). In the light of evolutionary theory, it may seem paradoxical that genetic diseases exist at all. Genes are supposed to be selected over thousands of generations to increase our ability to have healthy babies, not to give is terrible diseases. The fact that they do has several, often subtle, explanations. First, it should be noted that some genetic diseases are not hereditary. Downs syndrome is for the most part not hereditary. Here we focus on hereditary diseases so further discussion of non-hereditary is diseases is omitted. Further, we shall only discuss single gene genetic diseases. Diseases that depend on the expression of several genes are vastly more complicated and in most cases not worked out in sufficient detail to be useful for economic modelling. It should be noted that all diseases mentioned in this article belong to this type unless otherwise stated.

Every human receives one copy of the chromosomes of each parent. These chromosomes are paired together in most human cells. Located on the chromosomes are genes whose primary purpose is to act as recipes for proteins. Many genes come in different varieties, termed alleles, and we may or may not receive different alleles

from our parents. If we receive the same allele from both parents, we are homozygote with respect to that gene. If we receive different alleles from our parents, we are heterozygote. For homozygotes, the gene will, if expressed, induce the production of a protein, which determines some aspect of our physiology or behaviour. For heterozygotes the picture is more complicated. In many cases, one of the alleles will completely control what protein is produced. An allele that completely controls expression is termed dominant. The other allele is termed recessive. Intermediate cases where no allele dominates completely also exist. In these cases both alleles are active to some degree. The different combinations are termed genotypes. There are two homozygote genotypes and one heterozygote genotype.

Different protein expressions give rise to different phenotypes. Phenotypes are the physiological and/or behavioural manifestation caused by the proteins. If one allele is completely dominant, individuals may be one of two possible phenotypes. Individuals with two copies of the dominant allele are indistinguishable, except at the microbiological level, from individuals with one copy of the dominant allele and one copy of the recessive allele. These two genotypes constitute one phenotype. Individuals with only recessive genes constitute a different phenotype. If none of the alleles dominates completely, there will be three phenotypes, one for each genotype.

With single gene diseases, it will be the case that individuals with one or two copies of the gene will become ill. If only a homozygote gives rise to the disease, the disease is recessive. If both homozygote and heterozygote becomes ill, the disease is dominant. If there is not complete dominance, the homozygote and the

heterozygote may exhibit different degrees of illness. Dominant and recessive diseases will typically be very different with respect to when a disease presents over the individuals lifetime. A serious dominant disease will generally appear rather late in life. The ecological reason for this is that a dominant disease that presents early will impede the individual's reproduction. As all individuals with a disease inducing allele will become ill, this allele cannot hide by being paired with a non-inducing allele. Alleles coding for dominant diseases that presents before the individual has reproduced successfully will quickly be extinguished by Darwinian selection. This is why a disease such as Huntington's disease will on average present when the afflicted is in the late fourties/early fifties.[5] With recessive diseases, the picture is usually different. As the disease inducing allele can hide by pairing with another allele, the forces of selection are much weaker. Heterozygote individuals being able to reproduce without presenting the disease compensate for the early death of homozygote individuals. However, an allele which induces a disease when the individual is homozygote will also be selected out of the population although at a slower rate than dominant diseases. The persistence of alleles that recessively induce disease is usually explained by there being some sort of advantage to being heterozygote. Two copies of a disease inducing allele kills you whereas only one copy gives you an reproductive advantage relative to individuals with no copies of this allele. The classical example

[5] It should be noted that, although Huntington's disease is a single gene disease, the genetics is somewhat more complicated than the rough description of genetic diseases given here. The disease is also dominant and does not fit in the model presented below.

of this phenomenon is the Sickle Cell gene, which induces severe symptoms when two copies are present, but protects from malaria even with only one copy, Kwiatkowski (2005). It has been suggested that Tay-Sachs gives a heterozygous advantage because of increased resistance to tuberculosis, Rotter and Diamond (1987). It has also been suggested that Tay-Sachs endows heterozygous carriers with an increased IQ which presumably endows the carrier with a fitness advantage, Cochran *et al* (2006). In many cases, the specific advantage to a heterozygote may be hard to identify, but with a line of reasoning that should be familiar to economists, biologists argue that if there were no advantage in being a heterozygote and only disadvantage in being homozygote, the disease inducing allele wouldn't exist.

There is much more to genetic diseases than what has been covered here. The above only applies to genes not located on one of the sex chromosomes. Genes located on sex chromosomes adds a layer of complications. Further, in addition to our own genes, our cells have organelles called mitochondria. These organelles, believed to originate with a symbiotic relationship with "bugs" early in life's evolutionary history, have their own independent genome and come with their own set of genetic diseases.

The variation in types of genetic diseases implies that there really is no general model that can be applied to all varieties. Here we focus on recessive diseases induced by a variant of a single gene, which still covers dozens of diseases. In addition to the diseases mentioned above, the genetic model in the present paper also fits Cystic fibrosis, Gaucher's disease, Type 1 Galactosemia, Pendred disease,

Thalassemia, some types of Erb's muscular dystrophy and Werdnig-Hoffmann disease to mention a few.

2.1. The Genetic Model without intervention

We employ a standard Mendelian model of an autosomal recessive point mutation. There are two alleles, one mutant and one "normal", of a gene. Denote the mutant allele a, and the normal allele A. There are three possible combinations of alleles; AA, Aa and aa. Each of these combinations represents a distinct phenotype. The mutation is genetically beneficial, if it is coupled with the normal gene as a heterozygote, but harmful if appears as a homozygote. Using standard conventions from quantitative genetics we assume that the natural coefficient of selection for the AA phenotype is $s_1 \in (0, 1)$, and $s_2 \in (0, 1)$for the aa Phenotype. This implies that if each individual of the Aa phenotype has on average N offspring, phenotype AA has on average $N(1 - s_1)$ offspring and phenotype aa has on average $N(1 - s_2)$ offspring. Let q be the frequency of a in the population. From standard genetic models it follows that the number of aa phenotypes in the population is given by q^2, the frequency of AA phenotypes is $(1 - q)^2$ and the frequency of Aa phenotypes is $2q(1 - q)$. Given our assumptions it can be shown that the genetic dynamics are determined by:

$$\dot{q} = \frac{q(1-q)\left(s_1(1-q) - s_2 q\right)}{1 - s_1(1-q)^2 - s_2 q^2}, \quad q(0) \text{ given} \tag{1}$$

Equation (1) is derived in some detail in the Appendix. This equation has three steady states, $q = 0$, $q = 1$ and $q = s_1/(s_1 + s_2)$. It is easy to show that $q = 0$ and $q = 1$ are unstable and that q will converge to $q = s_1/(s_1 + s_2)$, denoted $\tilde{q}$, for all $q(0) \in (0, 1)$. This is illustrated by Figure 1, which shows a plot of $\dot{q}$ as a function of q.

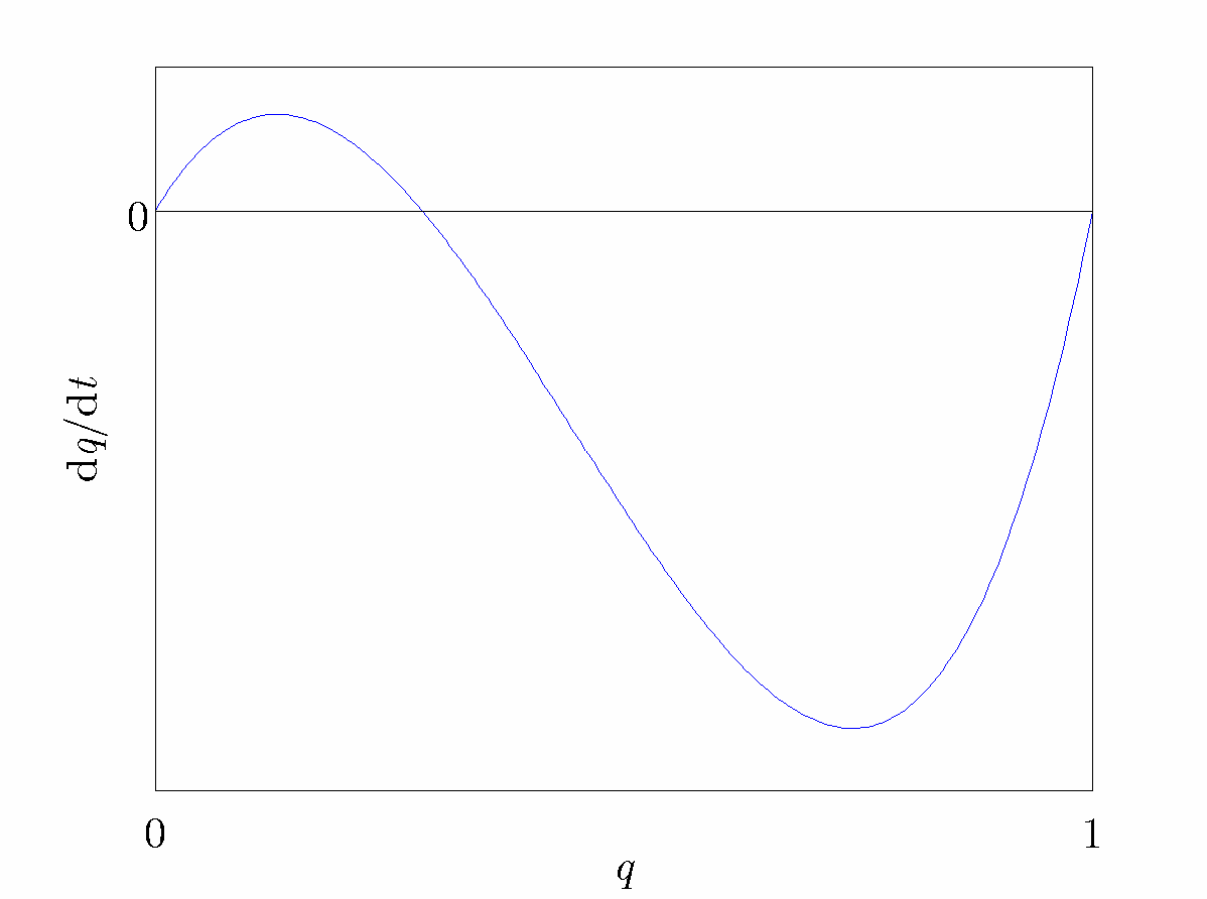

Figure 1, dq/dt for s1 = 0.1 and s2 = 0.25. Interior steady state value of q is 2/7.

As it is assumed that the a allele is the potentially malign gene variety, it is assumed that $s_1 < s_2$. This bounds $\tilde{q}$ between 0 and ½. One aspect suppressed in this setup is population dynamics. We will assume that the genetic disease does not affect the size of the population. This is a fairly innocuous assumption as mortality from most

genetic diseases is a small portion of the total mortality at any given time.

The Economics of Genetic Screening

The mathematical analysis of the present problem is somewhat involved. Therefore we first analyze the mathematical properties of the model and thereafter extract the economic implications. It is assumed that an effort $E(t)$ is exerted in order to screen individuals. $E(t)$ is to be interpreted as the fraction of the population screened at time t. The total number of screened individuals who are aa is then given by:

$$F = q^2 E \tag{2}$$

Assume further that individuals who are aa phenotypes are prevented from reproducing. The manner in which this is done is left unspecified, but one possibility that foetuses are screened and, if found to be aa, aborted. This process modifies the equation (1) to become:

$$\dot{q} = \frac{q(1-q)\big(s_1(1-q) - \big(s_2 + E - Es_2\big)q\big)}{1 - s_1(1-q)^2 - \big(s_2 + E - Es_2\big)q^2}, \quad q(0) \text{ given} \tag{3}$$

One can easily verify that the denominator will be non-negative for all $s_2 \in [s_1, 1)$ and $E \in [0, 1]$. This fact will be important later when second order conditions are discussed. At any time t, the fraction of the population in which the disease is expressed gives rise to disutility. It is assumed that this cost is given by b. This cost

is assumed to capture all aspects of the disease including treatment costs where treatments are available. Further, it is assumed that the marginal cost of screening effort is given by a constant K. Thus, the objective function for society is given by:

$$J\left(q\left(0\right)\right)=\max_{0\leq E(t)\leq 1}\int_{0}^{\infty}\left(-b\left(q\left(t\right)\right)^{2}-KE\left(t\right)\right)e^{-rt}dt \tag{4}$$

The problem for a regulator is then to find $E(t)$ to maximize equation (4), subject to the differential equation in (3). Proceeding with the Pontryagin Maximum Principle in the standard manner, we define the Hamiltonian:

$$H\left(E,q,\lambda\right)=-bq^{2}-KE+\lambda\frac{q\left(1-q\right)\left(s_{1}\left(1-q\right)-\left(s_{2}+E-Es_{2}\right)q\right)}{1-s_{1}\left(1-q\right)^{2}-\left(s_{2}+E-Es_{2}\right)q^{2}}$$

The necessary conditions are:[6]

$$E=\arg\max_{Y\in[0,1]}H\left(Y,q,\lambda\right) \tag{5}$$

In the event that the optimal E lies in (0, 1), E is the solution to:

$$\frac{\partial H}{\partial E}=-K-\lambda\frac{\left(1-q\right)q^{2}\left(1-\left(1-q\right)s_{1}\right)\left(1-s_{2}\right)}{\left(1-s_{1}\left(1-q\right)^{2}-\left(s_{2}+\gamma E-\gamma Es_{2}\right)q^{2}\right)^{2}}=0 \tag{6}$$

[6] Here Y is used to indicate E prior to optimization.

$$\begin{aligned}\dot{\lambda} &= r\lambda - \frac{\partial H}{\partial q} \\ &= r\lambda + 2bq + \lambda\left(1 + \frac{2(1-q)(1-(1-q)s_1)^2 - (2-q-2(1-q)s_1)\left(1-s_1(1-q)^2 - (s2+E-Es_2)q^2\right)}{q\left(1-s_1(1-q)^2 - (s2+E-Es_2)q^2\right)^2}\right)\end{aligned} \tag{7}$$

Optimal paths are given by solutions to (3), (5) and (7), which also satisfies the transversality conditions. Given the complexity of these conditions, there is no hope of finding a closed form solution. Further, explicit formulae for steady states, if they exist, will be quite cumbersome. We can however gain quite some insight by performing some simple algebraic manipulations. First, note that in steady state it must hold that $\dot{q} = 0$ which implies that

$$E = \frac{s_1(1-q) - s_2 q}{(1-s_2)q} \tag{8}$$

Inserting E from (7) into (8) and then solving for λ yields:

$$\lambda = -\frac{2bq\left(1 - s_1(1-q)\right)}{r + (1-q)(1-r)s_1} \tag{9}$$

Note that λ is negative for all $q \in (0, 1]$, reflecting that the marginal contribution of q is negative. Inserting from (8) and (9) into (3) gives the following expression:

$$\begin{aligned}\left(\frac{\partial H}{\partial E}\right)_{\dot{q}=0,\dot{\lambda}=0} &= -K + \frac{2b(1-q)(1-s_2)q^3}{r + (1-q)(1-r)s_1} = 0 \\ &\downarrow \\ -K\left((1-r)s_1 + r\right) + K(1-r)s_1 q + 2b(1-s_2)q^3 - 2b(1-s_2)q^4 &= 0\end{aligned} \tag{10}$$

The left hand side of the last equation in (10) will be of some importance in the

sequel and we denote this expression:

$$\Theta(q) = -K\left((1-r)s_1 + r\right) + K(1-r)s_1 q + 2b\left(1-s_2\right)q^3 - 2b\left(1-s_2\right)q^4 \tag{11}$$

Note that $\Theta(0) < 0$ and $\Theta(1) < 0$. $\Theta(q)$ has the following properties:

$$\left(\frac{\partial H}{\partial E}\right)_{\dot{q}=0,\dot{\lambda}=0} = 0 \Rightarrow \Theta(q) = 0, \left(\frac{\partial H}{\partial E}\right)_{\dot{q}=0,\dot{\lambda}=0} > 0 \Leftrightarrow \Theta(q) > 0 \tag{12}$$

Roots of $\Theta(q)$ are candidates for be steady states for the solutions to problem (4). However, it should be noted that all roots with values of q larger than $\tilde{q}$ are irrelevant as it would require that a genes be introduced into the population from an external source.

The general shape of $\Theta(q)$ is illustrated in Figure 2. It is straightforward to check that $\Theta(q)$ has either two or zero roots in [0, 1] and that one of these roots is located in [1/2, 1].

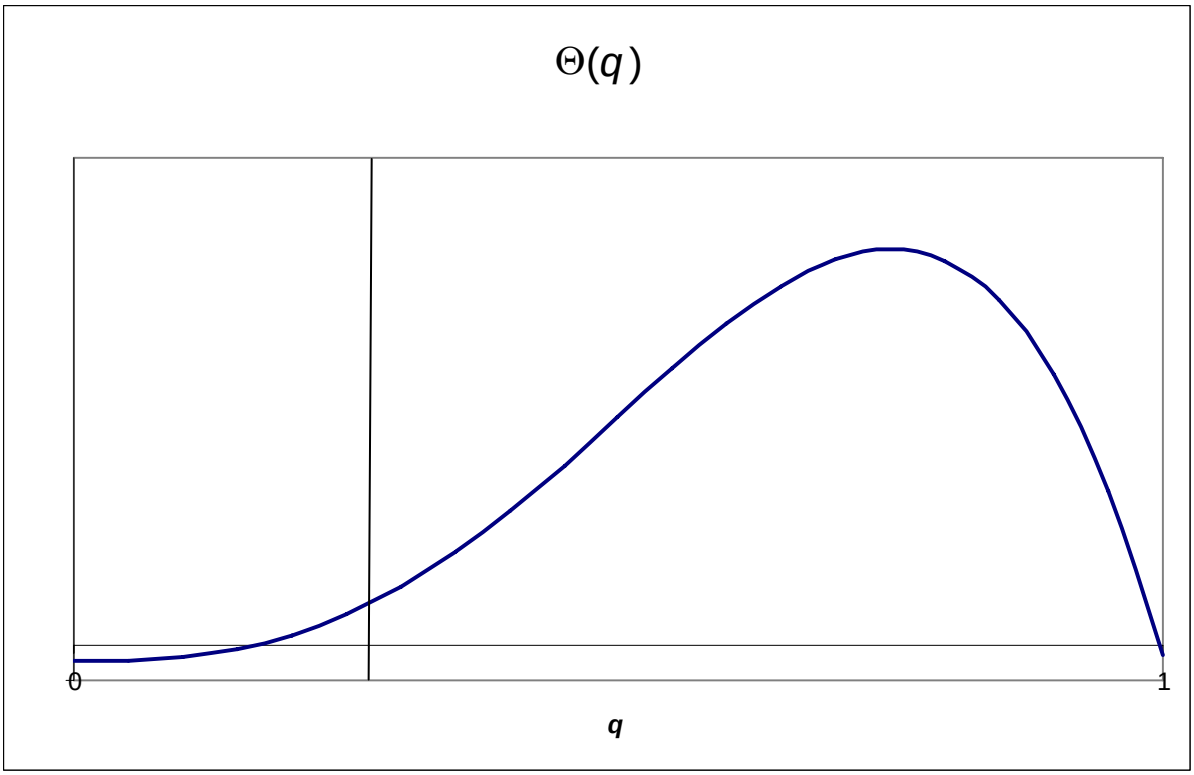

Figure 2, The polynomial $\Theta(q)$

If we focus on the shape of $\Theta(q)$ over $[0,\ \tilde{q}\,]$, there are two possibilities. Either there is a single root in $[0,\ \tilde{q}\,]$, or there is none. This is illustrated in Figure 3. A more formal phase diagram is shown below, but Figure 3 provides some good intuition for the qualitative nature of the solution.

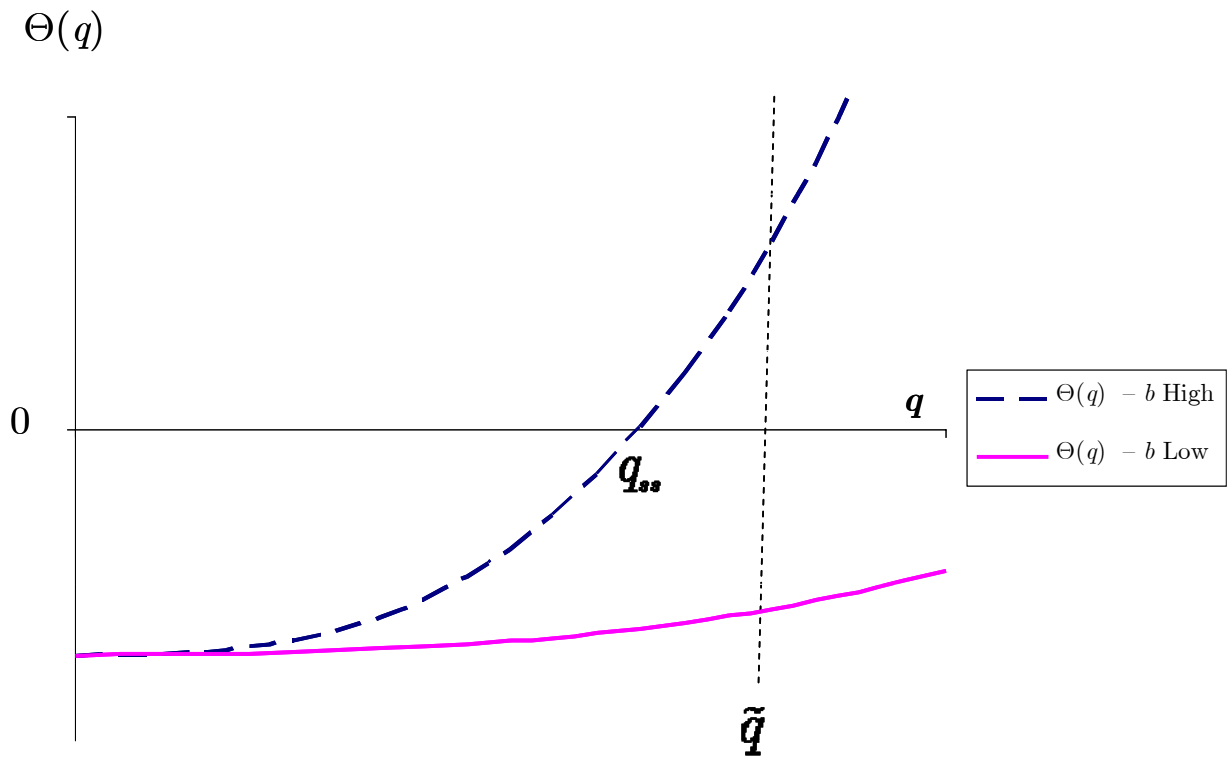

Figure 3, Roots of $\Theta(q)$ in $[0,\ \tilde{q}\,]$. The steady state is indicated by q_{ss}.

If b is "high", then there is an unique value of q, termed q_{ss}, which is a steady state. If b is "low," then there is none except possibly $\tilde{q}$. First we examine the case where b is high. Recall that $\Theta(q)$ is the derivative of the Hamiltonian with respect to the control variable, E, given that E is chosen so that $\mathrm{d}q/\mathrm{d}t = 0$. If $\Theta(q)$ is positive for a given, this implies that it would pay to increase E relative to the steady state level. But if E increases above the steady state level, then $\mathrm{d}q/\mathrm{d}t$ decreases. Conversely, if $\Theta(t)$ is negative, then it pays to reduce E relative to the steady state level and $\mathrm{d}q/\mathrm{d}t$ decreases. From Figure 3 it is clear that $\Theta(q)$ is positive for $q > q_{ss}$, so for any value of $q \in (q_{ss}\ ,\ \tilde{q}\,)$, q will be a decreasing function of time and converge towards q_{ss}. For

$q_{ss} \in [0,\ q_{ss})$ it will pay to decrease E as the marginal value of E is negative if E is set in order to keep $\mathrm{d}q/\mathrm{d}t = 0$. It follows that q will approach q_{ss} both from above and below and q_{ss} is therefore a stable steady state.

This was for a high value of b. The same reasoning can be applied to low values of b. Here $\Theta(q)$ is negative for all q. Thus for any value of q it is optimal to decrease E relative to the value of E that gives $\mathrm{d}q/\mathrm{d}t = 0$, and q will converge towards $\tilde{q}$.

In principle, it is possible to solve (12) and find q_{ss}. Unfortunately $\Theta(q)$ is a fourth order polynomial and any solution, if it exists, will be analytically cumbersome. However, it is a trivial task to find solutions with numerical methods for given values of the parameters. Inserting such a numerical value into (9) gives the steady state value of λ. One could then think that inserting the steady state value of q into (8) would yield the steady state value of E. This turns out *not* to be the case as the Hamiltonian is not concave with respect to E when $\mathrm{d}q/\mathrm{d}t = 0$. Calculating the second derivative of the Hamiltonian for interior steady states gives:

$$\left(\frac{\partial^2 H}{\partial E^2}\right)_{\dot{q}=0,0<q<1} = -\frac{2(1-q)q^4\left(s_2-1\right)^2\lambda}{\left(1-(1-q)s_1\right)^2} > 0 \tag{13}$$

This expression is positive for all negative values of λ. A value of E which solves the equation $\partial H/\partial E = 0$ in steady state will actually minimize the Hamiltonian! The value of E which maximizes the Hamiltonian in steady state must therefore be found on the boundary of the control region [0, 1]. But if one of these values is chosen, q will move away from the steady state towards either $q = 0$ or $q = 1$. And it has been

shown above that $q = q_{ss}$ is a steady and an attractor for all q in (0, 1). The solution to this conundrum is that in steady state, the optimal control function is a chattering control.[7] By letting E fluctuate between 0 and 1, the state and co-state variable remain in steady state. One can show that the Hamiltonian is in fact convex with respect to E for all admissible values of E, q and λ. (This is shown in the appendix.)

[7] This type of control turns up from time to time when the Hamiltonian is convex in the control variable. See Clark (1990), page 149 for an accessible discussion.

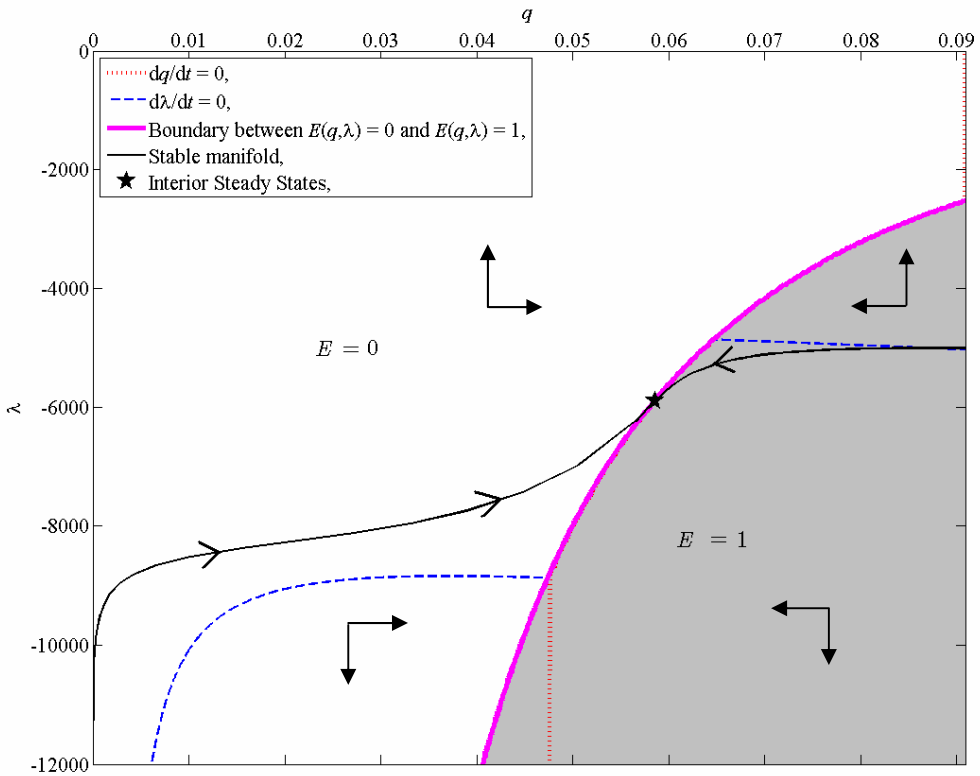

Figure 4, Phase diagram for q and λ. For all pairs of λ and q in the grey area the value of E that maximizes the Hamiltonian is 1. In the non-shaded area, the optimal value is 0. Note that the null-clines for $\mathrm{d}q/\mathrm{d}t$ and $\mathrm{d}q/\mathrm{d}t$ in a distinct segment lies very close to the boundary between $E = 0$ and $E = 1$. The null-clines only intersect at the steady state point indicated by the star. The black line is the stable manifold with the direction of the optimal path indicated by the arrows.

It follows that it will always be optimal to choose a value E on the boundary of the control region. In order to illustrate optimal paths further, a phase diagram is shown in Figure 4. The figure is drawn by a simple computer program, which is available from the author on request. The stable manifold that defines the optimal path is indicated by the black solid line. The direction of the optimal path is indicated by the arrows. The first thing to note from the figure is that everywhere along the

optimal path the optimal choice of E is found at the boundary of the control region. This is because the Hamiltonian being everywhere convex with respect to E. A consequence of this is that q will reach the steady state in finite time regardless of the starting value of $q(0)$. Careful examination of the phase diagram also reveals that there is no other path that satisfies necessary conditions. Thus, under the assumption that an optimal path actually exists, the path shown in Figure 4 is actually the optimal path.

Economic Implications

Despite its complexity, the model has some very distinct implications. If there is screening, it should be performed universally. This is proven in Proposition 1.

Proposition 1.

At any given point in time, it is either optimal to screen all foetuses or no foetuses.

Proof:

This is a mathematically trivial consequence of the Hamiltonian being convex with respect to E.

Although Proposition is mathematically trivial it should be noted that it follows

directly from the fundamental dynamic properties of the propagation of genes that induce autosomal recessive diseases. This result can be avoided by assuming that the second derivative of the screening cost function is sufficiently large. However, a convex screening cost function is only a necessary condition for a concave Hamiltonian and convexity may continue to be an issue.

One striking feature is that there is a lower bound on the cost of the disease where screening program is economically optimal. This is proven in Proposition 2.

Proposition 2.

There is a threshold value b^* such that if $b \leq b^*$, no screening should be conducted at any time. b^* is given by:

$$b^* = \frac{K\left((1-r)s_1 + r\right) - K(1-r)s_1\tilde{q}}{2\left(1-s_2\right)\left(\tilde{q}^3 - \tilde{q}^4\right)} \tag{14}$$

Proof:

The steady state level of q_{ss} is determined by solutions to the equation

$$\Theta(q_{ss}) = -K\big((1-r)s_1 + r\big) + K(1-r)s_1 q_{ss} + 2b\big(1-s_2\big)q_{ss}^3 - 2b\big(1-s_2\big)q_{ss}^4 = 0 \qquad (15)$$

This equation has at the most two roots of which a maximum of one root is located in [0, ½]. As $q(t) \leq \tilde{q} < ½$, there is at most one steady state. If $b = 0$, then $\Theta(q_{ss}) < 0$ for all q_{ss}. For b sufficiently large, there is a value $q_{ss} < \tilde{q}$ which solves (15). It follows that there exists a b^* such that $q_{ss} = \tilde{q}$. From the convexity of the Hamiltonian we know that optimal E must be 0 or 1. In order for q to converge to $q_{ss} = \tilde{q}$, it must hold that $E = 0$ for all t. b^* is then found by solving the equation $\Theta(\tilde{q}) = 0$ with respect to b.

This proposition trivially makes sense. Nobody in their right mind would start a screening and abortion program for e.g. a disease that gives a slight increase in the risk of blood clots in veins. (Factor V Leiden mutation.) However, many people would agree that it is justified for Tay-Sachs disease, which imposes severe suffering on the afflicted until death occurs, usually before the age of 5.

Taken together Propositions 1 and 2 have some important implications. If a particular genetic disease is considered severe enough to instigate a screening program, then the program should be universal. This may be taken to imply that if $b > b^*$, a genetic screening program is a public health issue and not a private matter. Many of the complexities discussed in the literature on genetic screening arise from

the potential conflict between private health care insurers and potential patients. These issues become less relevant when genetic information is gathered and acted on in a public health care system with the cost borne by the taxpayers. Also, the information aspects of screening becomes a somewhat moot point all individuals are screened. However, if $b < b^*$ there is still scope for conflict between insurers and insured and the insights from the health insurance literature applies.

Proposition 3.

Assume $s_1 > 0$. Then, for positive K it is never optimal to eradicate the disease-inducing gene completely.

Proof:

From the previous discussion it is clear that q_{ss} is a an attractor for all q in $(0, \tilde{q})$, it follows that $q(t)$ will be bounded away from 0.

One way of interpreting this is result is to note that the number of ill people is equal to q^2. As q becomes small, the number of ill goes rapidly towards zero. The number of ill caught for a given screening effort will therefore also decrease rapidly as q goes to zero. As the optimal screening effort is either $E = 0$ or $E = 1$, it makes sense that for sufficiently low q it makes sense to let $E = 0$. If $q < q_{ss}$ it follows that q will increase.

An unattractive feature of the model is that the steady state is maintained through a chattering control. As previously pointed out, this is caused by the dynamic properties of the model. It is hard to imagine a screening program which switches between screening all foetuses and screening none at virtually every point in time. One of the reasons such a system seems impractical is that there are costs incurred when starting up and closing down a screening program. Incorporating such costs will generally imply that optimal screening will be a pointvise continuous function of time and that the frequency of the a allele will oscillate, see Wirl (1991) for an example.[8]

Summary

Optimal pre-natal screening efforts for a certain class of genetic diseases has been investigated. It is found that there is a threshold cost of disease below which it is never optimal to screen. In this case, the frequency of the disease is determined by the evolutionary selection pressure on different genetic varieties. If the cost of disease is above this level, it is optimal to screen all foetuses until the frequency of the disease has reached a steady state level which is acceptable given the costs of screening and the cost of disease. Only pre-natal screening is examined. Alternative

[8] Obviously, if the switching costs are sufficiently high, it may be optimal to never start, or alternatively, stop a screening program.

screening strategies include genetic testing of parents and genetic counselling. An obvious extension of our model is to establish efficiency criteria for genetic counselling, possibly in conjunction with pre-natal screening.

References:

Clark, C.W. (1990), Mathematical Bioeconomics. 2nd ed. Wiley InterScience, Hoboken, 1990.

Cochran, G., J. Hardy and H. Harpending, (2006), Natural History Of Ashkenazi Intelligence, *Journal of Biosocial Science* , 38: 659-693.

Doherty N.A. and P. D. Thistle, Adverse selection with endogenous information in insurance markets, *Journal of Public Economics*, **63**:83–102.

Falconer, D. and T.F. Mackey, (1996), Quantitative Genetics. 4th ed. Longman, New York, 1996.

Hoel, M., T. Iversen, T. Nilssen, and J. Vislie, Genetic testing in competitive insurance markets with repulsion from chance: A welfare analysis, *Journal of Health Economics* **25**:847-860.

Kwiatkowski, D. P., (2005), How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria, *American Journal of Human Genetics*. **77**,2:171–192.

Rotter, J.I and J. M. Diamond, (1987), What maintains the frequencies of human genetic diseases?, *Nature*, 329:289-290

Rosner, F., (1998), Judaism, Genetic Screening and Genetic Therapy, *The Mount*

Sinai Journal of Medicine, **65**:406-413.

Wirl, F., (1991), Economics of (oil) price politics: Penalizing price changes, *Journal of Policy Modeling*, **13**, 4:515-527.

Appendix.

Deriving Equation (1).

This equation is a standard result in quantitative genetics. The exposition follows Falconer and Mackey (1996). A gene is located at a locus. At any time t, if the fraction of genes at a particular locus that belongs to the allele a is q_t and the fraction belonging to the allele A is $1 - q_t$, then the population can be divided into three phenotypes. The fraction of the population being aa is $Q_t = q_t^2$, the fraction being Aa is $H_t = 2q_t(1 - q_t)$ and the fraction being AA is $P_t = (1 - q_t)^2$. Assuming that reproduction is sexual and well-mixed, then in the absence of any selection, the q_{t+1} would be given by $q_{t+1} = (2Q_t + 2\ ½H_t)/(2Q_t + 2H_t + 2P_t) = q_t$. This formula simply counts the total number of a genes and divides them by the total number of genes. If the aa and AA individuals are selected against, this formula must be modified. Assume that for every Aa for 100 offspring that reproduces, only $100\times(1 - s_1)$ of the AA individuals and $100\times(1 - s_2)$ of the aa achieve the same.

$$q_{t+1} = \frac{(1-s_2)Q_t + \tfrac{1}{2}H_t}{(1-s_2)Q_t + H_t + (1-s_1)P_t} = \frac{q_t - s_2 q_t^2}{1 - s_1(1-q)^2 - s_2 q_t^2} \tag{A.1}$$

Subtracting q_t on both sides yields:

$$q_{t+1} - q_t = \frac{q_t - s_2 q_t^2}{1 - s_1(1-q)^2 - s_2 q_t^2} - q_t = \frac{q(1-q)\left(s_1(1-q) - s_2 q\right)}{1 - s_1(1-q)^2 - s_2 q^2} \tag{A.2}$$

This is a discrete time version of the differential equation for q presented in Equation 1.

Convexity of the Hamiltonian with respect to *E*.

The second derivative of the Hamiltonian with respect to E may be written:

$$\frac{\partial^2 H}{\partial E^2} = \frac{-2(1-q)q^4\left(1-(1-q)s_1\right)\left(1-s_2\right)^2 \lambda}{\left(1 - s_1(1-q)^2 - \left(s_2 + E - Es_2\right)q^2\right)^3} \tag{A.3}$$

λ is negative for all t. This follows directly from the fact that λ is the marginal contribution of q to the objective function at time t. Therefore the numerator is positive. The denominator is the denominator from equation (3) raised to the power of 3. As this expression is positive for all admissible q and E, it follows that the Hamiltonian is convex with respect to E over [0, 1] and strictly convex over (0, 1).