The Timing of First Sexual Intercourse, Heritability, and Social/demographic/Contextual Influences

Our proposed project has two objectives. First, using the twin adolescent samples collected by the National Longitudinal Study of Adolescent Health, we will examine whether or not genetic factors play an important role in influencing the timing of the first sexual intercourse among adolescents in the United States. Second, if we do find important genetic factors, we will proceed to examine how much the social, demographic, and contextual influences moderate the influences of genetic factors on the outcome.

There is a rapidly growing body of evidence pointing to an important role of genetics in the determination of human pathology, psychopathology, and physical traits. The genetic basis for such neurological disorders as Parkinson's disease, epilepsy, and Huntington's disease is well-established. Consistent with our direct observation of the resemblance in physical traits among biologically related relatives, genetic studies have found that more than 70% of variance in height and weight is attributable to genetic influences.

Compared with pathology and physical traits, complex human behavior appears much more environmentally determined; nevertheless, behavior geneticists have reported genetic effects on such seemingly environmentally-determined behavior as parenting style, rate of accident occurrence in childhood, television viewing habits, peer group selection, social support, marital disruption, educational attainment, and socioeconomic status.

Previous work has found the effects of a number of social economic factors on first sexual intercourse such as gender, ethnicity, family structure, and neighborhood characteristics. These social and contextual influences will be examined carefully for their potential moderating influences on the expression of genetic factors.

The analytical approach we will use is the random effects proportional hazards model with the gamma frailty. The random effects modeled by the gamma frailty will from the MZ twins and the DZ twins will reveal how important the genetic factors are. The gamma frailty will facilitate the interpretation of the results.

Methods

The statistical methods for studying heritability when the outcome variable is linear is well-developed. For the survival data such as the timing of first sexual intercourse, we propose a random-effect bivariate hazard model for survival data, in which the random effect W 1 is assumed to have a gamma distribution.[i] Let $T_{i1},\ldots,T_{in_i}$ 2 denote the n_i 3 lifetimes in cluster *i*. For twin observations, $n_i = 2$ 4 always. We assume that conditional on a common random effect W_i 5, the T_{ij} 6 are mutually independent, and satisfy the proportional hazards model

$$\lambda_{ij}(t|W_i = w_i) = w_i \lambda_{ij}(t;\theta) 1, \qquad (1))$$

where $\lambda_{ij}(t;\theta)$ 7 is a baseline hazard function depending on a vector-valued parameter θ 8. Note that the effect of frailty is multiplicative. The unmixed hazard may in turn follow a proportional hazards model with covariates 9, so that $\lambda_{ij}(t;\theta) = \lambda_o(t;\alpha)\exp\{$ 2 where $\lambda_0(t;\alpha)$ 10 is the new baseline hazard and α 11 and β 12 are vector-valued parameters. In addition, we assume that the random frailty effect W_i 13 has a gamma distribution with density $f(w_i) = w_i^{\phi-1} e^{-\phi w_i} \phi^{\phi} / \Gamma(\phi),$ 3 so the mean is unity and the variance is ϕ^{-1} 14.

We observe exposure times t_{ij} 15 and death indicators δ_{ij} 16, which take the value one if the observation is a death and zero otherwise. If frailty W_i 17 were observable, we could estimate the parameters ϕ 18 and θ 19 maximizing the full likelihood

$$L = \prod_{i=1}^{m} f(w_i) \prod_{j=1}^{n_i} \lambda_{ij}(t_{ij})^{\delta_{ij}} w_i^{\delta_{ij}} e^{-\Lambda_{ij}(t_{ij}) w_i}, 4 \qquad (2))$$

where *m* is the total number of clusters. Since frailty is not observable, however, we treat the w_i 20 as missing data and maximize the likelihood using the EM algorithm (Guo and Rodríguez 1992).

The heritability can be estimated using Clayton's characterization of the random-effect bivariate hazard model as follows

$$\frac{\lambda(t_1 \mid T_2 = t_2)}{\lambda(t_1 \mid T_2 > t_2)} = 1 + \phi^{-1} 5, \qquad (3))$$

where $T_1$21 and $T_2$22 are the survival times of the twins and $\phi^{-1} \geq 0$23. In words, the ratio of the hazard function of $T_1$24 at any duration $t_1$25 given $T_2 = t_2$26, to the hypothetical hazard of $T_1$27 given $T_2 > t_2$28, is a positive constant equal to one plus the gamma-variance of twin-pair-specific frailty. For example, an estimated ϕ^{-1}29 of 0.20 means that if one twin has experienced the event , the risk of experiencing the event for the index twin would be 20 per cent higher than that for the same index twin if the other twin had not experiencing the event. The size of ϕ^{-1}30 to be estimated from the twins will be an indicator of the level of heritability. Similar approaches can be applied to a number of environmentally-defined groups and estimating heritability for these groups separately would amount of estimates of the interactions between environments and genetic factors.

---

i. D.Clayton, 'A model for association in bivariate life tables and its application in epidemiological studies of familial tendencies in chronic disease incidence'. Biometrika **65**, (1978) pp.141-151. We distinguish 'random-effect bivariate model' and 'random effect multivariate model' from 'bivariate relationship model' and 'multivariate relationship model'. The first two terms describe the class of models that handle correlated survival data. In this paper, a random-effect bivariate model is used for twin survival data, which are pairwise correlated. Later in this paper, we use 'bivariate relationship' to refer to the relationship between child survival and one independent variable, and 'multivariate relationship' to refer to the relationship between child survival and more than one independent variables.