Longevity Studies in GenomEUtwin

Axel Skytthe[1], Nancy L. Pedersen[3], Jaakko Kaprio[4], Maria Antonietta Stazi[5], Jacob v.B. Hjelmborg[1], Ivan Iachine[2], James W. Vaupel[1], and Kaare Christensen[1]

[1] *The Danish Twin Registry, Epidemiology Unit, Institute of Public Health, University of Southern Denmark, Odense, Denmark*
[2] *The Department of Statistics, Institute of Public Health, University of Southern Denmark, Odense, Denmark*
[3] *The Swedish Twin Registry, Karolinska Institutet, Stockholm, Sweden*
[4] *The Finnish Twin Cohort Study, University of Helsinki, Finland*
[5] *The Italian Twin Registry, Instituto Superiore di Sanità, Roma, Italia*

Previous twin studies have indicated that approximately 25% of the variation in life span can be attributed to genetic factors and recent studies have also suggested a moderate clustering of extreme longevity within families. Here we discuss various definitions of extreme longevity and some analytical approaches with special attention to the challenges due to censored data. Lexis diagrams are provided for the Danish, Dutch, Finnish, Italian, Norwegian, and Swedish Twin registries hereby outlining possibilities for longevity studies within GenomEUtwin. We extend previous analyses of lifespan for the Danish 1870–1900 twin cohorts to include the new 1901–1910 cohorts, which are consistent with the previous findings. The size of the twin cohorts in GenomEUtwin and the existence of population-based, nationwide health and death registers make epidemiological studies of longevity very powerful. The combined GenomEUtwin sample will also allow detailed age-specific heritability analyses of lifespan. Finally, it will provide a resource for identifying unusual sibships (i.e., dizygotic twin pairs) where both survived to extreme ages, as a basis for discovering genetic variants of importance for extreme survival.

During the last decade a series of twin studies has shown that approximately 25% of the variation in lifespan is caused by genetic differences. This seems to be a rather consistent finding in various Nordic countries across different time periods among other species not living in the wild (Finch & Tanzi, 1997; Herskind et al., 1996; Iachine et al., 1998; Ljungquist et al., 1998).

For extreme longevity, moderate familial clustering has also been observed: Perls and co-workers found that the survival ratio for siblings of centenarians versus siblings of 73-year-olds was about four-fold for ages 80–94 (Perls et al., 1998). Kerber et al. (2001) also found, based on Mormon genealogies, an increased recurrence risk for siblings for surviving to extreme ages, although the estimate was somewhat lower than for the centenarians. Gudmundsson and colleagues (2000) using the population-based genealogy in Iceland, found that the first-degree relatives of the probands who live to an extreme old age (≥ 95 percentile) are twice as likely as the controls to survive to the same age. Our preliminary Danish work suggests that siblings born after centenarians have a four-fold increased chance of becoming centenarians compared to their birth-cohort (Skytthe et al., 2002).

A large ongoing research effort is underway to identify the genetic, environmental, and behavioral determinants of extreme survival by comparing centenarians with younger cohorts (association studies). However, such studies suffer from the lack of an appropriate comparison group as cohort specific characteristics may confound the comparison between the centenarians and younger cohorts. To date, only one common polymorphism, namely Apo-E e2 [ARG158CYS] of the ApoE e2/e3/e4 polymorphism, has consistently been shown to be associated with survival until extreme ages (Gerdes et al., 2000).

There are few genetic linkage studies of extreme and healthy survival. Perls and co-workers performed a genome-wide scan for detection of regions of significant linkage potentially harbouring one or more QTLs associated with extreme survival using 308 individuals belonging to 137 sibships demonstrating exceptional longevity. By using nonparametric analysis, borderline significant evidence for linkage was noted for chromosome 4 at D4S1564 (Puca et al., 2001).

As pointed out by Putter et al. in this issue the affected-sib pair method provides a powerful method for identifying genetic variants but at present there are no affected sib-pair studies of extreme survival apart from the study by Puca et al. (2001). In this paper we outline the possibilities for GenomEUtwin to identify rare sibships (i.e., dizygotic twin pairs) where both have survived to extreme ages, as a basis for identifying genetic variants of importance for extreme survival. Finally, this paper will expand previous analyses of lifespan in the Danish Twin Registry to include the 1901–1910 cohorts.

How to Define Longevity

Longevity can be defined in numerous ways. If the interest is normal variation in lifespan then age at death can be used as the phenotype (often with some lower limit to exclude child mortality and accidents from studies of adult mortality or "natural causes" of death).

Address for correspondence: Kaare Christensen, Epidemiology Unit, Institute of Public Health, University of Southern Denmark, Sdr. Boulevard 23A, DK-5000 Odense C, Denmark.
Email:kchristensen@health.sdu.dk

Often the interest will, however, be in extreme longevity (EL). The most appropriate definition of EL depends heavily on the research question. Below the strengths and weaknesses of different definitions are highlighted.

1. Age: The most obvious and simple definition for EL is those individuals who have survived past some certain age. However, this simple definition does not take into account factors affecting survival such as gender, population shifts, and secular trends, for example. The "exceptionality" of being a centenarian depends heavily on sex and birth cohort. For example, while the proportion of males born in 1900 who became a centenarian is 0.26% in the US, the expected proportion of females born in 1920 who will become centenarians is 10 times bigger (2.65%). Another practical consideration is the availability of a sufficient number of potential subjects to produce adequately powered studies; for example, surely a 99-year-old individual exhibits EL despite the fact that she/he may not have reached the 100 year cutoff for a particular study.
2. Sex- and cohort-specific top x%: That is, the x% most long-lived individuals of one sex within a certain cohort (e.g., the top 1% of males born in the United States [US] in 1900) or sub-cohort (e.g., the top 1% of Afro-American males born in the US in 1900). Selecting the age cutoff this way eliminates the effect of secular trends. Based on the US cohort life tables, the probability of surviving to age 95 for men is similar to the survival probability to age 100 for women for all cohorts. These data also indicate that survival to age 90 for women may not be "all that exceptional" at least for cohorts born in 1910 and 1920. Further, the secular decline in mortality during the 20th century has influenced cohort survival. By choosing appropriate sex and cohort specific percentile cutoffs, one can produce a definition of EL that reflects the actuarial data and controls for some important confounding factors. The actual value of "x" used depends again on the research question.
3. Deviation in lifespan: Which can be seen as a generalization of number 2. In this approach not only sex and cohort information is included, but also information on environmental and behavioral factors. The phenotype of interest will then be the difference between the observed lifespan and the expected lifespan (based on sex, birth cohort, environmental and behavioral factors). However, 1,2 and 3 focus on time of death and do not explore the potentials of survival analyses approaches.

   An analysis based on selection of extreme survivors followed by an application of an affected sib pair-method allows for the search for genes influencing survival until the extreme age. However, in this case it is not possible to point out whether the effects of the genes are present during the entire human life or whether they are specific to a particular age group. Therefore, it is of interest to consider age-specific hazard rates.
4. Age-specific hazard rates: The hazard rate approach focuses on age-specific genetic effects (NIA Aging and Genetic Working Group, 2000). The main idea is that the genes do not directly determine the human lifespan, but instead act by increasing or decreasing the age-specific force of mortality or hazard rate. In this case the variation of the risk of dying in a certain age group is the object of the study. In particular, models based on the concept of individual frailty have been used to analyse the genetic influence on age-specific susceptibility to death using Danish, Swedish and Finnish twin data (Iachine et al., 1998). These models were subsequently extended to allow for linkage analysis of survival data (Iachine et al., 1999; Li, 1999) and may be adapted to study age-dependent genetic effects.

   Another advantage of this methodology is the ability to utilize additional mortality follow-up data after DNA-sample collection. At advanced ages the number of non-survivors after a 1-year follow-up might be considerable. These additional censored lifespan observations may be included in the analysis to increase the power of a genetic study (Hsu et al., 2002; Li & Hsu, 2000).

Twin Cohorts Available in GenomEUtwin for Longevity Studies

Whether a twin cohort is suitable for longevity studies depends on a number of factors. First of all, a considerable proportion of the cohort individuals must have reached an age which may be considered as well above the normal lifespan. This means that the youngest individuals to be included in a study should have reached a minimum age around the life expectancy of newborns. For most European countries this corresponds to about 75 years. Secondly, the ability to maintain an updated cohort with respect to vital status enhances the usefulness of the cohort.

A number of the twin cohorts represented in the GenomEUtwin project fulfil these criteria and are suitable for several types of longevity studies. The Nordic twin registries are extremely useful due to the ease of follow-up of the cohort members by linking to the national population registers.

The structure of the different twin cohorts can be illustrated by Lexis diagrams. In a Lexis diagram the relation between age, time and cohort is visualized by having the Y-axis representing age, the X-axis, calendar time, and the diagonal in the up-and-right direction, the life course of the individuals. By drawing a horizontal line at a given age one can easily identify cohorts, which have members above this line.

In Figure 1 Lexis diagrams for twin cohorts from six participating countries in the GenomEUtwin project — Denmark, Finland, Italy, the Netherlands, Norway, and Sweden — are shown. It follows from the above that at present only the Nordic countries include birth cohorts old enough for studies of longevity, although the ongoing establishment of the Italian twin cohorts will add a considerable number of elderly twins and thereby enlarge the basis for longevity studies in the future (Stazi et al., 2002).

In Table 1 a list of available twin cohorts for longevity studies is given. The twin cohorts from Norway and the Netherlands are too young to be suitable for longevity

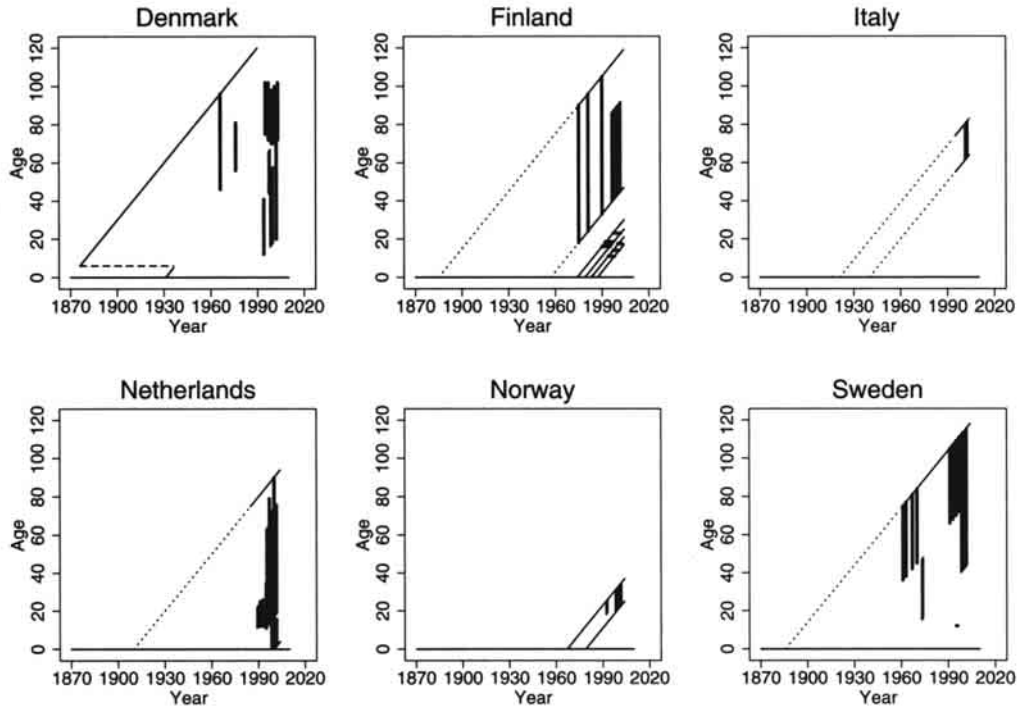

Figure 1a

Lexis diagrams for twin cohorts from six participating countries. The diagonal lines enclose cohorts included in the GenomEUtwin project. Shaded areas indicate when surveys have been conducted. Dotted lines delimit left-truncated cohorts (see text).

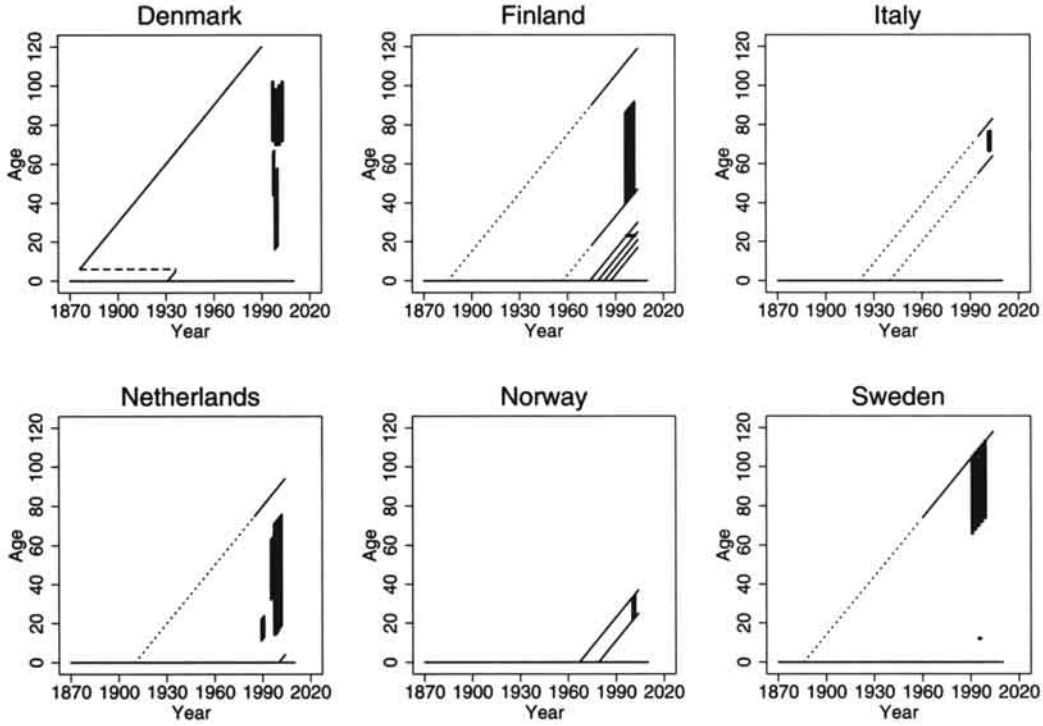

Figure 1b

Lexis diagrams for twin cohorts from six participating countries. The diagonal lines enclose cohorts included in the GenomEUtwin project. Shaded areas indicate surveys with sampling of biological material. Dotted lines delimit left-truncated cohorts (see text).

studies, since the oldest Norwegian twin cohorts are born in 1967, and the Dutch Twin Registry has primarily focussed on younger twins and comprises less than 100 twins aged 70 years or more (Boomsma et al., 2002).

In Table 2 the numbers of intact twin pairs are given for different cut-off ages, illustrating the potential for identifying intact twin pairs reaching an exceptional age. Ninety years may not seem that exceptional, but having both co-twins surviving to that age only happens to about 1% of all twin pairs. Having both co-twins surviving to 95 years is certainly exceptional, and the numbers in Table 2 indicate the need of combining twin cohorts from several countries in order to have sufficient power in studies using the affected sib-pair method (DZ pairs). The declining ratio of DZ:MZ intact pairs for increasing cut-off age is well in line with a higher correlation among MZ than DZ pairs for lifespan.

Combining cohorts from several countries can be advantageous in many respects, but also introduces analytical challenges. It may be necessary to take into account the left-truncation by date present in several cohorts (i.e., they only include pairs for which both members are alive after a particular date). This implies that both members have to survive to a certain age, the age at entry, in order to be included in the sample. The age at entry is determined by the date of birth and date of truncation. The analysis of the left-truncated samples involves bivariate survival analysis methods, which can handle delayed entry and right-censoring (Iachine et al., 1998). Both the Finnish and the Swedish cohorts exhibit left-truncation, and the analysis of each cohort as well as the combined sample requires advanced analytical methods. Application of these methods fall outside the scope of this paper, and only in the Danish cohort is the use of traditional methods for estimation of heritability of lifespan possible.

Lifespan of Danish Twins Born 1870–1910

Based on a follow-up until May 1, 1994 of Danish twins born 1870–1900, Herskind et al. (1996) found that the heritability of lifespan was 0.26 for males and 0.23 for females with no cohort effect. The analysis was based on same-sexed twin pairs in which both pairs survived to age 15 and in order to normalize the data, the correlation and heritability analyses used "lifespan squared". Some 0.6% of the 1870–1900 sample was still alive in 1994. We extended the follow-up of the 1870–1900 cohort up to January 8, 2003 and added the 1901–1910 cohort to the analyses. Table 3 gives the basic descriptives of the sample. From Table 4 it is seen that the 1901–1910 cohort had correlations similar to the previous cohorts. The correlations given here are the raw lifespan correlations and therefore not directly comparable to the "squared lifespan" correlation provided by Herskind et al. (1996).

We restricted these basic analyses to twins pairs where both survived age 15 and both died in Denmark, since the analysis of the complete sample of the Danish twin survival data is complicated by the right censoring problem — about 7% of all twin pairs born 1870–1910 contain one or more twin individuals with censored lifespan (see Table 3). The primary reasons for censoring are that the individuals

Table 1
Twin Cohorts Available for Longevity Studies

Country	Birth cohorts	Ascertainment method and period	Approx. number of twin pairs with known zygosity	Biological material available from approx. number of twin individuals	Notes
Denmark	1870–1930	Search of all parish registers 1953–1965	15300	3000	Both co-twins alive at age 6, 1911–1930 SS only
Finland	1885–1957	Selected from Central Population Register 1975	14000	6000	Both co-twins alive in 1975 SS only
Italy	1921–1940	Selected from regional population register 1995	500	150	Both co-twins alive in 1995
Sweden	1886–1925	Search of all parish registers 1959–1961	11000	1000	Both co-twins alive in 1959 SS only

Note: Only the oldest cohorts from each country are listed. Twin cohorts from Norway and the Netherlands are too young for longevity studies. Unless specified, both same-sexed (SS) and opposite-sexed (OS) twin pairs are included.

Table 2

Number of Twin Pairs Where Both Co-twins Are Alive at Age X

	Denmark[a]		Finland[b]		Sweden[c]	
X	MZ	DZ	MZ	DZ	MZ	DZ
85	267	365	104	149	546	754
90	66	65	25	27	128	145
95	9	8	2	3	15	15

Note: a. at January 8, 2003
b. at December 31, 2001
c. at March 16, 2003

Table 3

Status January 8, 2003 for Danish Twin Pairs Born 1870–1910 in Which Both Twins Survived to Age 15

Twin1 \ Twin 2	Alive	Emigrated	Dead	Total
Alive	10	0	57	67
Emigrated	0	70	89	159
Dead	57	97	4667	4821
Total	67	167	4813	5047

Table 4

Lifespan and Intra-pair Similarity for Danish Twin Pairs Born 1870–1910 — Completely Uncensored Pairs

	Males		Females	
	MZ	DZ	MZ	DZ
1870–1880				
N (pairs)	113	186	126	215
Mean age at death (*SD*)	71.5 (15.9)	70.1 (17.1)	73.1 (16.3)	70.7 (17.8)
Variance	251.6	291.2	266.7	318.1
Covariance	47.2	17.5	64.9	–7.4
*r**	0.19	0.06	0.25	0.00
1881–1890				
N (pairs)	168	306	184	330
Mean age at death (*SD*)	70.2 (16.1)	68.4 (18.1)	72.9 (17.8)	70.8 (19.1)
Variance	257.9	327.6	315.0	366.0
Covariance	51.7	37.1	70.2	43.4
r	0.20	0.12	0.23	0.12
1891–1900				
N (pairs)	236	410	217	417
Mean age at death (*SD*)	71.0 (16.9)	69.7 (17.9)	74.5 (17.9)	72.9 (18.4)
Variance	286.5	319.2	321.7	338.5
Covariance	25.9	29.2	70.2	20.7
r	0.09	0.09	0.19	0.06
1901–1910				
N (pairs)	319	561	298	581
Mean age at death (*SD*)	72.2 (15.0)	70.4 (16.1)	75.2 (16.7)	73.9 (17.7)
Variance	223.6	260.4	277.5	312.6
Covariance	42.2	27.2	44.4	21.1
r	0.19	0.10	0.16	0.07
All cohorts				
N (pairs)	836	1463	825	1543
Mean age at death (*SD*)	71.4 (15.9)	69.7 (17.1)	74.2 (17.2)	72.5 (18.3)
r	0.16	0.10	0.20	0.07

Note: * *r* = intraclass correlation coefficient

Table 5

Estimates of Correlation Coefficients for Danish Twin Pairs Born 1870–1910 — Including Censored Pairs, Adjusted for Birth Cohort

	r_{MZ}	r_{DZ}
Males	0.195 (0.124–0.264)	0.086 (0.038–0.134)
Females	0.233 (0.166–0.297)	0.075 (0.028–0.122)

are still alive (at January 8, 2003) or have emigrated from Denmark. In this case the register contains the age of the last observation of the living twin (i.e., the age on January 8, 2003, or the age at emigration).

An approach to the censoring problem, which is often applied when the number of censored observation is not too large, is to restrict the analysis to twin pairs where both individuals are deceased, as it was done for the descriptive analysis of the Danish twin survival data presented in this paper. However, this approach may lead to overestimated mortality rate and biased estimates of the correlation coefficients. Moreover, as younger twin cohorts contain more living individuals than the older ones, the censoring problem will become increasingly important if analysis of survival data from the younger twin cohorts has to be considered.

One approach to the analysis of censored bivariate survival data is based on parametric survival models (Hougaard, 2000). In this case the parameter estimation may be implemented using the Maximum Likelihood Estimation (MLE) procedure. If the goal of the analysis is the estimation of the correlation coefficient between the lifespan of the twins, it is convenient to apply a parametric survival model, which includes the correlation coefficient as one of the parameters (e.g., the bivariate normal distribution). In this case the likelihood function is constructed using the bivariate normal probability density function for the uncensored observations, the bivariate normal survival function, $S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2)$, for the completely censored observations and $P(T_1 > t_1, T_2 = t_2) = -\partial S/\partial t_2$ for pairs with one censored and one deceased twin individual. Using this approach we estimated lifespan correlations using the complete dataset, which included all censored observations (with adjustment for birth cohort and assuming equal means and variances for MZ and DZ twins). The results are given in Table 5. There is only a small difference between the correlations estimated from the complete sample and those estimated from the completely uncensored pairs due to a relatively small degree of censoring. However, the normal distribution does not provide a satisfactory description of the human lifespan distribution, which is skewed. A better approach may be based on fitting bivariate frailty models to the data (Iachine et al., 1998) followed by the calculation of the predicted correlation coefficient. These models allow for more accurate modelling of the marginal lifespan distributions (e.g., Gompertz or gamma-Makeham mortality trajectories). We will explore these approaches in future work on the heritability of lifespan.

Perspectives

The combined twin population in GenomEUtwin will provide the basis for a wide range of longevity studies. Due to the long history of many of the participating twin registries, traditional epidemiological studies will be very powerful, especially when combined with existing population-based health registries found in many of the participating countries. The ability to combine the oldest cohorts in the Swedish, the Finnish, and the Danish registries will provide an opportunity to improve heritability analyses of life span. The combined samples will allow for various left-truncations, that is to say heritability conditioned by survival to, for example, age 50, age 70, age 80.

Ascertaining and drawing blood from the oldest dizygotic twin pairs in the Finnish, the Swedish, and the Danish registries will provide a valuable start for collecting a sample of extremely long-lived siblings which can be used in affected sib-pair analyses of extreme longevity. As seen in the description of the twin cohorts, a substantial proportion of these pairs have already provided blood for the registries, and high priority will be given to collect blood samples from the oldest pairs who still not have provided blood. The establishment of the Italian twin registry will add substantially to the longevity studies in GenomEUtwin as the population in Italy is 5–10 times larger than most of the Nordic countries. The Italian population aged 85+ comprises more than one million and the number of intact twin pairs in that age group is expected to be around 2–3000.

There is a growing interest in ascertaining extremely long-lived sib-pairs in order to identify genetic factors associated with longevity. A network of European scientists, several of whom are also members of GenomEUtwin, has submitted a EU proposal to ascertain extremely long-lived "ordinary" siblings in Europe in the project GEHA. Also the National Institute of Aging in the US has launched a program for various centers to collaborate in ascertaining extremely long-lived relatives. The combined efforts of these initiatives will provide an excellent basis for identifying genetic variants of importance for longevity, and GenomEUtwin will be a valuable contribution to this research.

References

Boomsma, D. I., Vink, J. M., van Beijsterveldt, T. C. E. M., de Geus, E. J. C., Beem, A. L., Mulder, E. J. C. M., et al. (2002). Netherlands Twin Register: A focus on longitudinal research. *Twin Research, 5,* 401–406.

Finch, C. E., & Tanzi, R. E. (1997). Genetics of aging. *Science, 278,* 407–411.

Gerdes, L. U., Jeune, B., Andersen-Ranberg, K., Nybo, H., & Vaupel, J. W. (2000). Estimation of apolipoprotein E genotype-specific relative mortality risks from the distribution of genotypes in centenarians and middle-aged men: Apolipoprotein E gene is a "frailty gene", not a "longevity gene". *Genetic Epidemiology, 19,* 202–210.

Gudmundsson, H., Gudbjartsson, D. F., Frigge, M., Gulcher, J. R., & Stefansson, K. (2000). Inheritance of human longevity in Iceland. *European Journal of Human Genetics, 8,* 743–749.

Herskind, A. M., McGue, M., Holm, N. V., Sorensen, T. I., Harvald, B., & Vaupel, J. W. (1996). The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900. *Human Genetics, 97,* 319–323.

Hougaard, P. (2000). Analysis of multivariate survival data. New York: Springer.

Hsu, L., Li, H., & Houwing-Duistermaat, J. J. (2002). A method for incorporating ages at onset in affected sibpair linkage studies. *Human Hereditary, 54,* 1–12.

Iachine, I. A., Holm, N. V., Harris, J. R., Begun, A. Z., Iachina, M. K., Laitinen, M., Kaprio, J., & Yashin, A. I. (1998). How heritable is individual susceptibility to death? The results of an analysis of survival data on Danish, Swedish and Finnish twins. *Twin Research, 1,* 196–205.

Iachine, I. A., Christensen, K., & Yashin, A. I. (1999). Incorporating genetic marker information into the analysis of twin survival data: A simulation study (Research Report 2). Odense: Department of Statistics and Demography, SDU, Odense University.

Kerber, R., O'Brien, E., Smith, K., & Cawthon, R. (2001). Familial excess longevity in Utah genealogies. *Journal of Gerontology: Biological Sciences, 56A,* B130–B139.

Li, H. (1999). The additive genetic gamma frailty model for linkage analysis of age-of-onset variation. *Annals of Human Genetics, 63,* 455–468.

Li, H., & Hsu, L. (2000). Effects of age at onset on the power of the affected sib pair and transmission/disequilibrium tests. *Annals of Human Genetics, 64,* 239–254.

Ljungquist, B., Berg, S., Lanke, J., McClearn, G. E., & Pedersen, N. L. (1998). The effect of genetic factors for longevity: A comparison of identical and fraternal twins in the Swedish Twin Registry. *Journal of Gerontology A Biological Sciences Medical Science, 53,* M441–M446.

NIA Aging and Genetic Epidemiology Working Group (2000). Genetic epidemiologic studies on age–specified traits. *American Journal of Epidemiology, 152,* 1003–1008.

Perls, T. T., Bubrick, E., Wager, C. G., Vijg, J., & Kruglyak, L. (1998). Siblings of centenarians live longer [letter]. *Lancet, 351,* 1560.

Puca, A. A., Daly, M. J., Brewster, S. J., Matise, T. C., Barrett, J., Shea-Drinkwater, M., et al. (2001). A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4. *Proceedings of the National Academy of Science U.S.A., 98,* 10505–10508.

Skytthe, A., Jeune, B., Vaupel, J. W., & Christensen, K. (2002). Exceptional longevity in humans: The role of familial clustering. Poster at the 2nd Nordic Conference in Epidemiology, 9–12 June 2002. In Aarhus: The Danish Epidemiological Society.

Stazi, M. A., Cotichini, R., Patriarca, V., Brescianini, S., Fagnani, C., D'Ippolito, C., et al. (2002). The Italian Twin Project: From the personal identification number to a national twin registry. *Twin Research, 5,* 376–381.