

Genetics of Disability and Other **Chronic** Conditions:  
A Bivariate Model of Debilitation and Survival.

Ivan A. Iachine, Kaare Christensen, James W. Vaupel  
and Anatoli I. Yashin

Published in transactions of *Symposium i Anvendt Statistik*,  
25-27 January 1999, Copenhagen, Denmark, pp.96-105.

# Genetics of Disability and Other **Chronic** Conditions: A Bivariate Model of Debilitation and Survival

Ivan A. Iachine†, Kaare Christensen, James W. Vaupel  
and Anatoli I. Yashin

†Department of Statistics and Demography,  
Odense University  
e-mail: i-iachine@win-chs.ou.dk

## Abstract

In this paper a model of dependent health histories appropriate for genetic analysis of disability and other **chronic** conditions using cross-sectional survey data on twins is developed. The model allows us to handle selective drop-outs by including information on twin pairs with deceased individuals in the analysis. This information is usually ignored in traditional genetic analysis of such data. The approach allows for evaluation of genetic effects on the age at onset of disability. The properties of the new and the traditional approaches are compared. It is shown that the presence of genetic effects revealed in the traditional analysis of health status data may have nothing to do with transition to disability state. The results of a simulation study are discussed.

## 1 Introduction

The first wave of the Longitudinal Study of Aging in Danish Twins (LSADT) was completed in Denmark in 1995 (Christensen et al. [2]). In this study information on activity of daily living (ADL) for twins who are alive and older than 75 is collected. Many ADL variables measured in the study describe the individual as being "healthy" or "disabled" with respect to some functional ability. The goal of this study is to better understand the relative importance of genetic and environmental factors on functional disability and other **chronic** conditions.

Traditional methods of genetic analysis of categorical data are based on liability models (Neale and Cardon [6]). Analysis of ADL data using liability models with age-dependent thresholds indicates the presence of genetic influence on disability (Yashin et al. [9]). Does it mean that free of disability life span is heritable? In this paper we show that traditional methods cannot be used to

address this question since they are unable to identify the source of the correlation in health status data. As a consequence, the results of genetic analysis may be misinterpreted. For example, non-zero heritability estimates obtained by traditional liability analysis of disability may have nothing to do with genetic influence on the hazard of transition from the "healthy" to the "disabled" state. In addition, traditional methods ignore information on "broken" twin pairs (pairs with one deceased twin individual) and on pairs where both twins are deceased. This information may be crucial in the identification of the stage of the aging process where genetic influence plays a substantial role (e.g. disability or mortality transition).

To avoid these limitations, we suggest an approach based on the model of "correlated health histories". This model extends the idea of correlated frailty (Yashin and Iachine [8]) used in bivariate (multivariate) survival to the case of multistate transitions. The properties of this model are investigated. In particular, it is shown that the same correlation in the cross-sectional health status data measured in the traditional analysis may be generated either by correlations in hazards of transition to disability states or by dependence between mortality rates among individuals in a twin pair. A new parameter estimation procedure is suggested which allows for evaluation of genetic influence on transition rates. The procedure is tested on simulated data. The results show that parameters of the health history model for related individuals can be identified from cross-sectional ADL-type data.

## 2 Methods for Analysis of ADL Data

### 2.1 Health Histories and Current Status Data

Consider a model in which an individual at any time can be in one of the three states: *Healthy*, *Disabled* or *Dead*, the last one being a terminal state. Define a *health history* of an individual as a random process  $\{X(t)\}_{t \geq 0}$  such that:

$$X(t) = \begin{cases} 1, & \text{if the individual is } \textit{Healthy} \text{ at age } t \\ 2, & \text{if the individual is } \textit{Disabled} \text{ at age } t \\ 3, & \text{if the individual is } \textit{Dead} \text{ at age } t \end{cases} \quad (1)$$

A bivariate health history is used to describe health status trajectories a pair of related individual (e.g. twins). It can be defined as a bivariate random process  $\{X_1(t), X_2(t)\}_{t \geq 0}$  where  $X_i(t)$  is the health history of the  $i^{\text{th}}$  twin,  $i = 1, 2$ .

Complete observations of  $\{X_1(t), X_2(t)\}_{t \geq 0}$  may be obtained in a follow-up study. Health history information collected in a cross-sectional study is available in the form of *current status data* (Keiding [4]). For a twin cohort at age  $t$  this data is represented by single observation of the pair  $(X_1(t), X_2(t))$  that may be described by the *current status distribution* defined by:

$$P_{ij}(t) = P(X_1(t) = i, X_2(t) = j), \quad i, j = 1, 2, 3, \quad t \geq 0 \quad (2)$$

## 2.2 Liability Models for Current Status Data

Traditional methods of genetic analysis do not provide a way of handling selective drop-outs caused by non-survival, a limitation in cases where the disability status is a mortality risk factor. Instead, only health status data on the so-called "intact" twin pairs (i.e. where both twins are alive at the time of the study) is included in the analysis, which is focused on the properties of the conditional distribution:

$$\pi_{ij}(t) = P(X_1(t) = i, X_2(t) = j | X_1(t) \neq 3, X_2(t) \neq 3) \quad (3)$$

for  $i, j = 1, 2, t \geq 0$ .

For this purpose liability models are often used (Neale and Cardon [6]). In the case of twin data a pair of random variables  $(Y_1, Y_2) \sim BVN(0, 0, 1, 1, \rho)$  is introduced, where  $Y_i$  represents the liability of the  $i^{\text{th}}$  twin,  $i = 1, 2$ . The discrete trait  $X_i(t)$  (conditional on survival of the pair to age  $t$ ) is assumed to be related to the liability variable  $Y_i$  by means of a threshold  $y_0$ . When  $Y_i < y_0$ , the  $i^{\text{th}}$  individual is *Healthy*, otherwise the individual is *Disabled*. The correlation in liability  $\rho$  is used as an association measure between the conditional health status variables of the two twins. An age-specific estimate of the correlation in liability  $\hat{\rho}(t)$  may be obtained from the analysis of current status data on "intact" pairs. Alternatively, one might assume a common value of  $\rho$  for all  $t \geq 0$  and use the concept of age-dependent thresholds (Yashin et al. [9]) to obtain a single estimate  $\hat{\rho}$ .

The magnitudes of genetic and environmental influence on the trait may be estimated using data on identical (MZ) and fraternal (DZ) twins by comparing the correlation coefficients in liability  $\hat{\rho}_{MZ}$  and  $\hat{\rho}_{DZ}$  estimated for MZ and DZ twins respectively. In particular, the hypothesis of no genetic influence on disability  $H_0 : \rho_{MZ} = \rho_{DZ}$  may be tested versus the alternative  $H_A : \rho_{MZ} > \rho_{DZ}$ .

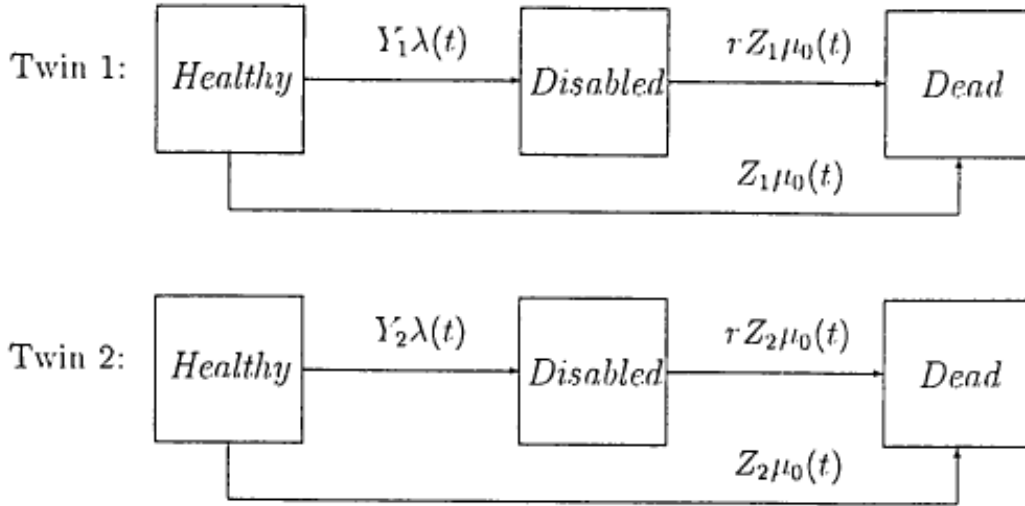
Based on additive decompositions of liability into additive genetic and environmental components one may define  $h^2$  as a percentage of variation in liability that is associated with the additive genetic component, also called *narrow-sense heritability* (Neale and Cardon [6]). An approximate heritability estimate may be computed as  $\hat{h}^2 = 2(\hat{\rho}_{MZ} - \hat{\rho}_{DZ})$  if  $\hat{\rho}_{MZ} < 2\hat{\rho}_{DZ}$ . The heritability estimate  $\hat{h}^2$  is frequently used as a measure of relative importance of genetic effects on the disability trait compared to the environmental effects.

## 2.3 Conditional Markov Frailty Models

A new approach for the analysis of twin ADL data suggested in this paper is based on a direct modeling of the bivariate random process  $\{X_1(t), X_2(t)\}_{t \geq 0}$  using dependent transition intensities.

In the following we assume that the debilitating transition *Healthy*  $\rightarrow$  *Disabled* is not reversible. Let  $Y_1, Y_2, Z_1, Z_2$  be non-negative random variables. Assume that given  $Y_i, Z_i, i = 1, 2$  the processes  $X_1(t)$  and  $X_2(t)$  are conditionally Markov and independent and such that  $X_i(t), i = 1, 2$  has conditional transition intensities

Figure 1: Bivariate Health History Model.



that follow a proportional hazards model, i.e.:

$$\begin{cases} Y_i\lambda(t) & \text{is the transition rate } \textit{Healthy} \rightarrow \textit{Disabled} \\ Z_i\mu_0(t) & \text{is the transition rate } \textit{Healthy} \rightarrow \textit{Dead} \\ rZ_i\mu_0(t) & \text{is the transition rate } \textit{Disabled} \rightarrow \textit{Dead} \end{cases} \quad (4)$$

for some baseline hazard functions  $\lambda(t)$ ,  $\mu_0(t)$  where  $\Lambda(t) = \int_0^t \lambda(u)du$  and  $H(t) = \int_0^t \mu_0(u)du$  are respective baseline cumulative hazard functions and  $r > 0$  is the conditional relative risk associated with the *Disabled* state. The multistate model may be represented graphically as shown in Figure 1.

Here the random variables  $Y_i, Z_i$  represent the individual susceptibility or frailty (Vaupel et al. [7]) to debilitation and mortality respectively for  $i^{\text{th}}$  twin,  $i = 1, 2$ . The conditional independence assumption is motivated by the following idea: the dependence between health status variables of two related individuals can only be explained by common genes or shared environmental factors influencing the health state. Consequently, if the respective genetic or shared environmental factors are observed the health histories of the related individuals become independent. A bivariate health history model based on these assumption which did not include a possibility for the transition *Healthy*  $\rightarrow$  *Dead* was suggested by Iachine et al., [3].

Define the conditional current status probability as:

$$P_j(t|Y_i, Z_i) = P(X_i(t) = j|Y_i, Z_i), \quad i = 1, 2, \quad j = 1, 2, 3, \quad t \geq 0 \quad (5)$$

Assuming that  $P_1(0|Y_i, Z_i) = 1$  (i.e. all individuals are *Healthy* at age  $t = 0$ ) the expressions for  $P_j(t|Y_i, Z_i)$  can be obtained by solving the system of Kolmogorov

equations associated with the conditional Markov process  $X_i(t)$ . The solution is:

$$\begin{cases} P_1(t|Y_i, Z_i) = e^{-Y_i\Lambda(t) - Z_iH(t)} \\ P_2(t|Y_i, Z_i) = \int_0^t e^{-Y_i\Lambda(u) - Z_i(H(u) + r(H(t) - H(u)))} Y_i\lambda(u) du \\ P_3(t|Y_i, Z_i) = 1 - P_1(t|Y_i, Z_i) - P_2(t|Y_i, Z_i) \end{cases} \quad (6)$$

Using the conditional independence assumption the probabilities  $P_{ij}(t)$  can be calculated as:

$$P_{ij}(t) = E[P_i(t|Y_1, Z_1)P_j(t|Y_2, Z_2)], \quad i, j = 1, 2, 3, \quad t \geq 0 \quad (7)$$

Using (6) the marginal current status probabilities (7) can be expressed in terms of the joint Laplace transform of  $Y_1, Y_2, Z_1, Z_2$  and its derivatives (Aalen [1]).

Questions about genetic influence on the frailty variables may be addressed by analyzing the  $4 \times 4$  covariance matrices of  $Y_1, Y_2, Z_1, Z_2$  for MZ and DZ twins by means of a *Cholesky* decomposition (Neale and Cardon [6]). When  $(Y_1, Y_2)$  and  $(Z_1, Z_2)$  are independent and  $Var(Y_i) = \sigma_\lambda^2$ ,  $Corr(Y_1, Y_2) = \rho_\lambda$ ,  $Var(Z_i) = \sigma_\mu^2$ ,  $Corr(Z_1, Z_2) = \rho_\mu$ ,  $i = 1, 2$  the genetic analysis can be carried out by a more simple approach similar to the analysis of correlations in liability described above.

We assume a *correlated gamma-frailty model* (Yashin and Iachine [8]) for  $(Y_1, Y_2)$  and  $(Z_1, Z_2)$  with respective variances and correlation coefficients. The bivariate Laplace transform for this model is given by:

$$L_{CF}(s_1, s_2; \rho, \sigma) = (1 + \sigma^2(s_1 + s_2))^{-\frac{\rho}{\sigma^2}} (1 + \sigma^2 s_1)^{-\frac{1-\rho}{\sigma^2}} (1 + \sigma^2 s_2)^{-\frac{1-\rho}{\sigma^2}} \quad (8)$$

where  $\sigma^2$  is the variance and  $\rho$  is the correlation coefficient of the bivariate frailty distribution. A *Gompertz* model is assumed for the baseline hazard functions:  $\lambda(t) = h_G(t; a_\lambda, b_\lambda)$  and  $\mu_0(t) = h_G(t; a_\mu, b_\mu)$  where  $h_G(t; a, b) = ae^{bt}$  and  $a_\lambda, b_\lambda, a_\mu, b_\mu \geq 0$  are parameters.

## 2.4 Dependent Hazards Explain Correlation in Liability

The multistate model of debilitation and mortality allows for two potential sources of genetic influence on the health status: the genetic influence on the debilitation process (transitions from *Healthy* to *Disabled*) and the genetic influence on the mortality process (transitions from *Healthy* and *Disabled* to *Dead*). Consequently, using this model for data analysis results in two heritability estimates. One is associated with the hazard to become disabled, another deals with the hazard of death conditional on being in a particular disability state. The traditional liability-based approach provides only one heritability estimate, i.e. the estimate of heritability in liability for the health status variable. It is therefore important to understand how this estimate is related to the two estimates of heritability in frailty obtained from the multistate model.

For this purpose we have calculated the correlations in liability that would be estimated by a traditional approach applied to the health status data produced



by the multistate model. Graphs of these correlations in liability as functions of age for MZ and DZ twins are presented in Figure 2 along with the correlation coefficients of the respective frailty distributions for MZ and DZ twins that were used to produce the graphs.

Two multistate models were used in the analysis: one model (Figure 2 (above)) illustrates the situation where genetic influence is only present in the frailty variable associated with the debilitation process (the respective correlations of frailty ( $\rho_\lambda$ ) for MZ and DZ twins are 1.0 and 0.5) and there is no genetic variation in the frailty variable related to the mortality process (the correlations in frailty are zero for both MZ and DZ twins). Figure 2 (below) depicts the opposite situation: there is no genetic influence on the debilitation process (the respective frailties for MZ and DZ twins are uncorrelated), but there are considerable genetic effects on the mortality process (the respective correlations in frailties ( $\rho_\mu$ ) are 1.0 and 0.5 for MZ and DZ twins). Other parameters of the model were chosen as follows:  $\sigma_\lambda = \sigma_\mu = 1$ ,  $a_\lambda = 10^{-5}$ ,  $a_\mu = 3 \times 10^{-5}$ ,  $b_\lambda = b_\mu = 0.1$  and  $r = 20$ .

One can see that these two radically different models with respect to the nature and source of the genetic influence on the health status trait produce virtually identical age-trajectories of correlations in liability. In both cases the heritability estimates in liability will be increasing functions of age since the difference between  $\rho_{MZ}$  and  $\rho_{DZ}$  increases with age. The increase of heritability estimates with age may be mistakenly interpreted as an increase of genetic influence on the age at onset of disability with age, even in the absence of any genetic influence on transition from "healthy" to "disabled" state. We can therefore conclude that the traditional liability-based approach to genetic analysis of health status data has a severe limitation: it cannot identify the source of correlation in liability and hence, it cannot provide a reliable evaluation of the roles of genes and environment in disability and other **chronic** conditions.

This result is not surprising: the traditional liability method excludes important health state information on twin pairs where one of the twins is deceased (the so-called "broken" pairs) from the analysis. The multistate model presented in this paper explicitly models the process of mortality and therefore allows "broken" pairs to be analyzed. Moreover, information about twin pairs where both individuals are deceased at the time of the study can also be used in the analysis. This information can be obtained from the Danish Twin Register (Kyvik et al. [5]). Thus, the multistate model allows us to integrate health status data obtained during cross-sectional studies with survival data from the Danish Twin Register.

## 2.5 Simulation Study

To verify the identifiability of model parameters in the multistate model from twin health status data, a simulation study was performed. For this purpose health status data were simulated using the multistate health-history model for related individuals. For each age between 75 and 100 two samples of data were

Table 1: Simulation Results

N	$\sigma_Y$	$\rho_Y$	$\sigma_Z$	$\rho_Z$	$a_\lambda \times 10^5$	$b_\lambda$	$a_\mu \times 10^5$	$b_\mu$	$r$
True	1.00	0.50	1.00	0.50	1.00	0.10	1.00	0.10	2.00
100	1.08 (0.48)	0.55 (0.31)	0.99 (0.24)	0.56 (0.19)	1.30 (1.37)	0.11 (0.03)	2.62 (4.91)	0.10 (0.02)	3.88 (7.21)
1000	1.00 (0.14)	0.53 (0.13)	1.00 (0.09)	0.51 (0.06)	1.03 (0.44)	0.10 (0.01)	1.02 (0.40)	0.10 (0.01)	2.00 (0.03)

generated by single year age groups. The size of the first sample is 100 twin pairs for each age. The second sample contains information about 1000 twin pairs for each age. The simulation procedure was repeated to obtain a series of 100 simulations each consisting of the two samples described above.

Maximum likelihood estimation procedure was used to obtain the estimates of the parameters for each sample. The expression for the likelihood function was derived using (7) and (8). The integrals resulting from (6) were evaluated numerically in each step of the maximization procedure. The results are presented in Table 1. The first row displays the true parameter values used in the simulation. The following rows show the empirical average and standard deviation (in parentheses) of the respective parameter estimates calculated from a series of 100 simulations for the two samples ( $N = 100$  and  $N = 1000$ ). The results indicate that estimates of the parameters are close to the true parameter values. However, the variances of the parameter estimates in the case of 100 observations per age are high, especially for the relative risk parameter  $r$ .

### 3 Discussion

Methods of genetic analysis of discrete traits based on liability models are a standard tool of genetic analysis of health status data. An appealing feature of these methods is their simplicity. Available statistical packages permit the estimation of the respective parameters of liability using current status data on twins.

The simplicity of these methods is also their weakness. Liability models are static — they describe the health state distribution at one fixed point of time. No mechanism is present in the model to provide a stochastic description for the evolution of the health state of a single individual. As a result, no conclusions about the future health trajectory of one twin individual can be made when the health history of his/her co-twin is known and the parameters of the liability model are estimated. For the same reason traditional liability methods cannot handle twin pairs with one deceased individual, since such pairs occur when a transition from "disabled" to "dead" state occurs and such transitions cannot be described by a traditional liability model. That is why "broken" twin pairs are



excluded from the analysis. For each discrete trait (i.e. health state) one liability variable is used. As a result only one heritability estimate for the liability variable can be obtained, even if our background knowledge about the aging process suggests two possible sources of genetic influence: the debilitation and mortality processes. No conclusions about the source of genetic influence can be made from one heritability estimate — the liability model for health status data tells us virtually nothing about the nature of the aging process.

In other words, traditional liability models are too simplified to provide a reliable tool for studying the processes that occur in an aging individual — the processes of transitions between the health states. To learn more about aging one must use a model that characterizes this process and provides means to estimate model parameters from health status data. The multistate model developed in this paper fulfills these criteria. It provides a convenient framework of a health trajectory by explicitly modelling the transitions between the health states. This framework allows us to use all available information on twins (including those who are deceased) and identify the source of genetic influence on the aging process.

The multistate model provides a way to extend the analysis of health status data using the information from the second wave of the Longitudinal Study of Aging in Danish Twins. This data will provide multiple observations of a health history for each individual and in this way allow us to obtain more precise estimates of the parameters. Liability models provide no method for dealing with repeated measurements, since they do not describe health changes over time.

The multistate health history models are more complicated than the traditional liability models. Consequently, more sophisticated estimation procedures have to be used to yield the estimates of the parameters. Additional difficulty is related to the estimation of the unknown underlying hazards  $\lambda(t)$  and  $\mu_0(t)$ . Semi-parametric estimation procedures need to be developed to provide estimates of these functions without the need to use an explicit parametric form for the two hazard functions. Such methods require collection of larger data samples than the traditional liability methods. Additional simulation studies are required to perform a comparison of the statistical power of the two methods.

A technical assumption that we have used in this model is the conditional Markov property. This property implies that given the frailty variables the risk of making a transition from one state to another depends on the individual's age and current state (i.e. it does not depend on the previous transition history). Other models may exploit the conditional semi-Markov property, where the risk of transition may in addition depend on the time spent in the current state. Whether one or another property holds for the health history data is an important question. Its solution can contribute to better understanding of the nature of the aging process.

## 4 Conclusion

This paper shows that genetic influence revealed in disability data by traditional liability analysis may have nothing to do with the heritability of the age of onset of disability. It can be induced by dependence between mortality rates conditional on disability status. A more sophisticated model of dependent health histories suggested in this paper may be used to identify the source of correlation in liability.

## Acknowledgement

The authors thank prof. Bent Jørgensen for valuable comments.

## References

- [1] Aalen, O.O. (1987): Mixing Distributions on a Markov Chain. *Scand. J. Statist.* 14: 281-289.
- [2] Christensen K., Holm N., McGue M., Corder L. and J. Vaupel (1997). A Danish Population-based Twin Study on General Health in the Elderly. Accepted by *J. Aging Health*.
- [3] Iachine I.A., Yashin A.I., Christensen K., and J.W. Vaupel (1997). Combining Incomplete Data in Disability Studies: An Approach to Genetic Analysis of Health Status Data. Presented at the Population Association of America 1997 Annual Meeting, March 27-29, Washington D.C.
- [4] Keiding, N. (1991). Age-Specific Incidence and Prevalence: A Statistical Perspective (with discussion). *J. Roy. Statist. Soc. A* 154: 371-412.
- [5] Kyvik, K.O., Christensen K., Skytthe A., Harvald B. and N.V. Holm (1996). The Danish Twin Register. *Dan. Med. Bull.* 1996 Dec; 43(5): 467-470.
- [6] Neale, M.C. and L.R. Cardon (1992). Methodology for Genetic Studies of Twins and Families. *NATO ASI series D: Behavioural and Social Sciences*. Kluwer Academic Publishers.
- [7] Vaupel J.W., Manton K.G. and E. Stallard (1979). The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography* 16, 439-454.
- [8] Yashin, A.I. and I.A. Iachine (1995). Genetic Analysis of Durations: Correlated Frailty Model Applied to Survival of Danish Twins. *Genetic Epidemiology*, 12: 529-538.
- [9] Yashin A.I., Iachine I.A., Christensen K., Holm N.V. and J.W. Vaupel (1998). The Genetic Component of Discrete Disability Traits: An Analysis Using Liability Models with Age-Dependent Thresholds. *Behavior Genetics*, 28(3):207-14.

Figure 2.

