Discussion Papers
Statistics Norway
Research department

No. 776
April 2014

Statistics Norway

Statistisk sentralbyrå

Edwin Leuven, Erik Plug and Marte Rønning

The relative contribution of genetic and environmental factors to cancer risk and cancer mortality in Norway

Discussion Papers No. 776, April 2014
Statistics Norway, Research Department

Edwin Leuven, Erik Plug and Marte Rønning

The relative contribution of genetic and environmental factors to cancer risk and cancer mortality in Norway

Abstract:
Using Norwegian cancer registry data we study twin and non-twin siblings to decompose variation in cancer at most common sites and cancer mortality into a genetic, shared environment and individual (unshared environmental) component. Regardless the source of sibling variation, our findings indicate that genes dominate over shared environment in explaining relatively more of the variation in cancer at most common cancer sites (but lung and skin cancer) and cancer mortality. The vast majority of the variation in cancer and cancer mortality, however, is explained by individual (unshared environmental) factors.

Keywords: Cancer, Twins, Heritability, Environment

JEL classification: I12, J62

Acknowledgements: The research on this paper was funded by the Norwegian Research Council. Some of the data in this article are from the Cancer Registry of Norway. The Cancer Registry of Norway is not responsible for the analysis or interpretation of the data presented.

Address: Edwin Leuven, Department of Economics, University of Oslo.
E-mail: Edwin.leuven@econ.uio.no.

Erik Plug, University of Amsterdam.E-mail: e.j.s.plug@uva.no.

Marte Rønning, Statistics Norway, Research Department. E-mail: mro@ssb.no

Discussion Papers comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no
http://ideas.repec.org/s/ssb/dispap.html

For printed Discussion Papers contact:
Statistics Norway
Telephone: +47 62 88 55 00
E-mail: Salg-abonnement@ssb.no

ISSN 0809-733X
Print: Statistics Norway

Sammendrag

Ved å benytte individdata fra Kreftregisteret koblet til SSB's befolkningsregister og tvillingdata fra Folkehelseinstituttet, utfører vi en såkalt variansdekomponeringsmetode for å skille gener og oppvekts miljø som distinkte årsaker for å utvikle kreft samt dø av kreft.

Våre funn tyder på at gener dominerer over oppvekstmiljø, både når det gjelder de vanligste kreftformene og kreftdødelighet. De enste unntakene er lunge og hudkreft hvor oppvekstmiljø dominerer over gener.

1 Introduction

In this paper we examine how much of the variation in cancer risks and cancer mortality can be explained by genetic and environmental factors. Specifically, we estimate correlations for common cancer types and cancer mortality among twin and non-twin sibling pairs and exploit the variation in which they are genetically connected and exposed to family environments to identify genetic and environmental influences on cancer risks and mortality.

The data we use are a combination of multiple administrative registers in Norway. Cancer information comes from the Norwegian Cancer Registry which holds records of any cancer diagnosis and, in case of death, whether cancer has been the leading cause. Sibling information comes from the Norwegian Population Registry. Twin information comes from the Norwegian Twin Registry. We have matched these registers using personal identification numbers of all Norwegian citizens between the years 1954 and 2007.

We consider two sources of identifying variation: twin sibling and non-twin sibling correlations. When relying on twin siblings, we find that correlations in cancer incidence (at most common sites but lung cancer) and cancer mortality are higher among monozygotic twins (who share all genes) than among dizygotic twins (who share some but not all genes). For non-twin siblings we find that correlations in cancer incidence (at most common sites but lung and skin cancer) and cancer mortality are very similar among both closely spaced siblings (who share more shared environment) and widely spaced siblings (who share less shared environment). While these findings suggest that genes dominate over shared environment (with the exception of lung and skin cancer), the larger part of the variation in cancer and cancer mortality is driven by individual (unshared environmental) factors.

The remainder of the paper proceeds as follows. Section 2 provides the background and motivation behind this study. Section 3 introduces the standard behavior genetic methodology to separate genetic from environmental influences. Section 4 describes

our data set and the Norwegian cancer and population registers upon which our data set is built. The main results are presented in Section 5. And finally Section 6 highlights the implications and conclusions of this study.

2 Background and Motivation

Many people die of cancer. Recent mortality statistics in the EU (including Norway) as well as the US indicate that about one out of every four deaths is caused by cancer (Ferlay et al., 2007). In addition, many people who are diagnosed with cancer have a family history of cancer. For most common cancer forms (including breast cancer, colorectal cancer and prostate cancer), the risk of developing cancer is two to four times higher when a child, sibling or parent is also diagnosed with cancer (Steinberg et al., 1990; Pharoah et al., 1997; Eberl et al., 2005). It is therefore not surprising that scientists, and medical scientists in particular, are interested in the degree to which genetic and environmental family factors determine cancer risks and cancer mortality.

The stakes in the debate on the biological and environmental origins of cancer are high and provide clues on how to reduce cancer risks and increase cancer survival. On one hand, if there are genetic risks of cancer, it would justify wide-scale genetic testing to detect responsible genes (Hopper et al., 2005). On the other hand, if there are environmental risks of cancer, it would provide us with the rationale to look for possible environmental inputs that matter for the development of cancer. A recent literature on candidate inputs, such as smoking, drinking, diet and socioeconomic background, shows that these personal behaviors can have profound effects on both cancer risk and cancer survival (e.g. Eloranta et al. 2010).

There are only a handful of studies that try to separate hereditary and environmental factors as distinctive causes for cancer (Holm et al., 1980; Verkasalo et al., 1999; Lichtenstein et al., 2000). With information from Danish, Swedish and Finnish twin and cancer registries, these studies estimate the heritability of cancer

by comparing correlations of common cancers among monozygotic and dizygotic twin pairs. The basic idea is fairly simple. If cancer correlations are higher among monozygotic twins (who share all genes) than among dizygotic twins (who share some but not all genes), genetic factors are in part responsible for causing cancer. If, on the other hand, cancer correlations are comparable among monozygotic and dizygotic twins, environmental factors are the more likely determinants of cancer. With key assumptions on how much monozygotic and dizygotic twins share genes and environments, twin cancer correlations can then be used to estimate the relative contribution of genetic and environmental cancer risks. The main result in these twin studies is that environmental factors explain most of the variance in the most common cancers (including breast, prostate, ovarian and uterine cancer). Non-shared environmental factors explain about 60 to 80 percent of all the variation in cancer risks while genetic factors as well as shared environmental factors explain only little of the variation in cancer risks.

Not everyone, however, is convinced that the decomposition method using twin data leads to unbiased estimates of the contribution of heritability and environmental effects. In particular, critics have raised three important reservations (Goldberger, 1979; Jencks, 1980; Manski, 2011). First, twins with favorable genes tend to grow up in families with favorable endowments for child development. Second, monozygotic twins tend to share more of the same family environment than dizygotic twins because they are more influenced by each other, and are treated more similar by their parents and others. Third, dizygotic twins share more than half of their genetic material in the presence of assortative mating, genetic dominance, and gene-gene interactions. These reservations are rather unfortunate for a twin decomposition method that relies heavily on genes acting independently of the family environment, twins facing equal family environments, and dizygotic twins sharing half of their genes. We should keep these reservations in mind when interpreting the decomposition results.

In this paper we examine how much of the variation in cancer risks and cancer

mortality can be attributed to genetic and environmental factors in Norway. As a starting point, we begin to replicate previous twin studies on the heritability of cancer using another data set on twins. In view of the sparse literature, it is certainly useful to have more than one study using comparable methodologies with different data sources. But our paper also complements previous work in at least two important directions.

First, we extend the twin sibling sample with non-twin siblings, which allows us to separate heredity and environmental factors under much weaker identifying assumptions. If monozygotic and dizygotic twins share the same shared environment, we can rank the degrees to which monozygotic and dizigotic twin pairs are genetically related to establish whether genetic factors matter for cancer and cancer mortality (and not how much genetic factors matter). If non-twin siblings share, on average, the same share of genes, we can rank the degrees to which closely and widely spaced siblings are exposed a shared environment to establish whether shared environmental factors matter for cancer and cancer mortality (and not how much a shared environment matters). Such comparisons together are informative about the origins of cancer and cancer mortality, with less room for misinterpretation.

Second, we extend our set of cancer outcomes with cancer mortality, for which heritability has never been analyzed. Empirical studies on the genetic and environmental influences on cancer risks are scarce, but less is known about cancer survival (Lindström et al. 2007). Our explicit focus on cancer mortality might shed some light on this.

3 Empirical approach

In order to separate hereditary and environmental factors as distinct causes for cancer and cancer mortality we start out by a linear additive model of genetics and environmental influences for two siblings

$$Y_{1i} = gG_{1i} + sS_{1i} + uU_{1i} \tag{1}$$

and

$$Y_{2i} = gG_{2i} + sS_{2i} + uU_{2i} \tag{2}$$

where subscripts 1, 2 and i denote the first and second sibling in the ith pair. The model is assumed identical for the two siblings. In this model, additive genetic factors G, common environmental factors S and specific environmental factors U account for all the individual differences in the outcome variable of interest Y. All factors, including the outcome variable Y, are standardized and expressed as deviations from zero with a variance of one. The factors G, S and U are unobserved. In this study the parameters of interest are g and s, which measure the influence of G and S on Y.

With the unobservable individual specific components U_{1i} and U_{2i} assumed uncorrelated with each other and with G_{1i}, G_{2i}, S_{1i} and S_{2i}, we can express the outcome correlation (which, due to the normalization is also an outcome covariance) between the observed outcomes of the two siblings as follows

$$\text{Corr}(Y_1, Y_2) = g^2\text{Corr}(G_1, G_2) + 2gs\text{Corr}(G_1, S_2) + s^2\text{Corr}(S_1, S_2) \tag{3}$$

where we impose sibling similarity of the correlation between the siblings' genes and shared environment ($\text{Corr}(G_1, S_2) = \text{Corr}(G_2, S_1)$). Sibling data can then be used to measure the outcome correlation and, together with assumptions on how G and S are related within different sibling pairs, identify the parameters g and s.

As a starting point, we consider the correlation relationship for monozygotic and dizygotic twins and assume (for now) that G and S act independently of each other. If monozygotic (MZ) twins share all of the same genes ($\text{Corr}(G_1, G_2)_{MZ} = 1$) and same family environment ($\text{Corr}(S_1, S_2)_{MZ} = 1$), the correlation between the outcomes of the two twins is given by $\text{Corr}(Y_1, Y_2)_{MZ} = g^2 + s^2$. If dizygotic (DZ) twins

share half of the same genes ($\text{Corr}(G_1, G_2)_{DZ} = 1/2$) and same family environment ($\text{Corr}(S_1, S_2)_{DZ} = 1$), the correlation between the outcomes of the two twins is given by $\text{Corr}(Y_1, Y_2)_{DZ} = g^2/2 + s^2$. It then follows that g^2 is identified by taking (twice) the difference in outcome correlation between monozygotic and dizygotic twins

$$\Delta\text{Corr}(Y_1, Y_2) = g^2/2.$$

The parameter s^2 can be recovered from any of the two twin correlations. In this simple framework, the parameters g^2 and s^2 (but also u^2) allow for an easy interpretation and measure how much outcome variation is due to genetic, common environmental and specific environmental variation.[1]

While twins are often used this way to distinguish genetic from environmental causes, there is a sizable literature that calls into question the twin method stressing that corresponding twin decompositions rely on strong assumptions. First, the assumption that G and S are independent goes against the widespread believe that children with favorable health endowments tend to grow up in families with favorable environments ($\text{Corr}(G_1, S_2) = \text{Corr}(G_2, S_1) \geq 0$). Second, the assumption that family environments of monozygotic and dizygotic twins are comparable does not hold if families treat monozygotic twins more similarly than dizygotic twins ($\text{Corr}(S_1, S_2)_{MZ} \geq \text{Corr}(S_1, S_2)_{DZ}$). Third, the assumption that dizygotic twins (and full siblings) share half of their genes is problematic if there is assortative mating of parents, genetic dominance, and gene-gene interactions ($\text{Corr}(G_1, G_2)_{DZ} \geq 1/2$). The main gist of the twin critique is that the decomposition estimates are likely biased due to questionable assumptions. The bias, however, can occur in any direction and depends on which assumption is violated (most). If family genes and environment interact, or if monozygotic twins share more family environment than dizygotic twins do, the decomposition estimates wrongfully favor genes. If dizygotic twins share

[1] With assumed indepence between G, S and U, we can decompose the outcome variance into three additive components ($1 = g^2 + s^2 + u^2$).

more than half their genes, then this biases in favor of family environment.

To meet the critics, we explore what twins and full siblings (raised together) can identify when we relax some of the assumptions and rely on more general models. Again, the standard procedure to obtain estimates of g and s is to measure the correlation between the outcomes of two siblings and to compare these correlations between different types of siblings pairs. A general representation of this correlation comparison can be written as

$$\Delta\text{Corr}(Y_1, Y_2) = g^2\Delta\text{Corr}(G_1, G_2) + 2gs\Delta\text{Corr}(G_1, S_2) + s^2\Delta\text{Corr}(S_1, S_2). \quad (4)$$

Without further assumptions, it follows that the difference in correlation between the outcomes of different sibling pairs is uninformative about g and s.

With twin data it is possible to relax two of the three restrictive assumptions and obtain an alternative estimate of g. If we let G and S be correlated with each other, dizygotic twins share some but not all of their genes, but assume homogenous family environments for monozygotic and dizygotic twins ($\text{Corr}(S_1, S_2)_{MZ} = \text{Corr}(S_1, S_2)_{DZ}$), then the difference in correlation between the outcomes of monozygotic and dizygotic twins in (4) simplifies to

$$\Delta\text{Corr}(Y_1, Y_2) = g^2\Delta\text{Corr}(G_1, G_2) + 2gs\Delta\text{Corr}(G_1, S_2).$$

since the homogenous family environment assumption removes $s^2\Delta\text{Corr}(S_1, S_2)$ from equation (4). The equation that remains can be interpreted as a test for genetic effects; that is, if the data show that monozygotic twins are more highly correlated on observed outcomes than dizygotic twins, it must be that genes matter ($g \geq 0$).[2]

[2]A positive correlation difference imply $\Delta\text{Corr}(G_1, G_2) \geq 0$ and $\Delta\text{Corr}(G_1, S_2) \geq 0$. This holds whenever ($a$) monozygotic twins share more genes than dizygotic twins; and (b) genes and common environments covary more within individuals than within sibling pairs $\text{Corr}(G_1, S_1) \geq \text{Corr}(G_2, S_1)$. With monozygotic twins ($G_1 = G_2$) and dizygotic twins ($G_1 \neq G_2$) , these two conditions automatically hold.

With sibling data, we can follow a similar decomposition strategy and compare full siblings who are close and far apart in age to obtain an alternative estimate of s. Sibling data allow us to relax all three restrictive assumptions. Instead, we assume that sibling spacing is exogenous. If we let closely spaced (CS) and widely spaced (WS) siblings share the same amount of genes ($\text{Corr}(G_1, G_2)_{CS} = \text{Corr}(G_1, G_2)_{WS}$), families treat siblings close in age more similar than (they would treat) siblings far apart in age ($\text{Corr}(S_1, S_2)_{CS} \geq \text{Corr}(S_1, S_2)_{WS}$), and part of the common family treatment reflect the interaction between the shared genes and family environment ($\text{Corr}(G_1, S_2)_{CS} \geq \text{Corr}(G_1, S_2)_{WS}$), we can express the difference in outcome correlation between closely and widely spaced siblings as

$$\Delta\text{Corr}(Y_1, Y_2) = 2gs\Delta\text{Corr}(G_1, S_2) + s^2\Delta\text{Corr}(S_1, S_2).$$

Although not all sibling correlations between G and S have been specified, we can still test for a common environmental effect; that is, if the data show that siblings who are close in age are more alike on observed outcomes than siblings who are far apart in age, it must be the common environment shared by siblings matters and varies with the sibling's age difference ($s \geq 0$).

The results produced by restrictive twin decomposition methods are obviously difficult to interpret. We show, however, that it is possible to give an alternative, potentially more meaningful, interpretation to comparable decomposition estimates combining data on twins and full siblings (raised together).

4 Data

In this paper we use data on twin and non-twin siblings. The sample of twin siblings stems from the Norwegian Twin Registry, which houses the main twin panels associated with different research institutes in Norway. In this study we use information on twin siblings for the twin cohorts born between 1915 and 1960

collected by the Norwegian Institute of Public Health (NIPH). Information on zygostity is available for approximately 78 percent of the twin pairs. These data has been described in more detail in Harris et al. (2002).

The sample of (non-twin) siblings is drawn from the Norwegian Population Register. This register contains information on all Norwegian citizens who were alive in 1954. This amounts to about 7.3 million individuals born between 1855 and 2008. Sibship, which is established through the mother, is identified for those individuals whose mother was alive in 1954 or later.[3] Siblings born in the same month to the same mother are classified as twins. We restrict the sample to non-twin siblings born between 1915 and 1960.

Both sibling samples are then matched to the Cancer Registry of Norway. This cancer registry collects individual level data from 1954 to 2007. Reporting to the cancer registry is mandatory (and done by clinicians and pathologists) and the completeness of registrations for solid tumors is close to 100 percent (Cancer Registry of Norway, 2007; Larsen et al., 2009). Information is available on the date of diagnosis, location of the tumor (encoded by ICD-10), stage at diagnosis (metastasis), the date the death certificate was issued (if the patient has died) and whether cancer was the main cause of death.

Figure 1 shows the prevalence of the ten most common cancer types in our sample encoded by the first three digits of the ICD-10 (International Classification of Diseases, 10th revision) codes. Breast and prostate cancers are clearly the most common cancer types among women and men. About 30 percent of all women have been diagnosed with breast cancer. About 20 percent of all the men in our sample have been diagnosed with prostate cancer. If we ignore that breast and prostate cancers are gender specific, breast and prostate cancers remain the most common

[3]Because women survive to older ages, more women than men have missing mothers. In the whole population file 50.6 percent of the individuals are men and 49.4 are women. When conditioning on non-missing mother, the fraction of men increases to 52.4 whereas the fraction of women decreases to 47.6. When looking separately at the 1855-1950 birth cohorts, only 41.4 percent of the women have non-missing mothers compared to 58.6 percent of the men.

cancer types. As a comparison, we see that about 10 to 11 percent of all the men and women have been diagnosed with either colorectal or skin cancer, which are the second most common cancer types among all women and men.

We work with samples in which the average age when diagnosed with cancer is quite young. Since we rely on twins born between 1915 and 1960 with cancer data available up to 2007, the youngest twins diagnosed with cancer are 47 years old. This is comparable to the age composition of the twin birth cohorts used in Lichtenstein et al. (2000).[4] In our non-twin sibling sample the average age when diagnosed with cancer for the first time is 51 years. Since siblings enter the sample when their mother was alive in 1954 or later, young siblings with cancer are systematically oversampled. This drives the average age at first cancer diagnosis downwards. In the entire cancer population we find that the average age when diagnosed with cancer for the first time is 67 years.

5 Results

In our empirical analysis we will primarily focus on sibling correlations of cancer mortality, overall cancer risk and cancer risks for the most common cancer sites; that is, we will report twin and sibling correlations of breast and cervical/uterine cancer for women, prostate and testicular cancer for men, and colorectal, skin, lung and leukemia cancer for men and women pooled together. All correlations for monozygotic twins, dizygotic twins and (non-twin) full siblings are contained in Table 1, where correlations for full siblings are also estimated separately for full siblings with different age differences ranging from siblings who are very close in age (Δa <2.5) to siblings who are far apart in age ($2.5 \leq \Delta a < 10$). In addition, we estimate correlations based on either twin or sibling samples in which there are at

[4]Lichtenstein et al. (2000) use data on Swedish twins born between 1886 and 1958; on Danish twins born between 1970 and 1930; and on Finnish twins born between 1880 and 1958. Moreover, they have cancer data from Sweden until 1995; from Denmark until 1993; and from Finland until 1996. This implies that their youngest twins diagnosed with cancer are 37 years old in Sweden; 63 years old in Denmark; and 38 years old in Finland.

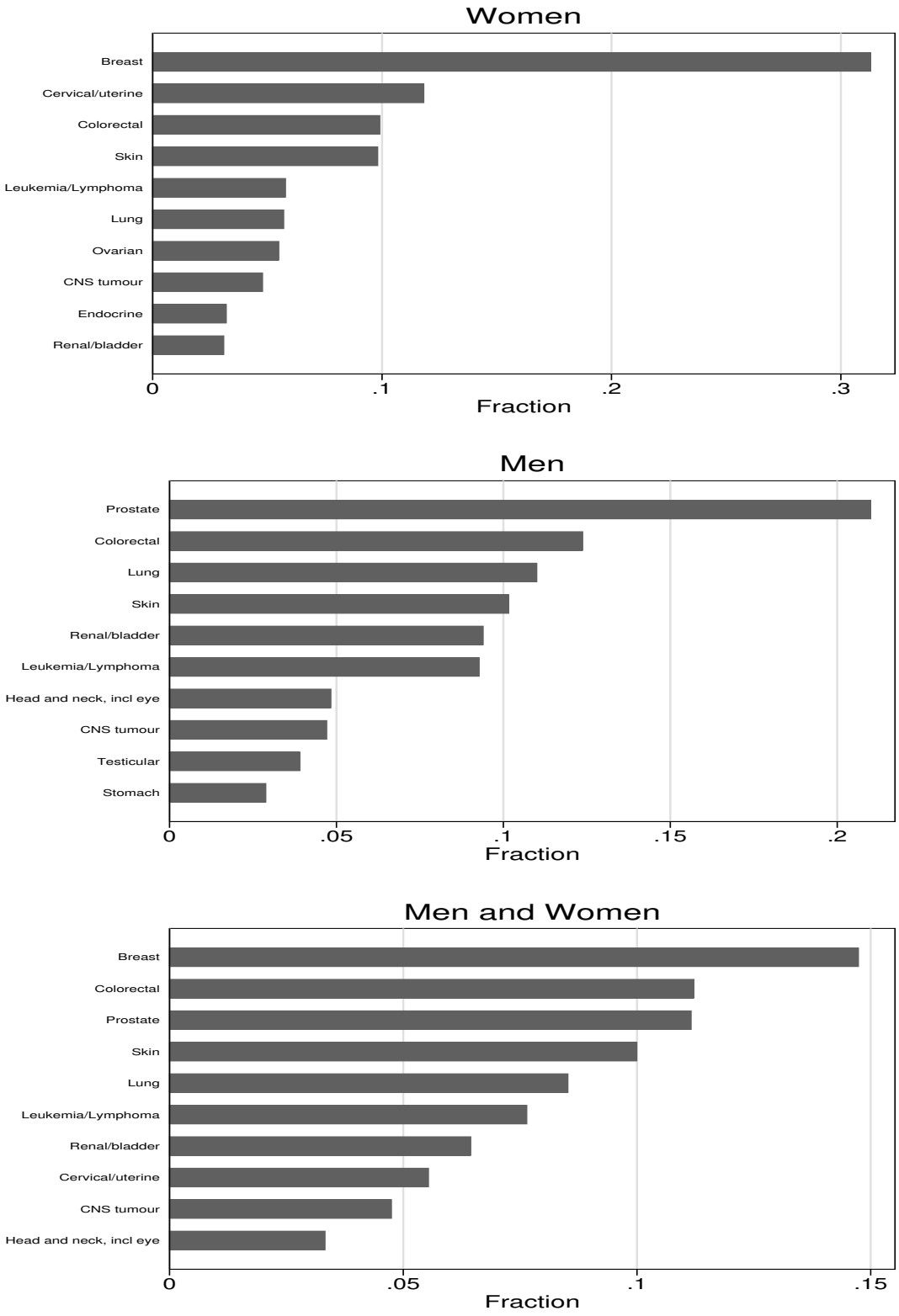

Figure 1. Distribution of the cancer types in our sample

least 4 pairs with the same cancer type (as in Lichtenstein et al., 2000), and write those correlation coefficients not statistically significant at the 10 percent level in italics.

Twin correlations are shown in the first two columns. Cancer mortality and overall cancer risk are clearly correlated among twins. All the correlations we find (twelve in total) are positive and statistically significant at the ten percent level. It is apparent that cancer mortality and overall cancer risk are more correlated among monozygotic twins than among dizygotic twins, regardless of whether we look at twin brothers, twin sisters or all twins together. When we compare these cancer correlations between twin brothers and sisters, however, we do find that cancer mortality and overall cancer risk correlations, as well as the difference in correlations between monozygotic and dizygotic twins, are larger for men than for women. When we zoom in on the most common cancer types, we find again that the correlations for monozygotic twins are amost always (twice or more times) larger than for dizygotic twins, with three exceptions. The correlations for lung and skin cancer among twins, including twin brothers and sisters, are significantly positive and similar for monozygotic and dizygotic twins. The correlations for cervical cancer among twin sisters are also similar, but very close to zero and not statistically significant at the ten percent level.

If monozygotic and dizygotic twins share the same common environment, these findings allow us to draw three empirical conclusions: (i) genetic factors play a prominent role in predicting overall cancer risks (most notably for men), breast, prostate and colorectal cancer as well as cancer mortality; (ii) environmental factors shared among twins seem to play a more prominent role for lung and skin cancer; and (iii) because of correlations close to zero, cervical/uterine cancer is neither driven by genetic factors nor by environmental factors shared among twins. If monozygotic and dizygotic twins do not share the same common environment, however, we should be more careful in drawing conclusions.

Table 1. Sibling correlations in cancer for siblings born between 1915 - 1960.

				Age difference between full siblings (in years)		
	MZ (1)	DZ (2)	FS (3)	Δa <2.5 (4)	$2.5 \leq \Delta a < 5$ (5)	$5 \leq \Delta a < 10$ (6)
Brothers and Sisters						
N sibpairs	4,275	9,926	743,260	184,181	250,290	231,233
Cancer mortality	0.1820	0.0751	0.0359	0.0370	0.0348	0.0384
All cancers	0.1890	0.1058	0.0500	0.0524	0.0500	0.0497
Colorectal	0.1211	0.0352	0.0173	0.0170	0.0195	0.0180
Skin	0.0310	0.0362	0.0165	0.0233	0.0159	0.0138
Lung	0.0712	0.0708	0.0217	0.0255	0.0269	0.0171
Leukemia	0.0925	*-0.0087*	0.0055	0.0029	0.0046	0.0050
Sisters						
N sibpairs	2,225	3,132	135,506	39,680	53,195	46,320
Cancer mortality	0.1226	0.0711	0.0276	0.0295	0.0276	0.0287
All cancers	0.1175	0.0851	0.0434	0.0435	0.0451	0.0430
Breast	0.1416	*0.0166*	0.0344	0.0353	0.0360	0.0341
Cervical/uterine	*0.0073*	*0.0020*	0.0071	0.0013	0.0136	0.0043
Brothers						
N sibpairs	2,050	3,269	224,559	54,184	73,810	71,287
Cancer mortality	0.2280	0.0734	0.0496	0.0542	0.0489	0.0488
All cancers	0.2578	0.1219	0.0709	0.0747	0.0721	0.0713
Prostate	0.2102	0.1033	0.0844	0.0831	0.1065	0.0871
Testicular	0.1142	0.0569	0.0120	0.0158	0.0114	0.0094

Note: The correlations written in italics are not statistically significant at the 10 percent level.

Sibling correlations can be used to assess (in part) the relevance of twins sharing the same common environment. The idea is fairly simple. If siblings close in age encounter more common family influences than siblings far apart in age, we can test whether wider age gaps between siblings correspond to smaller siblings correlations. In Table 1 we report sibling correlations in the last four columns, where the correlations in the last three columns are calculated on sibling samples stratified on the siblings' age difference (measured in years). When we treat dizygotic twins, who are born exactly the same age, as the reference group and compare cancer correlations of siblings with zero age gap to those of siblings with an age gap of at least one year, we find that cancer mortality, overall cancer risk and most of the common cancer forms are more correlated among dizygotic twins than among non-twin siblings. Notable exceptions are the correlations for leukemia, breast and cervical/uterine, which are insignificantly smaller for twins. When we compare correlations between non-twin siblings across different age gaps, however, most of the sibling correlations for cancer mortality, overall cancer risk and common cancer forms are very similar. The correlations in skin and testicular cancer appear slightly higher, the smaller the age difference, but never in a meaningful way (i.e., the difference is never statistically significant). All these sibling correlations indicate that a common environment matters, but only when it is shared among siblings who are born exactly the same age.

At first sight, these sibling results appear puzzling. On one hand, we find larger cancer correlations for dizygotic twin siblings than for non-twin siblings, which suggests that common environmental factors matter for most of the cancer risks we observe in our data. On the other hand, we find similar cancer correlations for siblings close and far apart in age, which suggests that environmental factors shared among siblings do not matter, at least not those common environmental factors that vary by the siblings' age gap. Since genetic resemblance is similar for dizygotic twins and non-twin siblings, we should look for other environmental factors (independent

Table 2. Variance decompositions with MZ and DZ twins – Results for the 1915-1960 birth cohorts

	g^2	s^2	u^2
Brothers and Sisters			
Cancer mortality	0.2138	0	0.7862
All cancers	0.1664	0.0226	0.811
Colorectal	0.1718		0.8282
Skin	0	0.031	0.969
Lung	0.0008	0.0704	0.9288
Leukemia	0.2024	0	0.7976
Sisters			
Cancer mortality	0.103	0.0196	0.8774
All cancers	0.0648	0.0527	0.8825
Breast	0.25	0	0.75
Cervical/uterine	0.0106	0	0.9894
Brothers			
Cancer mortality	0.3092	0	0.6908
All cancers	0.2718	0	0.7282
Prostate	0.2138	0	0.7862
Testicular	0.1146	0	0.8854

of how close siblings are in age) responsible for the observed difference in cancer correlations. One likely candidate is the prenatal environment, which only twin siblings share. If experiences in utero (and the first few months of live) are important for the further development of siblings, as the work of Almond and Currie (2011) suggests, we should find larger cancer correlations for dizygotic twin siblings and smaller, but similar, cancer correlations for siblings close and far apart in age.

This brings us back to the twin model. If the common environmental component has prenatal origins, it is not a priori clear whether monozygotic twins encounter more common influences in utero than dizygotic twins. If we assume they are not, which seems not unreasonable, we can attribute the observed difference in cancer correlations between monozygotic and dizygotic twins to the influence of genes and how genes interact with the common environment. If we further impose the initial

twin assumptions $\text{Corr}(S_1, G_1)_{MZ,DZ} = \text{Corr}(S_1, G_2)_{MZ,DZ} = 0$, $\text{Corr}(G_1, G_2)_{MZ} = 1$ and $\text{Corr}(G_1, G_2)_{DZ} = 1/2$, we can estimate the variance decomposition model as explained above and provide a basis to better compare our cancer results to those reported in other cancer twin studies (Holm et al. 1980; Verkasalo et al. 1999; Lichtenstein et al. 2000). The estimates are presented in Table 2.[5] Most of the estimates confirm our previous findings. Genetic factors dominate common environmental factors in explaining cancer mortality, overall cancer risks and most common cancer forms, including breast, prostate and colorectal cancers. The opposite is true for lung and skin cancer. Most of the cancer variation, however, we attribute to the unshared environment.[6] These findings are much in line with Lichtenstein et al. (2000), who also report that heritability dominates shared environment in explaining variation in lung cancer.

6 Conclusion

In this paper we have presented correlations for common cancer types and cancer mortality among twin and non-twin sibling pairs and used the variation in which they are genetically connected and exposed to family environments to identify genetic and environmental influences on cancer risks and mortality. Our results indicate that genes dominate over shared environment in explaining relatively more of the variation in cancer at most common cancer sites (but lung and skin cancer) and cancer mortality. The vast majority of the variation in cancer and cancer mortality, however, is explained by individual (unshared environmental) factors.

Notwithstanding our attempt to decompose observable cancer risks into unobservable genetic and environmental components, we (as social scientists) would rather know why these unobservable components affect cancer risks and cancer mortality.

[5]The reader not interested in these comparisons may turn to the next concluding section.

[6]The level of statistical significance is calculated by using a linear probability model where the dependent variable is a dummy variable which equals one if individual i had/has cancer and zero otherwise, and the only explanatory variable is a dummy variable which equals one if your sibling had/has cancer and zero otherwise.

While our decomposition estimates tell us little about what it is that constitutes these unobservable genetic and environmental components, they do provide us with some clues on where to look for potentially successful inputs. In this context, biologists and medical researchers started to explore genes as explanatory variables in regression models of heritable cancer risks. One example is the influential study of Ford et al. (1998), which successfully links mutations in BRCA1 or BRCA2 genes to increased breast cancer risk. This is also how we see our work; as a first step towards a better understanding of the environmental origins of cancer providing clues on potential covariates that may (or may not) help to reduce cancer risks and increase cancer survival. Our future work will focus attention to the socio-economic channels and explore whether measurable (and potentially successful) environmental inputs such as individual education and/or transitory shocks in family income when children are in utero have an impact on cancer risks. Our unique data certainly allow us to investigate these links.

References

Almond, D. and Currie, J. (2011). Human capital development before age five. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4, chapter 15, pages 1315–1486. Elsevier Science Publishers B.V.

Eberl, M., Sunga, A., Farrell, C., and Mahoney, M. (2005). Patients with a family history of cancer: Identification and management. *The Journal of the American Board of Family Medicine*, 18(3):211.

Eloranta, S., Lambert, P., Cavalli-Bjorkman, N., Andersson, T., Glimelius, B., and Dickman, P. (2010). Does socioeconomic status influence the prospect of cure from colon cancer – A population-based study in Sweden 1965–2000. *European Journal of Cancer*, 46(16):2965–2972.

Ferlay, J., Autier, P., Boniol, M., Heanue, M., Colombet, M., and Boyle, P. (2007).

Estimates of the cancer incidence and mortality in Europe in 2006. *Annals of oncology*, 18(3):581.

Ford, D., Easton, D., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D., Weber, B., Lenoir, G., Chang-Claude, J., et al. (1998). Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *The American Journal of Human Genetics*, 62(3):676–689.

Goldberger, A. S. (1979). Heritability. *Economica*, 46(184):327–347.

Holm, N., Hauge, M., and Harvald, B. (1980). Etiologic factors of breast cancer elucidated by a study of unselected twins. *Journal of the National Cancer Institute*, 65(2):285–298.

Hopper, J., Bishop, D., and Easton, D. (2005). Population-based family studies in genetic epidemiology. *The Lancet*, 366(9494):1397–1406.

Jencks, C. (1980). Heredity, environment, and public policy reconsidered. *American Sociological Review*, pages 723–736.

Lichtenstein, P., Holm, N., Verkasalo, P., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A., and Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer: Analyses of cohorts of twins from Sweden, Denmark, and Finland. *New England Journal of Medicine*, 343(2):78.

Lindström, L., Hall, P., Hartman, M., Wiklund, F., Grönberg, H., and Czene, K. (2007). Familial concordance in cancer survival: a Swedish population-based study. *The Lancet Oncology*, 8(11):1001–1006.

Manski, C. F. (2011). Genes, eyeglasses, and social policy. *The Journal of Economic Perspectives*, 25(4):83–93.

Pharoah, P., Day, N., Duffy, S., Easton, D., and Ponder, B. (1997). Family history

and the risk of breast cancer: A systematic review and meta-analysis. *International journal of cancer*, 71(5):800–809.

Steinberg, G., Carter, B., Beaty, T., Childs, B., and Walsh, P. (1990). Family history and the risk of prostate cancer. *The Prostate*, 17(4):337–347.

Verkasalo, P., Kaprio, J., Koskenvuo, M., and Pukkala, E. (1999). Genetic predisposition, environment and cancer incidence: A nationwide twin study in Finland, 1976-1995. *International Journal of Cancer*, 83(6):743–749.

Statistics Norway

Postal address:
PO Box 8131 Dept
NO-0033 Oslo

Office address:
Akersveien 26, Oslo
Oterveien 23, Kongsvinger

E-mail: ssb@ssb.no
Internet: www.ssb.no
Telephone: + 47 62 88 50 00

ISSN: 1892-753X

Statistisk sentralbyrå
Statistics Norway

Design: Siri Boquist