# RISKS ASSESSMENT - ROLE OF PRE-EXISTING GENETIC VARIATION

**Paul Unrau and Katazyna Doerffer**
Radiation Biology and Health Physics Branch
Atomic Energy of Canada Limited
Chalk River Laboratories

To assess risk from an external source in a population, first the effects of inherited or internal risks need to be identified so that we can discriminate those imposed by the external risk source. Fundamentally, risk must be differentiated into those pre-existing components that are part of the biological system, those components that act on the system from the environment, and the interactions of these components in the biological system.

When the genetic program (the genotype) which codes for an individual is expressed, the observed result is called the phenotype. Each person has a unique genotype and phenotype. The phenotype may be modified by the environment. In radiation risk assessment we are concerned mainly with the effects of the environmental agent, ionizing radiation. The effect on the phenotype we are concerned with mainly involves higher risks of cancer, or of the induction and inheritance of new mutations. It should be emphasized that usually the genes are received unchanged in form from the parents. When a new phenotype arises in the children, these need not be because of new mutations, because the new gene combinations in the offspring may give rise to apparently new characteristics.

In general, each phenotype results from a particular genotype under the influence of the environment as follows [1]:

$$P = G + E$$

where: P is the phenotype,
G is the genotype, and
E is the effect of the environment.

The interaction of these two variables, "G" (with the pre-existing complexity inferred from human genetic uniqueness), and the environment, "E", defines this relationship. There is ample evidence that most human disorders, and much susceptibility to cancer, arises from new combinations of genetic variants already extant in the population. We need to identify the sources of these genetic variants in order to identity new variants. Only new variants can arise from the action of the environment. Disentangling these risk sources is thus required for risk assessment. This presentation describes our approach to analyze these interactions.

In Mendelian genetic studies disorders are defined by their inheritance patterns. For example, simple Mendelian monogenic disorders are exemplified by such diseases as retinoblastoma [2], or Tay-Sachs disease [3]. In more complex disorders, either one genotype can result in different phenotypes due to the influence of the environment, or different genotypes can result in the same phenotype, because different genetic causes can lead to the same disorders. Examples of the latter class include the observation that alternative comple-

mentation group alleles lead to the same disease, as in Xeroderma Pigmentosum [4], or as we discuss below, different mutations in the same gene can lead to the same disease, as in Tay-Sachs, even though the mutations and their origins differ.

Some genetic risks are simply inherited in the new zygote. In other cases, the potential for change is inherited, and may lead to genetic and somatic changes. Independent mutations can also occur, depending upon the nature of the environmental insult and the genetically determined response of the individual.

For risk assessment, this means it is essential that we are able to find out what the prior genetic state of an organism was, in order to determine if the risk source has caused a particular outcome. In essence, this means that we have to assess whether the "change" was present in one or other parent. Only if it is not, can a change be "new mutation". So the minimum requirement is to distinguish old genetic variations from new ones as only the new may have been caused by the putative risk source.

We are therefore: developing molecular biological tools to distinguish inherited genetic variants from new mutations; analyzing molecular biology databases [5,6, 7] to determine biological sources of pre-existing variation; and integrating these sources of information into Molecular Epidemiology.

Biological sources of genetic and cancer risks expressed in the initial genotype of individuals result from the reproductive behaviour and genotypes of their parents. Two reproductive behavioral variables have been studied as biological sources expressed in some genetic and cancer risks. (Other risks may also occur for the same reasons.) We used "ethnicity" as a surrogate for the interbreeding and shared genetic risks of self defined breeding groups containing potentially related individuals. Consanguinity was used as a direct indicator of marriage of related individuals.

If either ethnicity or consanguinity can be related to risk changes (either increased or decreased), it means that epidemiological studies that do not take these factors into account may be misleading with respect to the attribution of risk source to an environmental variable, when a biological variable is the actual risk source.

For population based risk assessment, it is a simplifying assumption that people choose their mates randomly. (In fact, it is required in epidemiological studies that people choose their mates randomly with respect to some disease or disorder.) Humans are as discriminating in their mating behaviour as any other animal, and they discriminate in ways that lead to non-random breeding, which has profound effects on the distribution of inherited genetic risks in populations. This non-randomness may underlie much of the variation in disease distribution presently being associated with external risk sources by epidemiological means. In actual risk assessment explicit measures need to be taken to address the genetic effects of non-random mating. This cannot be done in current epidemiology, but will be the inevitable product of the new Molecular Epidemiology.

Table 1 is based on an analysis of ethnicity indicators from the Online Mendelian Inheritance in Man (OMIM) database as an indicator of genetic risk. It summarises the impact of ethnicity on risk profiles of 51 different ethnic groups examined. What we observed is that many of these disorders are associated with specific ethnic groups. This means that the genetic risks conferred by the segregation of the causative genes are largely limited to those ethnic groups. Of course most genetic diseases are not analyzed with respect to ethnicity, so this represents a minimum estimate of the effects of ethnicity in determining genetic risks. Even for diseases shared across ethnic groups, (as described in Table 2 below), the molecular biological basis of these risks can be different. That is, different mutations in the same genes might be involved, each specific to one of the ethnic groups carrying the disorder.

Table 1. Number of OMIM entries associated with ethnicity indicators and the distribution of specific versus disorders shared by ethnic groups.

| | |
|---|---|
| Number of genetic disorders in OMIM associated with ethnicity. | 713 (12%) of OMIM |
| Number of genetic disorders specific to ethnic groups. | 591 (83%) |
| Number of genetic disorders shared across ethnic groups. | 122 (17%) |

Table 2 shows genetic risk profiles for different ethnic groups represented in Canadian populations. This is based on analysis of the genetic disorders described in OMIM. In these three ethnic groups various disorders are specific to the groups, giving a sort of genetic risk profile of those groups. Those disorders which are shared between groups are unlikely to denote interbreeding, because analysis at the DNA level indicates that the genetic changes differ at the molecular level.

Table 2. Number of specific versus shared genetic disorders for three different ethnic groups in Canadian populations.

| | | | |
|---|---|---|---|
| Ashkenazic Jews | 39 (specific) | | |
| French Canadians | 9 | 68 (specific) | |
| Mennonites | 3 | 2 | 26 (specific) |
| | Ashkenazic Jews | French Canadians | Mennonites |

Amongst the three ethnic subsets in our population, the inherited genetic risks differ from one group to another. Shared diseases such as Bloom's syndrome or cystic fibrosis may reflect either the segregation of ancient genes from before the initiation of ethnic divisions, or separate new mutations. When we look at the molecular basis of the shared diseases, we often find that they differ from one group to another. Thus, even common diseases cannot be held to show common causes at the DNA sequence level in these ethnic subsets. The evidence for this conclusion is given in Table 3, in which the allelic variants of Tay-Sachs disease in two groups indicates totally different molecular, biological and social origins of these ethnically shared diseases.

Table 3. Common allelic mutations of Tay-Sachs disease in two identifiable subgroups of the Canadian population. Data abstracted from [3].

| | Ashkenazic Jews | French Canadians |
|---|---|---|
| Changes in DNA sequence | 73% - Base Pair Insertion<br>15% - G-to-C substitution<br>4% - GLY269 substitution by Serine<br>8% - others | 82% - 7, 6KB deletion, Southeastern Quebec, Gaspe<br>18% - ARG170TRP mutation, Quebec, Estrie<br>- G-to-A transition, Quebec, Saguenay-Lac<br>- others |

What this means is that if one were to look for an association of Tay-Sachs with any arbitrary environmental risk source in the Gaspe, using the population of Quebec as a reference, one would find a positive association. This association could be with anything - consumption of oysters, water sources, or any other differential. But in fact the risk source is intrinsic, and determined at the genetic level.

Consanguinity is a well recognized source of genetic risks in human families. The risk comes from the amount of shared DNA from common ancestors, and the odds that deleterious genes are in the shared DNA. In Table 4 data from a consanguinity survey indicates that a portion of genetic risk is related to consanguinity as such, so any ethnic groups or individual parents in which consanguineous marriage is common, further increase the genetic risks to their offspring. This is not entirely bad, as any group in which certain risk genes are lacking cannot pass these absent risks to their children.

Table 4. Number of OMIM entries associated with both consanguinity and ethnicity indicators and the distribution of specific versus shared disorders.

| | |
|---|---|
| Number of genetic disorders in OMIM associated with consanguinity. | 830 (14%) of OMIM |
| Number of genetic disorders in OMIM associated with both consanguinity and ethnicity indicators. | 352 (42%) |
| Number of genetic disorders specific to ethnic groups. | 211 (60%) |
| Number of genetic disorders shared by ethnic groups. | 141 (40%) |

What this means is that consanguinity as such increases genetic risks, and in combination with ethnicity leads to even greater risks to children. And if both these indicators reflect shared genetic material, and thus increased odds of sharing pre-existing genetic variants, this is not surprising. We conclude that the genetic risks shared among the members of a breeding group, however defined, are specific to that group and have no transference to the rest of the population.

So the picture we have of the Canadian population is of a collection of breeding groups sharing life-styles, locations, genes, and genetic risks among themselves, amidst a small proportion of individuals approaching true random breeding, at least genetically. This is not similar to the picture of our population as an interbreeding and homogeneous gene pool.

As an example of the application of such knowledge to risk assessment, two studies, one in Kuwait and one in the Shetlands, indicates that the risk of Down syndrome (trisomy 21) is much higher in the offspring of consanguineous marriages than in those between unrelated individuals. The relative risk for offspring, largely of second cousins or closer in Kuwait, is 4.3 after correction for maternal age, birth order, gravidity and previous reproductive history [8]. The children of second cousins, it should be pointed out, share only about 1/64 (1/8 from each grandparent) of their DNA in common. This is one clear example of the huge effects on risk which may be caused by consanguinity.

To reiterate: sharing 1.5% of genetic material in common in this particular group has given rise to a four-fold increased risk of Down syndrome. If such a group were to be located in a particular study area, comprising 10% of the population, this could give rise to a 1.4-fold excess relative risk.

We think that from this limited survey of genetic disorder OMIM it is clear that biological risk factors have major effects when considering disease aggregations in populations. These biological risk factors arise from the underlying structure of the population(s) at risk and need direct testing to determine their magnitude and effects.

**Analysis of the effects of ethnicity and consanguinity should be a primary concern after the identification of non-random aggregations of genetic or cancer risks by epidemiological means.**

For risk assessment, it should be pointed out that the above surveys include both dominant and recessive diseases. Dominant mutations are often thought to be new mutations, requiring special explanations such as newly induced mutation caused by the putative risk source. But if both consanguinity and ethnicity are associated with dominant disorders, they must reflect the segregation and expression of pre-existing variations from within the ethnic group or reproducing couple. This rules out new mutation as the major source of dominant diseases, reducing the uncertainty of the effects of external risk sources in risk assessment.

The classification of genetic disorders into such neat boxes as recessive, dominant and multifactorial is steadily eroding as our knowledge grows that almost all gene products interact with many other gene products [9]. If all proteins interact with another 10 to 35 other proteins, it is expected that all genetic disorders will be multifactorial in fact, despite classifications of convenience. In a while, with a few exceptions, truly recessive disorders will vanish. All mutations are likely to possess some features of dominance.

We have discussed some of the biological factors which are important in risk assessment. To evolve Molecular Epidemiology to a useful science in risk assessment, the effects of these factors need to be distinguished at the DNA level. How can this be done operationally? We will discuss one example which indicates the direction we wish to drive this analysis.

As we argued above, even when the genotype permits a genetic risk, not all people may express that risk. For example, acute myelogenous leukemia (AML) [10] is one outcome of Fanconi's anemia (FA) [11], which is observed in about 15% of all Fanconi's homozygotes. Yet from population genetic principles, the numbers of carriers must vastly outnumber the number of patients with AML. And the number of patients vastly outnumbers the expected number of homozygotes.

A testable prediction arising from our approach to molecular epidemiology is that some of the AML patients may reflect additional risk to carriers of FA gene variants, possibly in combination with some environmental etiological agent, such as a virus. We are starting an analysis of the Fanconi's Anemia complementation group C (FACC) [12] locus in AML patients. The hypothesis tested is that carriers of FA gene variants are more at risk than the rest of the population of developing AML. If this is true, then the associations between ionizing radiation and leukemia need to be re-examined in greater depth.

To test this connection direct analysis of FACC carriers in the population of AML patients will need to be performed. We are developing a non-cloning procedure for DNA analysis which may have advantages in diagnostic work, especially for new mutations. The preferred methods for analyzing genetic variants associated with risk, at the DNA level, will arise out of progress in the Human Genome Project.

Why will we want to do this? At the moment, we have very small uncertainties in our dose estimates, and very large uncertainties in our assessment of the concomitant risks. These uncertainties are made up largely of biological and genetic uncertainties. Biological uncertainties can be addressed by determining the actual patterns of breeding which give rise to ensuing generations. Genetic un-

certainties, and in fact the foregoing biological uncertainties as well, can be addressed by direct analysis of the DNA. By directing risk assessments to the DNA we focus on the underlying sources of our genetic and cancer risks. Until these methods are developed and validated, we will be unable to estimate the relative contributions of prior history to the risks which may be caused by ionizing radiation.

**Acknowledgements**: We wish to thank Norman Gentner for his helpful comments. This work was supported by the CANDU Owners Group.

**References**:

(1) Morton, N.E. 1982. Outline of Genetic Epidemiology. Basel, Karger, New York.

(2) McKusic, V.A. March 1994. On-line version of Mendelian Inheritance in Man. John Hopkins Univ. Press, Baltimore, MIM No. 180200.

(3) McKusic, V.A. March 1994. On-line version of Mendelian Inheritance in Man. John Hopkins Univ. Press, Baltimore, MIM No. 272800.

(4) McKusic, V.A. March 1994. On-line version of Mendelian Inheritance in Man. John Hopkins Univ. Press, Baltimore, MIM No. 278700, 133510, 278720, 278730, 278740, 278750, 278760, 278780, 278810.

(5) Person, P. et al. 1994. The status of online Mendelian in man (OMIM) medio. 1994. Nucleic Acids Research 22:3470-3473.

(6) Benson, D.A. et al. 1994. GenBank, Nucleic Acids Research 22:3441-3444.

(7) Emmert, D.B. et al. 1994. The European Bioinformatics Institute (EBI) databases. Nucleic Acids Research 22:3445-3449.

(8) Khoury, M.J., Beaty, T.H., Cohen, B.H. 1993. Fundamentals of Genetic Epidemiology, Oxford University Press, New York.

(9) Mayr, E. 1982. The Growth of Biological Thought. Harvard University Press, Cambridge.

(10) McKusic, V.A. March 1994. On-line version of Mendelian Inheritance in Man. John Hopkins Univ. Press, Baltimore, MIM No. 151385.

(11) McKusic, V.A. March 1994. On-line version of Mendelian Inheritance in Man. John Hopkins Univ. Press, Baltimore, MIM No. 227650, 227660, 227645, 227646.

(12) McKusic, V.A. March 1994. On-line version of Mendelian Inheritance in Man. John Hopkins Univ. Press, Baltimore, MIM No. 227645.