

A Correlated Frailty Model with Long-Term Survivors for Estimating the Heritability of Breast Cancer

*Un Modello Correlated Frailty con Long-Term Survivors per la Stima
della Componente Genetica del Tumore al Seno*

Isabella Locatelli

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

Alessandro Rosina

Institute of Population and Geographical Studies,
Università Cattolica, Milano, Italia (alessandro.rosina@unicatt.it)

Paul Lichtenstein

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

Anatoli I. Yashin

Max Planck Institute for Demographic Research, Rostock (yashin@demogr.mpg.de)

Riassunto: Allo scopo di determinare il ruolo giocato da fattori genetici e ambientali nell'influenzare la propensione a contrarre un tumore al seno ricorriamo a modelli di analisi della sopravvivenza con *frailty*, applicati ai dati su coppie di gemelle. Una stima della componente genetica della *frailty* individuale nei confronti della malattia, ottenuta introducendo nel modello considerazioni provenienti dalla genetica quantitativa, è fornita anche nel caso in cui si tenga conto dell'eventualità che vi siano, nella popolazione studiata, dei soggetti immuni (*unsusceptible fraction, long term survivors*). Il modello è espresso in un contesto Bayesiano e le stime dei parametri sono ottenute con l'ausilio di tecniche MCMC.

Keywords: Frailty models, correlated frailty, unsusceptible fraction, quantitative genetics

1. Introduction

The aim of this study is to investigate the role of genes and environment in determining the individual propensity to develop a breast cancer, and calculate an estimate of *heritability* in the propensity to develop this disease. 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, and thus an estimate of the

correlation between cotwins' frailties is available for both monozygotic and dizygotic twins, it is possible to use genetic arguments in order to give a genetic-environmental decomposition of the frailty variance and, by consequence, an estimate of *heritability* in the individual susceptibility towards the disease (Yashin and Iachine 1995).

In Section 2 we will present the data used in our analysis. More details about the statistical methods can be found in Section 3. In particular we will discuss the introduction into the model of the possibility that a fraction of the population under study is not susceptible to experience the event. Working with an event which is not necessarily experienced by all individuals in the population, as the case of the onset of a disease, it could be important to take into account the possibility that some individuals in the population are actually not susceptible to it. Does the introduction of an *unsusceptible fraction* lead to different results in terms of frailty variance and correlation and in terms of the estimated heritability? In the last section we will briefly present some results which can somehow represent an answer to this question.

2. The data

In this analysis we use breast cancer data from the Swedish Twin Registry. This one 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 both cohorts in our analysis and looked at a total of 12568 pairs of female twins. For a comprehensive description of the Swedish Twin Registry database, with a focus on the recent data collection efforts and a review of the principal findings that have come from the Registry, see Lichtenstein et al. (2002).

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 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 .

This two assumptions lead to the following expression of the conditional likelihood of the model:

$$L(\mathbf{X}, \Delta | \mathbf{Z}) = \prod_{i=1}^n \prod_{j=1,2} [Z_{ij} \mu_0(x_{ij})]^{\delta_{ij}} \exp(-Z_{ij} H_0(x_{ij}))$$

where H_0 indicates the cumulative baseline hazard; x_{ij} , δ_{ij} and Z_{ij} represent the duration time, the censoring indicator and the frailty variable for the individual j ($j = 1, 2$) in the pair i ($i = 1, \dots, n$), respectively; and the vector of frailties (Z_{i1}, Z_{i2}) is supposed to follow some bivariate distribution, whose parameters are estimated by the model together with parameters of the baseline hazard. For identifiability reasons, the mean of the frailty variables is supposed to be equal to one. We do not take into account observed covariates information because we are especially interested in estimating frailty parameters and giving them a genetic interpretation.

The correlated frailty model can easily be accommodated in order to incorporate the possibility that a fraction of the population under study is not susceptible to experience the event. In the literature the so-called *unsusceptible fraction* (or *long term survivors*) has been incorporated into the analysis of duration data with the use of *mixture models*, combining a logistic formulation for the probability of experiencing the event of interest (i.e. of being susceptible) with a survival model describing the time of occurrence (given that the individual is susceptible). Farewell (1982) estimates such a mixture model assuming a Weibull shape of the baseline hazard for those who are susceptible to experience the event. McDonald and Rosina (2001) propose a mixture model that combines a discrete-time survival model (with constant baseline hazard) with a logistic regression model for the probability of never experiencing the event of interest. They also introduce an unobserved heterogeneity term (frailty) in their discrete-time event-history model. Chatterjee and Shih (2001) give an extension in the bivariate setting, estimating a shared frailty model with an unsusceptible fraction. Wienke *et al.* (2003) provide a further generalization with a correlated gamma frailty model for those who are susceptible to experience the event.

As in the paper of Wienke *et al.* (2003), we adopt here a *correlated frailty-mixture model*, in order to give an estimate of the frailty variance and correlation, taking into account the unsusceptible fraction. However, in our work the model is specified in a Bayesian framework and estimated with the help of Markov Chain Monte Carlo techniques instead of the maximum likelihood method applied by the cited authors. We also adopt a different hypothesis on the frailty distribution, which, in our case, is assumed to be a bivariate lognormal distribution with vector of means equal to one. With respect to the gamma assumption, the lognormal distribution is more flexible when one is interested to specify different - but correlated - variables, as in the case of a correlated frailty model (Locatelli *et al.* 2003). Finally, a genetic decomposition of the frailty variance is given via application of quantitative genetics models, in order to investigate how the introduction of the unsusceptible fraction into the model affects the estimate of heritability in the propensity to develop a breast cancer.

4. Results

When we do not take into account the unsusceptible fraction, results give a very large estimate of the population heterogeneity ($\sigma = 6.7$), relatively small correlations between cotwins' frailties – around 0.3 for monozygotic and 0.1 for dizygotic twins – and a

heritability estimate around 30% of the total variability of frailty (Locatelli *et al.* 2003). It is interesting to observe that such results, especially in terms of the heterogeneity and correlation estimate, differ drastically from the ones which have been obtained in several works on overall mortality (Yashin *et al.* 1999). This could mean that the structure of frailty is completely different when the propensity to develop a specific disease instead of the risk of death is analyzed. Nevertheless, the unsusceptible fraction idea can represent another explanation to the different behavior of the estimates when one is working with an event, like the onset of a disease, which is not necessarily experienced by all individuals in the population. When the unsusceptible fraction is introduced into the model, we obtain a smaller heterogeneity estimate ($\sigma = 5.4$), much larger correlation estimates (around 0.8 for monozygotic and 0.6 for dizygotic twins) and an estimate of heritability which is more than 50 % of the total variability of frailty. These results show that, taking into account the possibility that a fraction of a population is 'immune' to the event, remaining individuals are estimated to be less heterogeneous in their propensity to develop the disease; within a pair, the two twins show a higher correlation in their susceptibility towards breast cancer; the role played by genetic factors in determining the individual propensity to develop the disease becomes more important.

References

- Chatterjee, N., Shih, J. (2001) A bivariate cure-mixture approach for modeling familial association in diseases. *Biometrics* 57, 779 – 786.
- Farewell V.T. (1982) The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* 38, 1041 – 1046.
- 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, A.I (2003) A Bayesian Correlated Frailty Model Applied to Swedish Breast Cancer Data. Rostock, MPIDR Working Paper WP-2003-025 (to be published in *Twin Research*, April 2004).
- McDonald, J.W., Rosina, A. (2001) Mixture modelling of recurrent event times with long-term survivors: Analysis of Hutterite birth intervals. *Statistical Methods & Applications* 10, 257 – 272.
- 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 (accepted by *Biometrics*).
- 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.
- Yashin, A.I., Iachine, I.A., Harris, J. (1999) Half of the Variation in Susceptibility to Mortality is Genetic: Findings from Swedish Twin Survival Data. *Behavior Genetics* 29, 11 – 19.