Estimating heritability of different types of cancer

Isabella Locatelli

Istituto di Statistica e Matematica Applicata alle Scienze Umane, Università degli Studi di Torino, Italia (isabella.locatelli@uni-bocconi.it)

Andreas Wienke

Institute of Medical Epidemiology, Biostatistics and Informatics, Medical Faculty, Martin-Luther-University Halle-Wittenberg (andreas.wienke@medizin.uni-halle.de)

Paul Lichtenstein

Karolinska Institutet, Stockholm, Sweden (Paul.Lichtenstein@mep.ki.se)

1. Introduction

The aim of this study is to investigate the role of genes and environment in determining the individual propensity to develop three different kinds of cancers, namely breast, prostate and colon cancer. Our study is mainly concerned with so-called *frailty models*, which represent a special area of survival analysis and are characterized by the introduction of an unobserved heterogeneity term (the frailty) into the usual multiplicative structure of a survival analysis model (Vaupel *et al*. 1979). We adopt here a particular frailty model, which is called *correlated frailty model*. This is a bivariate model, allowing to estimate, in addition to the variance of the unobserved heterogeneity term, the correlation coefficient between frailties of related individuals (Yashin *et al.* 1995). When the model is applied to twin data, an estimate of the correlation between cotwins' frailties is available for both monozygotic and dizygotic twins. In this case, it is possible to adopt genetic arguments (namely, quantitative genetic equations) in order to give a genetic-environmental decomposition of the frailty variance and, by consequence, an estimate of *heritability* in the individual susceptibility towards a disease (Yashin and Iachine 1995, Wienke *et al.* 2003, Locatelli *et al.* 2004).

2. The data

In this analysis we use cancer data from the Swedish Twin Registry (Lichtenstein *et al.* 2002). First established in the late 1950s to study the importance of smoking and alcohol consumption on cancer and cardiovascular diseases whilst controlling for genetic propensity to disease, it has today developed into a unique source. Since its establishment, the Registry has been expanded and updated on several occasions, and the focus has similarly broadened to most common complex diseases.

At present, the Swedish Twin Registry contains information about two cohorts of Swedish twins referred to as the 'old' and the 'middle' cohort. The old cohort consists of all same-sexed pairs born between 1886 and 1925 where both members in a pair were living in Sweden in 1959. In 1970 a new cohort of twins born between 1926 and 1967, the middle cohort, was compiled. We have included only the old cohort in our analysis and looked at a total of 5857 pairs of female twins and 4618 pairs of male twins.

3. Statistical methods

The approach adopted in our study in order to investigate the role of genetic and environmental factors in the propensity to develop a breast, prostate or colon cancer is an interdisciplinary approach. This method consists in merging models developed in the field of survival analysis, that is frailty models, and models coming from quantitative genetics and epidemiology, i.e. genetic models. In particular, we apply a bivariate frailty model – the so-called *correlated frailty model* – and, in order to come to a genetic interpretation of the results, we use the special information deriving from twin data.

The correlated frailty model is based on two main assumptions. First, the frailty variable is supposed to act multiplicatively on the baseline hazard. That is, the hazard for an individual $\mu(x)$ can be expressed by the product between the baseline hazard $\mu_0(x)$ and the frailty Z. Second, the duration times X_1 and X_2 of the two individuals in a pair are supposed to be conditionally independent giving the frailty variables Z_1 and Z_2.

Typical models of quantitative genetics can easily be incorporated into the correlated frailty model described above. Quantitative genetics models (Falconer 1990) are based on the decomposition of a phenotypic trait into a sum of different components, which are supposed to be independent. Using this approach, it is possible to estimate the proportion of the total variability of the phenotype which is related to genetic factors. This proportion is defined as the 'heritability' of the phenotype. In particular, when we work with duration data (e.g. the age at the onset of a disease) a heritability estimate can be calculated by identifying the phenotype with the duration variable (McGue *et al*. 1993). The problem which often arises in this context is that observations of that variable are in general censored.

We follow here the definition of heritability given by Yashin and Iachine 1995. They suggest an approach based on the frailty variable instead of the duration variable itself. The phenotype is thus identified with the unobserved heterogeneity term. With this approach the problem of censoring in the estimate of heritability does not arise. This is because heritability is calculated as a function of the correlation coefficient between cotwins' frailties - estimated via application of a correlated frailty model - instead of the correlation between observed duration times. In this context, heritability is defined as the proportion of the total variability of frailty explained by genetic factors and it is thus obtained via decomposition of the frailty variance.

4. First results and work in progress

The model described above has been applied to four different datasets from the Swedish Twin Registry. Limiting the analysis to the old cohort of the Registry, we consider (female) data about the onset of a breast cancer, (male) data about the onset of a prostate cancer, and separately female and male data containing information about the onset of a colon cancer. We observe 715 cases of breast cancer and 373 cases of colon cancer above the 5857 pairs of female twins; 703 cases of prostate cancer and 304 cases of colon cancer above the 4618 pairs of male twins.

We adopt here the genetic decomposition of frailty which includes three different effects: an additive genetic, a dominance genetic and an uncommon environmental effect (ADE model). What we found is that, in all the four cases, environmental effects are more important than genetic effects for cancer susceptibility. Factors related with the environment seem to have the principal role in causing sporadic cancer, although the relatively large heritability, especially for prostate and colon cancers, confirms the importance of focusing on the genetic aspects of these diseases.

In further developments, we wish to continue the study by extending the described model in different directions. First, results obtained by assuming different frailty distributions, in particular gamma and lognormal, will be compared; second, various genetic models, besides the ADE model, will be applied to the four datasets, especially with the aim of testing for the existence of two different environmental effects: shared and not shared by the two twins in each pair; third, we wish to verify if results can change substantially when we introduce into the model the possibility that a fraction of the population under study is not susceptible to experience the event of interest (long-term survivors).

References

Falconer, D.S. (1990) Introduction to Quantitative Genetics, Longman Group, New York.

Lichtenstein, P., de Faire, U., Floderus, B., Svartengren, M., Svedberg, P., Pedersen, N.L. (2002) The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. Journal of Internal Medicine 252, 184 - 205.

Locatelli, I., Lichtenstein, P., Yashin, I.A. (2004) The heritability of breast cancer: a Bayesian correlated frailty model applied to Swedish twins data. Twin Research 7, 182 - 191.

McGue, M., Vaupel, J.W., Holm, N., Harvald, B. (1993) Longevity is moderately heritable in a sample of Danish twins born 1870 - 1880. Journal of Gerontology: Biological Sciences, B 48: B237 - B244.

Vaupel, J.W., Manton, K.G., Stallard, E. (1979) The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. Demography 16, 439 – 454.

Wienke, A., Lichtenstein, P., Yashin, A.I. (2003) A bivariate frailty model with a cure fraction for modeling familial correlations in diseases. Biometrics 59, 1178 - 1183.

Yashin, A.I., Iachine, I.A. (1995) Genetic Analysis of Durations: Correlated Frailty Models Applied to Survival of Danish Twins. Genetic Epidemiology 12, 529 - 538.

Yashin, A.I., Vaupel, J.W., Iachine, I.A. (1995) Correlated Individual Frailty: an Advantageous Approach to Survival Analysis of Bivariate Data. Mathematical Population Studies 5, 145 – 159.