

LOW TOBACCO-RELATED CANCER INCIDENCE IN OFFSPRING OF LONG-LIVED SIBLINGS

Web Appendix 1: The Study population

The identification of long-lived families in Denmark was undertaken systematically in five steps. Initially, all individuals born before April 2, 1918, and alive in 2004 were identified in the Danish Civil Registration System (DCRS), which covers all persons alive and living in Denmark on or after April 2nd, 1968 [1]. Among other variables, this registry includes information on personal identification number, gender, date and place of birth, continuously updated vital status, and the identity of parents and spouses; the latter, though, only systematically from 1968. After matching the individuals identified in step 1 in clusters of individuals with the same birth parish and surname, the second step consisted of selecting male-male pairs, female-female pairs and male-female pairs from each cluster of matched individuals. When a cluster consisted of more than two persons of the relevant sex-combination, the oldest two individuals in the cluster were chosen. The third step was to identify which of these potential pairs of siblings were actual siblings by tracing the names of their parents in church records from each birth parish. This resulted in the identification of 3,636 families with at least two siblings being alive and having reached the age of 86+. As the fourth step, siblings were invited to participate in one of three studies at the Epidemiology unit of the University of Southern Denmark, for which recruitment ran sequentially in the following order: a pilot study (Danish Oldest Siblings = DOS), the Genetics of Healthy Aging (GeHA) study [2] and the Long Life Family (LLFS) study [3]. Priority was given to inviting the oldest pairs and then to brothers. Recruitment to DOS was conditional on both siblings being alive and 88 years or older; recruitment to GeHA required both siblings to be alive and above age 90, and the LLFS recruited only families with a family longevity index (FLoSS) score above 7 [3]. While higher FLoSS scores are closely (positively) correlated with

exceptional longevity in families, they also correlate with the families being large and having many living members at high ages at the time of recruitment. A family could only be in one of the three studies: DOS, GeHA or LLFS. Recruitment in the GeHA study led to subsequent exclusion from recruitment in the LLFS. Finally, as the fifth step, siblings from sib-ships with at least two siblings alive and willing to participate were enrolled in the respective studies until the target number of families was reached. This resulted in a total of 659 families in the study population, of which 114 were from the pilot study DOS, 469 were recruited in the GeHA study and 76 families came from the LLFS study, so that each family belonged uniquely to one of the three studies. For each of the three studies, a structured interview took place in the homes of the participants, where information on their siblings as well as on the offspring of the participants and the participants' siblings was gathered. Interviews were performed by five trained interviewers located in different parts of Denmark, thereby covering the entire country. It cannot be ruled out that unidentified offspring may have worse health profiles, but we expect that few offspring remain unidentified, as this means that their participating parent was unable/unwilling to provide name and birth date.

A total of 1,511 siblings from the 659 families were interviewed, resulting in the identification of a total of 3,972 siblings and 5379 offspring of siblings by providing names and birth dates of these siblings and offspring. To further ensure reliable family information and that the families in this study were strongly enriched for longevity, we restricted our study population to the offspring of those siblings who 1) participated in an interview in either DOS, GEHA or LLFS 2) survived to age 90+ before July 1, 2010, and 3) had at least another participating sibling surviving to age 90+ before July 1, 2010. This means that the population under study consisted of the offspring of those sets of siblings who survived to age 90+ and participated in an interview (in DOS, GEHA or LLFS). So rather than using cohort and sex-specific cut points for classifying siblings as being long-lived, we used an operational criterion of both

siblings having attained age 90 regardless of sex and cohort. Although this means that some families, i.e. those from older birth cohorts and with one or both siblings being male, have been selected for a higher degree of longevity enrichment, all families are enriched for longevity. Names and dates for those alive in 1968 or later were later verified through DCRS, thereby providing the personal identification number and enabling the linkage of each individual to the various registries used in the study. A total of 1405 siblings from 628 families fulfilled this criterion: 99, 454 and 75 families in DOS, GeHA and LLFS, respectively. Of the 1405 siblings, 264 had no offspring, so for the remaining 1141 siblings from 611 families, 3297 offspring were identified. Of these offspring, six had unknown vital status in the DCRS, a further 17 had a status as emigrants at end of study in 2009 but with an unknown date of emigration, and one with emigration status had emigrated before April 2, 1968; four offspring had died at an unknown date, and two had died before April 2, 1968. The remaining 3,267 offspring from 610 families comprise our study population with 448 offspring from 97 DOS families, 2262 offspring from 438 GeHA families, and 557 offspring from 75 LLFS families.

The similarities and differences between families recruited in the different studies are shown in Tables A1-A3. In Table A1, we see that while the sex distribution among the siblings varied from about 50% males in the DOS study to about a third in the GeHA and LLFS studies, the sex distribution of the offspring was close to fifty-fifty for all three studies. This table also shows the tendency of larger sib-ships in families enrolled in the LLFS study, a tendency that is accentuated by the somewhat larger number of offspring in these families. Table A2 shows that the age distributions in the three studies were very similar. Similarity is also found when comparing cancer incidence in these three studies to the background Danish population stratified on sex, five-year age bands, and five-year calendar periods, as shown in Table A3, especially when comparing the overall SIR between the three studies. The sex-specific SIRs vary moderately, but especially the DOS and LLFS studies have small samples

within sex strata resulting in the somewhat wide confidence intervals. Taking sampling variation into account, we find observed cancers to be about 3/4 of the expected number based on cancer rates in the Danish population.

Figures:

Figure A1: Flow chart describing the sampling procedure leading to the recruitment of offspring of nonagenarian siblings in the study.

Table A1. Descriptives of Family Characteristics for Danish Families Recruited in the Three Consecutive Studies:

	DOS-siblings		GeHA-siblings		LLFS-siblings	
N	209		969		227	
Males						
Number and percent	102	48.8%	321	33.1%	82	36.1%
Sibsize						
Median	2		2		3	
Q1 and Q3	2	2	2	2	2	4
Min and max	2	4	2	4	2	6
	DOS-offspring		GeHA-offspring		LLFS-offspring	
N	448		2262		557	
Males						
Number and percent	234	52.2%	1120	49.5%	287	51.5%
Offspring per family						
Median	5		5		6	
Q1 and Q3	3	6	3	7	3	10
Min and max	1	12	1	17	1	25

Abbreviations: DOS, Danish Oldest Siblings; GeHA, Genetics of Healthy Aging; LLFS, Long Life Family Study; Q1, lower quartile; Q3, upper quartile.

Table A2. Birth Year Distribution in Danish Long-lived Siblings, and Birth Year Distribution and Age Distribution of Risktime in Offspring of Danish Long-lived Siblings for the Three Consecutive Studies:

	DOS-siblings			GeHA-siblings			LLFS-siblings		
N	209			969			227		
Families	99			454			75		
Birth year	1911			1912			1913		
Median	1911			1912			1913		
Range	1897	1916		1901	1917		1902	1921	
	DOS-offspring			GeHA-offspring			LLFS-offspring		
N	448			2262			557		
Birth year	1944			1944			1945		
Median	1944			1944			1945		
Range	1921	1969		1925	1970		1926	1969	
	Time at risk		Cancers	Time at risk		Cancers	Time at risk		Cancers
	<i>person-years</i>	<i>percent</i>		<i>person-years</i>	<i>percent</i>		<i>person-years</i>	<i>percent</i>	
Age 0 - 30	3,249.2	17.7%	2	15,833.1	17.4%	3	4,431.4	19.7%	1
Age 30 - 50	8,543.5	46.7%	6	43,217.9	47.6%	52	10,639.1	47.3%	15
Age 50 - 60	3,930.9	21.5%	24	19,840.7	21.8%	92	4,765.1	21.2%	24
Age 60 - 70	2,207.3	12.1%	21	10,655.3	11.7%	120	2,336.0	10.4%	19
Age 70 - 80	369.1	2.0%	9	1,309.1	1.4%	27	325.3	1.4%	8
Age 80+	13.0	0.1%	0	20.5	0.0%	0	9.3	0.0%	0
Total	18,313.1		62	90,876.6		294	22,506.3		67
Affected persons	57			274			66		

Abbreviations: DOS, Danish Oldest Siblings; GeHA, Genetics of Healthy Aging; LLFS, Long Life Family Study.

Table A3: Cancer Incidence in Offspring of Danish Long-lived Siblings, From 1968 to 2009, Stratified on Study of Enrollment.

Study	Gender	Number of families	Number of individuals	Observed number of cancers	Expected number of cancers	SIR (obs/exp)	95% CI
DOS	Males	91	234	25	42.4	0.59	0.39 , 0.89
	Females	86	214	37	36.9	1.00	0.72 , 1.40
	All	97	448	62	79.3	0.78	0.60 , 1.03
GeHA	Males	398	1,120	116	172.4	0.67	0.56 , 0.81
	Females	397	1,142	178	205.4	0.87	0.74 , 1.02
	All	438	2,262	294	377.7	0.78	0.69 , 0.88
LLFS	Males	66	287	34	43.0	0.79	0.56 , 1.12
	Females	70	270	33	45.7	0.72	0.49 , 1.06
	All	75	557	67	88.7	0.76	0.57 , 1.00

Abbreviations: CI, confidence interval; DOS, Danish Oldest Siblings; exp, expected number of cancers; GeHA, Genetics of Healthy Aging; LLFS, Long Life Family Study; obs, observed number of cancers; SIR, Standardized Incidence Ratio.

- [1] Pedersen CB. The Danish Civil Registration System. *Scandinavian journal of public health*. 2011;39(7 Suppl):22-5.
- [2] Skytthe A, Valensin S, Jeune B, Cevenini E, Balard F, Beekman M, et al. Design, recruitment, logistics, and data management of the GEHA (Genetics of Healthy Ageing) project. *Experimental gerontology*. 2011;46(11):934-45.
- [3] Sebastiani P, Hadley EC, Province M, Christensen K, Rossi W, Perls TT, et al. A family longevity selection score: ranking sibships by their longevity, size, and availability for study. *American journal of epidemiology*. 2009;170(12):1555-62.

Recruitment flow chart

