The X Chromosome and the Female Survival Advantage:

An Example of the Intersection between Genetics, Epidemiology and Demography.

Kaare Christensen [1], Karen-Helene Ørstavik [2], James W.Vaupel [3]

[1] The Danish Twin Registry, Epidemiology, Institute of Public Health, and the Danish Center for Demographic Research, University of Southern Denmark, DK-5000 Odense, Denmark.

[2] Department of Medical Genetics, National Hospital, Oslo, Norway.

[3] Max Planck Institute of Demographic Research, Rostock, D-18057, Germany

Corresponding author:

Kaare Christensen, MD, PhD

The Danish Twin Registry

University of Southern Denmark:

Main Campus Odense University

Sdr. Boulevard 23A

DK-5000 Odense C

Phone: +45 6550 3049

Fax: +45 6590 6938

E-mail: kchristensen@health.sdu.dk

Abstract

Despite differences in research traditions, the disciplines of genetics, epidemiology, and demography are becoming increasingly integrated in health related research. The enormous development within genetic technology with the possibility of genotyping thousands of variants from small samples of biological material obtained by non-invasive methods now makes it feasible to include genetic information in epidemiological and demographic studies. Simultaneously, new insight can be obtained from hybrids of methods and data from the three disciplines. This chapter will illustrate how a genetic observation combined with demographic insight and a modified genetic-epidemiological design (a twin study) provides evidence that part of the sex difference in survival can be attributed to the fact that females have two X-chromosomes and males have only one, a result which is of potential interest for genetics, epidemiology, and demography.

At a first glance a merging of genetics with epidemiology and demography seems difficult. Traditionally, genetics has primarily focused on rare genetic variants with strong effects on health, e.g. genetic diseases like Huntington's disease, and the number of individuals included in genetic studies has typically been less than 100. Epidemiologists have traditionally looked for a wide variety of environmental causes ("risk factors") for variations in health outcome and behavior, and the typical size of epidemiological studies is in the range of 100-10,000. While there is a considerable overlap between demographic household surveys and epidemiological studies, the more traditional demography has been focusing on large (often nationwide) studies including only a few key co-variates such as age, sex, race and socio-economic conditions. However, the development within genetics during the last decade combined with a greater integration of epidemiology and demography has brought the three disciplines closer.

With 30,000-40,000 genes now being identified, many of which are likely to have several variants, an enormous number of new explanatory variables are suddenly appearing in the arena of health related research. These "common genetic variants" are in nature often similar to the "risk factors" studied in epidemiology, i.e. the various common genetic variants are typically associated with a modest increase in risk and a new risk factor identified in one setting often proves difficult to corroborate in another setting. Hence, it is not surprising that epidemiologists have assimilated such common genetic variants into their analyses side-by-side with traditional risk factors such as

smoking, alcohol, diet etc. and with a special emphasis on the interaction between genes and environment.[1,2]

The rapid development in genetic techniques now makes it possible to obtain DNA from very small samples of body fluids or tissues with non-invasive methods – progress that has made inclusion of biological material in epidemiological studies feasible. A recent book from the National Research Council "Cells and Surveys: Should Biological Measures be Included in Social Science Research?" demonstrates that the techniques are now reaching a level where inclusion of genetic markers in very large studies is becoming possible. Hence, it seems likely that both epidemiology and demography will be able to include genetic factors as co-variates in the future. Simultaneously, new methods that integrate background demographic data into genetic-epidemiological studies of survival are emerging.[3] As illustrated by a concrete example in this chapter, the greater integration of genetics, epidemiology and demography are likely to generate new methods and insight from combining data and knowledge from all the three disciplines.

Why do females have lower mortality than males throughout life?
- a research question common for genetics, epidemiology and demography.

The male disadvantage in survival is a consistent finding in contemporary industrialized countries. The difference is so marked (a factor 1.5-2.0 throughout most of life) and so consistent that practically all demographic and

epidemiological studies of survival stratify the analyses by sex. However, understanding of why males have poorer survival compared to females is sparse.[4] Only a fraction of the difference can be explained by the often more hazardous lifestyle of males,[4] and it seems likely that also biological factors contribute to the different survival pattern. This is supported by the fact that the males' disadvantage begins in uterus.[5] One may even have expected females to have poorer survival than males due to their biological investment in reproduction [6] and based on the observation of a consistently lower mean physical performance than males throughout life.[7,8,9]

The genetic difference between males and females

The fundamental biological difference between males and females is to be found in the sex chromosomes: females have two X chromosomes, males have one X chromosome and a small Y chromosome (which directs the development into the male phenotype). Although females have two X chromosomes, only one of them is active in each cell of the body, the other is inactivated in early embryonic life and stays so throughout life. This means that females are a mosaic of two different cell lines: one cell line with the X chromosome from the father being active and another cell line with the X chromosome from the mother being active. Males have only one cell line because they receive only one X chromosome (from their mother) and one Y chromosome (from their father) (Figure 1).

It seems likely that the female scenario with two cell lines in all organs provides a health advantage, which is clearly the case for X-linked diseases, such as color blindness, hemophilia and Duchenne's disease.[10] These diseases caused by mutation on the X chromosome nearly always affect men, because

females have two cell lines (with different X chromosome active), and if just one of the X chromosomes is "healthy" this one cell line can prevent the expression of most X-linked diseases. Males with a "sick" X chromosome will be affected, as they have only one cell line and no "back-up" from a "healthy" cell line.

It seems plausible, not only in the case of such X-linked diseases but also for mutations with more subtle effects, that having two cell lines offers an advantage to females compared to males. This advantage will tend to be greatest in tissues with many cell divisions such as blood cells or mucosa cells. Furthermore, the predominance could be tissue-specific with one cell line being predominant in one organ due to better survival while the other cell line could be more frequent in another organ. Overall, this may contribute to the longer lifespan of females. This hypothesis has indirect support from animal lifespan studies. In mammals the male is heterogametic (XY) and has a shorter lifespan than the female. In birds it is the females that are heterogametic (ZW), and the available data suggest that male birds tend to live longer than female birds.[11] However, it has been difficult to test this hypothesis more directly in humans.

The genetic observation

For females, the inactivation of one of the X chromosomes in each cell occurs very early in embryonic life around the time where the embryo is only 16-32 cells. The inactivation seems to occur by a random process and usually results in roughly a 50:50 distribution. In rare cases females have just by chance a "skewed distribution" i.e. a predominance of one of the cell lines already from birth.[12] Cross-sectional studies have shown that among younger females it is

very rare to have a skewed distribution of X inactivation, while for females over age 60, more than a third of the females have a predominance of one of the cell lines in their blood, and among centenarian females the majority have a predominant cell line (Table 1).[12,13,14,15]

The observation that with age more females have a predominant cell line instead of about a 50:50 distribution of the two cell lines has led to two main hypotheses:

Hypothesis (i): The predominance of one cell line in many elderly females is a random event due to the fact that the number of stem-cells is small.

Hypothesis (ii): The predominance of one cell line in many elderly females is due to selection i.e. a growth or survival advantage conferred by one of the parental X chromosomes.

The random event hypothesis has some support from data from autologous marrow transplantation studies of female Safari cats heterozygous for glucose 6-phosphate dehydrogenase mutants (Safari cats are a good animal model because the two X can be discriminated). Mathematical modeling based on these data indicated that a predominance of one cell line can occur simply by chance when the number of blood stem-cells is small.[16] Chance events could act on many levels, e.g. stem-cell replication, apoptosis (cell death), and the initiation of differentiation or maturation. Depletion of stem-cells and random differentiation of the few residual stem-cells is a possible explanation.[12,13,16]

On the other hand, another study of female Safari cats, although only 11 cats were included, showed evidence that excessive age-related skewing could be due to a growth advantage conferred by one of the parental X chromosomes.[17] If this is the case it is likely that the X chromosome harbours a gene which affects cell survival and hence potentially overall survival.

However, animal data to test the two hypotheses are sparse and no human data had been used for testing the hypotheses.

The demographic parallel to the genetic observation

The genetic observation that there is a roughly 50:50 distribution of the two cell lines at birth, and that for most elderly women there is a predominance of one of the cell lines is an observation parallel to the sex ratio in a closed population: at birth there is about a 50:50 distribution of males and females, by age 85 there is a 1:2 male-female ratio in most developed countries which at age 100 has increased to at least 1:4.

In this case we know that it is not a random effect but a result of a selection against males. A general method, the "Survival-Attributes Assay", is being developed which aims at using cross-sectional data on "fixed traits" (gender, genotype, etc.). Based on such data it is possible to estimate the effect of a fixed trait on survival.

The Survival-Attributes Assay is a demographic method that can be illustrated by a simple example. Let N_{80} be the number of people aged 80. Let p_{80} be the proportion of 80-year-olds who have some fixed attributes, such as some genetic variant at the X chromosome. Let p_{100} be the proportion at age 100. Let s be the conditional survival probability from age 80 to age 100 for the elders who have the fixed attribute. Let S be the conditional survival probability from age 80 to age 100 for the entire cohort. Then because

$$p_{80} N_{80} s = N_{80} S p_{100},$$

it follows that

$$s = S\, p_{100} / p_{80}.$$

Suppose that 40% of a cohort has some fixed attribute at age 100 and that 3 per thousand of the cohort survived from age 80 to age 100. Furthermore suppose it can be estimated that 10% of the cohort had the attribute at age 80. Then the formula implies that 12 per thousand of the people with the attribute survived from 80 to 100. That is, the attribute quadrupled the chance of surviving from 80 to 100 compared with the average for the entire cohort. This is a very simple and useful method but it is based on some crucial assumptions. Most importantly because we rely on cross-sectional data, we have to assume that 80-year-olds today are similar to the centenarians when they were 80. From the formula, relative risks can readily be estimated, with or without assumptions about selectivity due to hidden heterogeneity.

To use the Survival-Attributes Assay in the case of cell line predominance we need a specific genetic variant. Before further pursuing this it would be desirable to get some evidence that the predominance of one cell line indeed was a selection process and not a random event.

Testing the hypotheses: The modified genetic-epidemiological design

Demographic methods did not allow us to distinguish between the two hypotheses. However, a study of monozygotic twins provided an opportunity. Traditionally, testing the overall influence of genetic factors on a trait variance requires both monozygotic and dizygotic twins. Investigations of monozygotic twins alone are usually made in studies of environmental factors where the effect of genetic factors are controlled for. However, in this case where each of

the two twins had two competing cell lines, the "monozygotic-twins-only-design" provided an opportunity to test the influence of genetic factors on cell line survival. If the often observed predominance of one of the two cell lines in peripheral blood in elderly females was determined by a stochastic process with no selection, one would expect little correlation in the X inactivation patterns between two monozygotic co-twins. A selection process based on X-linked genetic factors, on the other hand, would create a tendency for the same cell line to become predominant in two monozygotic co-twins.

The study results

We studied peripheral blood cells from elderly female monozygotic twin pairs.[15] The samples were obtained through The Longitudinal Study of Aging Danish Twins (LSADT) which comprises twins aged 73 or older in the nationwide Danish Twin Registry. In 1997, 2,172 individuals – 79% of the twins – were interviewed, regardless of whether their co-twin was alive or not.[6] A total of 71 monozygotic female twin pairs were available for X inactivation analysis. The sample of twin pairs was unselected in so far as no cases were excluded due to diseases. The X chromosome inactivation pattern was determined by a PCR analysis of a polymorphic CAG repeat in exon 1 of the androgen receptor gene.[18]

Table 1 shows that the elderly monozygotic twins did not differ from the elderly singletons in terms of overall X inactivation patterns ($p>0.5$) and that, as expected, elderly twins and singletons had a significantly more skewed pattern than younger individuals ($p<0.01$).

The intraclass correlation within monozygotic twin pairs for percentage of inactivation of an X chromosome was 0.57 ($p<0.01$). When the 8 outliers (see

Figure 2) were excluded from the analyses, the intraclass correlation rose to 0.84 ($p<0.01$). Some of these outliers most likely result from non-concordant X inactivation: Monteiro *et al*. [19] found that 13% (3/23) of monozygotic twin pairs aged 0-24 had large absolute differences in X inactivation pattern. If there is a similar selection for the two cell lines in such pairs, they can remain discordant for X inactivation pattern throughout life.

The strong tendency for the same blood cell line to become predominant in elderly female monozygotic co-twins provides evidence that human stem-cell kinetics is influenced by X-linked genetic factors. Furthermore, our finding of an increased prevalence of skewed X inactivation among centenarians compared to 73-93 year olds ($p<0.01$, Table 1) is compatible with a selection process depending on X-linked genetic factors. The first preliminary follow-up results from this study confirms that the presence of two cell lines is associated with a better survival among elderly females.

On this background we are now using sibs studies (elderly female dizygotic twins) to identify specific genetic markers of importance for cell survival. If we identify such genetic markers, we intend to use the Survival-Attributes Assay to quantify the effect of the genetic variant on a population level. If X-linked genetic factors that lead to a better survival for a blood cell line can be determined, this will have major implications for marrow transplantation and gene therapy on hematopoietic stem-cells. Also it may give us insight into some of the biological mechanisms behind the sex difference in survival.

Comments

This research on X inactivation in elderly monozygotic female twins lead to new insight in cell survival. The starting point was an observation of

increasing prevalences of skewed X inactivation among elderly and analytically parallel scenarios in demography (Figure 3). We speculate, based on this study and animal studies, that X inactivation may also play a role for overall survival. If so, and if we succeed in identifying specific genetic factors influencing (cell) survival by using demographic methods, then this may have an influence not only on the fathering disciplines of this study, but maybe also on medicine.

Reference List

1. Khoury, M. J. 1998. Genetic and epidemiologic approaches to the search for gene-environment interaction: the case of osteoporosis [editorial; comment]. Am. J. Epidemiol. **147:** 1-2.

2. Christensen, K. *et al.* 1999. Oral clefts, transforming growth factor alpha gene variants, and maternal smoking: a population-based case-control study in Denmark, 1991-1994. Am. J. Epidemiol. **149:** 248-255.

3. Yashin, A. I. *et al.* 1999. Genes, demography, and life span: the contribution of demographic data in genetic studies on aging and longevity. Am. J. Hum. Genet. **65:** 1178-1193.

4. Kraemer, S. 2000. The fragile male. BMJ **321:** 1609-1612.

5. Hansen, D., H. Moller, and J. Olsen. 1999. Severe periconceptional life events and the sex ratio in offspring: follow up study based on five national registers. BMJ **319:** 548-549.

6. Christensen, K. *et al.* 1998. A tooth per child? [letter] Lancet **352:** 204

7. Montoye, H. J. and D. E. Lamphiear. 1977. Grip and arm strength in males and females, age 10 to 69. Res. Q. **48:** 109-120.

8. Christensen, K. *et al.* 2000. Genetic and environmental influences on functional abilities in Danish twins aged 75 years and older. J. Gerontol. A Biol. Sci. Med. Sci. **55:** M446-M452.

9. Nybo, H. *et al.* 2001. Functional capacity and self-rated health in 2,262 nonagenarians - the Danish 1905-cohort survey. J. Am. Geriatr. Soc. **in press**.

10. Puck, J. M. and H. F. Willard. 1998. X inactivation in females with X-linked disease [editorial; comment]. N. Engl. J. Med. **338:** 325-328.

11. Payevsky, V. A. *et al.* 1997. Sex-specific survival rates in birds. Zhurnal Obschei Biologii 5-5.

12. Busque, L. *et al.* 1996. Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. Blood **88:** 59-65.

13. Gale, R. E. *et al.* 1997. Acquired skewing of X-chromosome inactivation patterns in myeloid cells of the elderly suggests stochastic clonal loss with age. Br. J. Haematol. **98:** 512-519.

14. Champion, K. M. *et al.* 1997. Clonal haemopoiesis in normal elderly women: implications for the myeloproliferative disorders and

myelodysplastic syndromes. Br. J. Haematol. **97:** 920-926.

15. Christensen, K. *et al.* 2000. X-linked genetic factors regulate hematopoietic stem-cell kinetics in females. Blood **95:** 2449-2451.

16. Abkowitz, J. L., S. N. Catlin, and P. Guttorp. 1996. Evidence that hematopoiesis may be a stochastic process in vivo. Nat. Med. **2:** 190-197.

17. Abkowitz, J. L. *et al.* 1998. An X chromosome gene regulates hematopoietic stem cell kinetics. Proc. Natl. Acad. Sci. U. S. A **95:** 3862-3866.

18. Allen, R. C. *et al.* 1992. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. Am. J. Hum. Genet. **51:** 1229-1239.

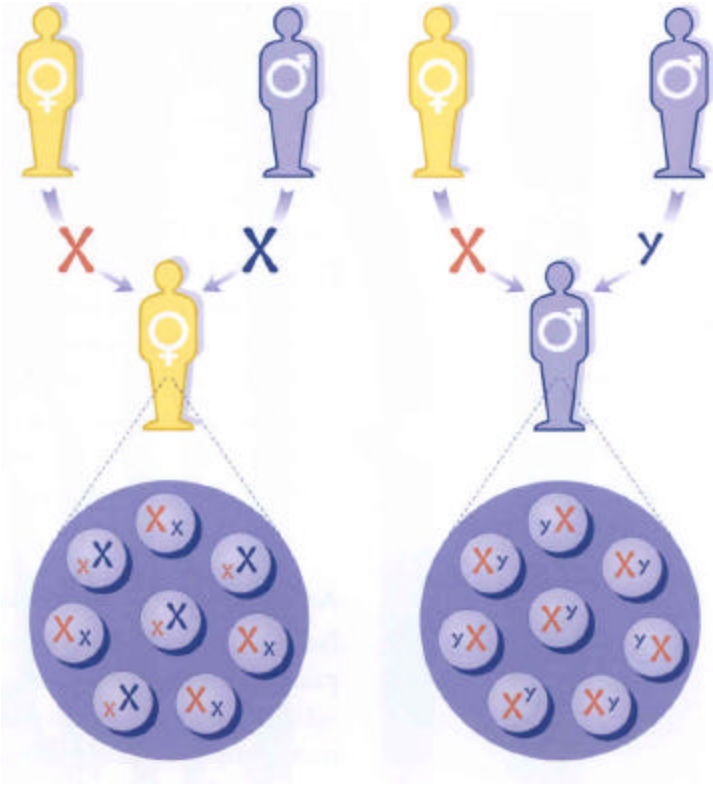
19. Monteiro, J. *et al.* 1998. Commitment to X inactivation precedes the twinning event in monochorionic MZ twins [see comments]. Am. J. Hum. Genet. **63:** 339-346.

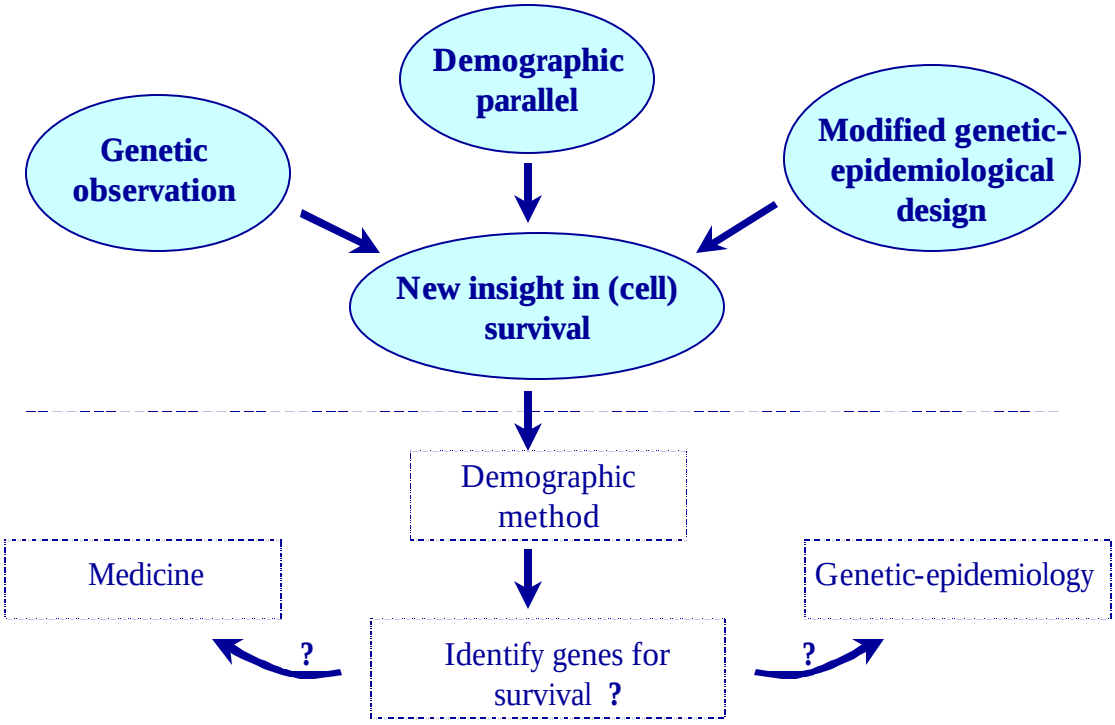
Legends

Figure 1. The sex chromosomes are the fundamental difference between males and females. Males have in all their cells an active X chromosome and an active Y chromosome. Although females have two X chromosomes, only one of them is active in each cell of the body, the other is inactivated in early embryonic life and stays so throughout life. This means that females are a mosaic of two different cell lines: one cell line with the X chromosome from the father being active and another cell line with the X chromosome from the mother being active.

Figure 2. X inactivation patterns in female monozygotic twins aged 73-93. Percentage of inactivation of an X chromosome (the measurements are truncated at 5% and 95%). [15]

Figure 3. A schematic presentation of the current study. The completed part of the study is shown above the dotted line, ongoing efforts and potential influences below.


X
X
X
y


Demographic parallel
Genetic observation
Modified genetic-epidemiological design
New insight in (cell) survival
Demographic method
Medicine
Genetic-epidemiology
?
Identify genes for survival ?
?

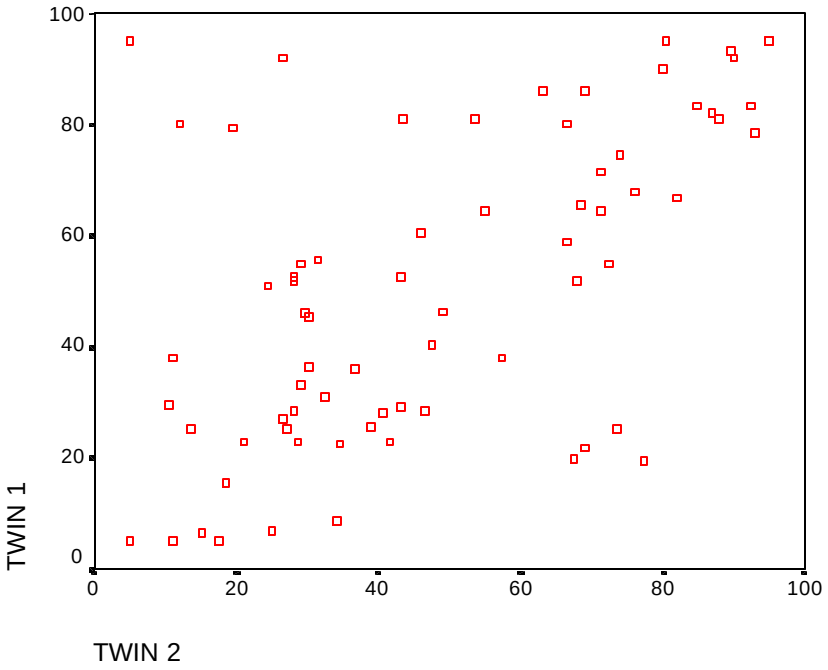

100
80
60
40
20
0
TWIN 1
0
20
40
60
80
100
TWIN 2

Table 1. Distribution of X inactivation patterns in Scandinavian twins and singletons. [15]

	Blood Donors	Elderly Singletons	Elderly Twins	Centenarians
	148	43	142	33
e range	19-65	83-101	73-93	101
nactivation pattern (% of individuals)				
Random (most common X ? [50%; 80%[)	93	65	65	33
Skewed (most common X ? [80%; 95%[)	7	26	27	49
Extremely skewed (most common X ? [95%; 100%])	0	9	8	18