Twin Research and Human Genetics (2019), **22**, 817–823
doi:10.1017/thg.2019.71

CAMBRIDGE UNIVERSITY PRESS

Article

The Nordic Twin Study on Cancer — NorTwinCan

Jennifer R. Harris[1], Jacob Hjelmborg[2], Hans-Olov Adami[3,4], Kamila Czene[3], Lorelei Mucci[5], Jaakko Kaprio[6] and Nordic Twin Study of Cancer (NorTwinCan) Collaboration

[1]Division of Health Data and Digitalisation, Norwegian Institute of Public Health, Oslo, Norway, [2]Danish Twin Registry, Institute of Public Health, University of Southern Denmark, Odense, Denmark, [3]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, [4]Clinical Effectiveness Research Group, Institute of Health, University of Oslo, Oslo, Norway, [5]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA and [6]Department of Public Health and Institute for Molecular Medicine, University of Helsinki, Helsinki, Finland

Abstract

Nordic twin studies have played a critical role in understanding cancer etiology and elucidating the nature of familial effects on site-specific cancers. The NorTwinCan consortium is a collaborative effort that capitalizes on unique research advantages made possible through the Nordic system of registries. It was constructed by linking the population-based twin registries of Denmark, Finland, Norway and Sweden to their country-specific national cancer and cause-of-death registries. These linkages enable the twins to be followed many decades for cancer incidence and mortality. To date, two major linkages have been conducted: NorTwinCan I in 2011–2012 and NorTwinCan II in 2018. Overall, there are 315,413 eligible twins, 57,236 incident cancer cases and 58 years of follow-up, on average. In the initial phases of our work, NorTwinCan established the world's most comprehensive twin database for studying cancer, developed novel analytical approaches tailored to address specific research considerations within the context of the Nordic data and leveraged these models and data in research publications that provide the most accurate estimates of heritability and familial risk of cancers reported in the literature to date. Our findings indicate an excess familial risk for nearly all cancers and demonstrate that the incidence of cancer among twins mirrors the rate in the general population. They also revealed that twin concordance for cancer most often manifests across, rather than within, cancer sites, and we are currently focusing on the analysis of these cross-cancer associations.

Keywords: Twins; cancer; NorTwinCan

(Received 30 June 2019; accepted 3 July 2019; First Published online 12 September 2019)

For more than half a century, data from the Nordic twin cohorts have helped elucidate the importance and nature of familial effects on the development of cancers (Ahlbom et al., 1997; Harvald & Hauge, 1963; Lichtenstein et al., 2000; Mucci et al., 2016). Earlier twin studies confirmed reports of familial clustering for cancers at common sites (Easton, 1994; Lynch et al., 1995) and further extended these findings to conclude that familial effects played a major role for cancers at all sites and for total cancer (Ahlbom et al., 1997). The prevailing interpretation, based on data from single countries, was that these familial influences primarily reflected the effects of shared environments (Ahlbom et al., 1997; Harvald & Hauge, 1963). However, to fully exploit the twin design for probing the nature of familial effects on cancer requires larger samples of concordant and discordant pairs than are typically available through single-country studies.

Findings from a landmark study (Lichtenstein et al., 2000) that leveraged twin and cancer data from the Swedish, Danish and Finnish registries and included nearly 45,000 twin pairs also emphasized the primacy of environmental influences for most types of cancers. However, that study also noted an increased risk of cancer to twins whose co-twin had developed certain types of cancers, including stomach, colorectal, lung, breast and prostate cancer. Furthermore, heritability estimates were moderately large for prostate, colorectal and breast cancer. Although the population-based sample of twins studied in this multicountry initiative was quite large and the confidence intervals for the heritability estimates were wide, heritability could not be calculated for the less common cancers, and statistical power to parse familial sources of cancer clustering into genetic and shared environmental influences was limited.

Moreover, a common methodological limitation of most previous twin cancer studies was that analyses did not take into account the considerable amount of censoring that can occur at both the beginning and end of follow-up, as well as competing causes of death — which is particularly important in epidemiological studies of cancer, given the relatively late-life incidence of most forms of cancer. Ignoring such censoring can severely bias the incidence and risk concordance estimates and can affect estimates of heritability.

To overcome these shortcomings and enable more in-depth analyses of the genetic and environmental influences on cancer, we established the NorTwinCan (Nordic Twin Study of Cancer) consortium (http://nortwincan.org). This collaborative effort

Author for correspondence: Jennifer R. Harris, Email: Jennifer.Harris@fhi.no

Cite this article: Harris JR, Hjelmborg J, Adami H-O, Czene K, Mucci L, and Kaprio J Nordic Twin Study of Cancer (NorTwinCan) Collaboration. (2019) The Nordic Twin Study on Cancer — NorTwinCan. *Twin Research and Human Genetics* **22**: 817–823, https://doi.org/10.1017/thg.2019.71

Table 1. Overview of follow-up dates for NorTwinCan I and II and current number of incident cases

	Denmark	Finland	Norway	Sweden
Follow-up				
Cancer registration start date	January 1943	February 1974	January 1964	April 1961
End of follow-up NorTwinCan I	December 2009	December 2010	December 2008	December 2009
End of follow-up NorTwinCan II	December 2016	December 2016	December 2016	December 2016
N incident cases NorTwinCan II	16,172	9061	5638	26,365

Note: NorTwinCan II is an updated database of NorTwinCan I, and therefore, includes all number of incident cases as of December 2016.

builds heavily on the previous Nordic twin registry collaboration (Lichtenstein et al., 2000) and unites cancer epidemiologists, biostatisticians and twin researchers from multiple institutes spanning the Nordic countries and the United States. NorTwinCan capitalizes on unique research advantages made possible through the Nordic system of registries in Denmark, Finland, Norway and Sweden. In its initial phase, NorTwinCan established the world's most comprehensive twin database for studying cancer. It expanded the previous large Nordic twin study base (Lichtenstein et al., 2000) with the addition of the Norwegian twin cohort, and 10 more years of follow-up for cancer incidence, providing a median follow-up of 32 years. We refer to these data as NorTwinCan I; subsequent updates to the data are described later as NorTwinCan II.

An important new feature of NorTwinCan is that we developed novel analytical models tailored to address specific methodological considerations within the context of the Nordic data. This included accounting for competing risks of death, for left-censoring that arises due to variable initiation of the time of cancer registration, and right-censoring whereby different potential outcomes could explain an individual's status at follow-up (Scheike et al., 2014, 2015). These approaches were applied in analyses that provide the most accurate estimates of heritability (Hjelmborg, Korhonen et al., 2017; Hjelmborg et al., 2014; Moller et al., 2016; Mucci et al., 2016) and familial risk (Mucci et al., 2016) of cancers reported in the literature to date.

One of the most compelling findings to emerge from our initial analysis of the NorTwinCan data revealed that twin concordance for cancer most often manifests across, rather than within, cancer types (Mucci et al., 2016); namely, if one twin has cancer, the co-twin is at an increased risk to develop cancer, but this cancer usually occurs at a different site. In NorTwinCan, cancer was diagnosed in both twins among 1383 monozygotic (MZ) pairs and among 1933 dizygotic (DZ) pairs. However, only 38% of the MZ pairs and 26% of the DZ pairs were diagnosed with the same type of cancer. These findings prompted a new NorTwinCan substudy to investigate familial risk of cross-cancer associations.

To help maintain NorTwinCan as a world-class research resource, and with novel plans to analyze the cross-cancer associations, we expanded the NorTwinCan database through a renewed round of registry linkages that were completed from 2016 through 2018. The national twin registries were linked to the national cancer registries in Denmark, Finland, Norway and Sweden. These updates provided additional years of follow-up across the four countries and more than doubled the number of incident cancers to more than 57,000. A preliminary description of the updated data is provided in Table 1. Analyses of the cross-cancer associations are now underway using the updated data, which will greatly enhance our statistical power.

NorTwinCan Cohorts

The NorTwinCan study was constructed by linking the population-based twin registries of Denmark, Finland, Norway and Sweden to their country-specific national cancer and cause-of-death registries. These Nordic twin registries are continually evolving as new studies, data collections and data updates are conducted. A recent NorTwinCan publication (Skytthe et al., 2019) describes information about the participating twin registries including the dates of establishment, birth cohorts and sample sizes. Further, more detailed information is also provided in this special issue of the journal for the country-specific registries.

Each country participating in NorTwinCan maintains a system of national registers through which information about vital status, date and cause of death, migration and systematic registration of cancer diagnosis and other cancer-related variables are registered. Record linkage between the information in the twin registries with data in the cancer and other population registries can be performed because every citizen in the Nordic countries is assigned an individually unique national registration number. Conducting such linkage for research purposes requires full compliance with the General Data Protection Regulation of the European Union (EU-GDPR, 2016) and typically involves obtaining permissions from the relevant entities including the national data protection authorities, regional ethics committees and the boards of the twin registers.

The first linkages for NorTwinCan I were conducted in 2011–2012. At that time, the cancer registration was complete through 2008 for Norway, 2009 for Denmark and Sweden and 2010 for Finland (Mucci et al., 2016). The NorTwinCan I database included information on more than 350,000 twin individuals, including 80,309 MZ twins, 123,382 same-sex DZ (SSDZ) twins and 96,499 opposite-sex DZ (OSDZ) twins. Through linkage to the cancer and cause-of-death registries in the participating countries the median follow-up time for cancer incidence and mortality was 28.3 years from 1943 to 2010. There were 62,522 individuals who died of any cause, 3804 individuals who emigrated and were lost to follow-up, and zygosity information was missing for 5375 individuals. Altogether, we identified 27,156 incident cancers among 23,980 individuals.

The NorTwinCan II linkage update was completed in 2018 and included all cancer registrations through 2016 for Denmark, Finland and Norway, and Sweden. Table 1 shows the number of additional years of follow-up this encompassed for each country. NorTwinCan II includes 315,418 eligible twins and 57,236 incident cancer cases, and 58 years of follow-up, on average, calculated as the weighted number of contributing cases from each cohort.

The core data obtained from the cancer registries consists of information about the occurrence of cancer covering 40 different and most common cancer sites based on the NordCan

classification of sites (Engholm et al., 2010). All diagnoses are based on the International Classification of Diseases, 10th Revision (ICD-10) coding from the population-based registries, with the NordCan grouping of ICD codes enabling comparability across countries (http://www-dep.iarc.fr/nordcan.htm). In addition, the participating twin registries have exposure measures, including information on body composition at multiple ages for each individual and lifestyle factors known to affect cancer risk, such as tobacco smoking status and alcohol consumption, are also available in at least half of the cohort.

Development of Statistical Models

The first important consideration in modeling twin data to quantify genetic and environmental contributions to cancer occurrence is to decide on how to treat the timing of events within the pair. The initiation of population-wide registration of cancer diagnoses and the collection of retrospective and prospective data on twins in each population provides critical information regarding the timing of events, diagnosis and vital status at follow-up for each member of the pair. The information available from the beginning of cancer registration is the date of diagnosis and status of the twins, and whether the co-twin of a diagnosed twin is alive at follow-up, dead or also diagnosed with a certain cancer, and the age at the co-twins' diagnosis.

The classic approach to parsing sources of variation into genetic and environmental effects relies on methods that compare within pair dependence between MZ and DZ pairs. This was described over a century ago in a landmark paper (Fisher, 1918) and applies naturally to (continuous) traits observed completely in all individuals. Historically, as shown in the first twin study of cancer outcomes with Nordic twin data from the Danish cohort (Harvald & Hauge, 1963), initial modeling attempts of the observed data relied on dichotomizing the time to event as either having a specific cancer diagnosis or not and comparing rates of concordance by zygosity. In that paper the authors noted that results may be influenced by unobserved cancer occurrences due to censoring at follow-up (Harvald & Hauge, 1963).

Although Harvald and Hauge (1963) are inconclusive regarding genetic and environmental influences for the risk to develop cancer at various anatomical sites, the notion that the risk of cancer in a twin given that their co-twin has had the cancer (before follow-up or death) highlights an important measure of dependence that has been applied in subsequent studies in the field. In the later Nordic follow-up (Lichtenstein et al., 2000) in which some 90,000 Nordic twins were studied, the sources of variation in liability to develop cancer was studied through the (bivariate) liability-threshold model in which prevalence (again at follow-up) determines the threshold on the latent trait of liability to cancer. This approach provides within-pair correlations, termed *tetrachoric correlations*, having the properties that within-pair dependence is independent of the prevalence of the cancer. The tetrachoric correlations may be compared by zygosity and further modeled directly by the polygenic biometric model from quantitative genetics, often referred to as the 'ADCE-model'. This classic biometric modeling decomposes the variance of the trait, here liability to develop cancer, into independent additive (A) and dominant (D) genetic contributions to the variance, and environmental contributions of shared effects, C, and unique individual effects including measurement error, E. This model estimates heritability in the sense described by Fisher in 1918 as the amount of variation in liability to develop cancer that is explained by genetic variation. However, this variation can change during the course of follow-up because cancer onset varies, for instance, by age. Thus, it is critical to include time to diagnosis in the biometric modeling of twin cancer data.

The biometric models developed for the NorTwinCan study allows for studying how genetic and environmental sources of variation to risk of cancer diagnosis vary over time, typically by age (Scheike et al., 2014, 2015). All results take censored observations into account and hence will be independent of time of follow-up. Further, the survival analysis approach allows for competing risks typically death before diagnosis, the 'alive-illness-death' scenario. Cancer-site specific characteristics — for instance, the cumulative incidence function by age — may be compared to that of the background population. Within-pair dependence is qualified by the measures of case-wise concordance and of relative recurrence risk that varies with age for MZ and DZ pairs. Hence, the approach to model the risk of cancer determinants described by Harvald and Hauge (1963) is extended to include variation by age. This allows the estimation of the time varying heritability of risk of cancer as, for example, in prostate cancer by age (Hjelmborg et al., 2014). The choice of risk scale allows for directly comparable familial risk; for example, the case-wise concordance in pairs that reflects the risk of cancer before a certain age conditional on cancer onset in the co-twin before that age is also applicable to siblings. Furthermore, the classic biometric modeling approach described earlier in terms of ADCE variance components is applied to the risk of cancer occurrence, taking time to diagnosis, censoring or death into account (Holst et al., 2016), and has been applied in the NorTwinCan analyses (Mucci et al., 2016). The key assumption in the bivariate time-to-event modeling of twin cancer data is that pairs are censored at the same time, which is indeed the case for the registry data. The theory, including examples, is discussed in Scheike et al. (2014, 2015), and the accompanying R-package 'mets' implementing the survival analysis methods is available from the CRAN library (Holst et al., 2016).

The matched case co-twin design allows for studying the association of the risk of cancer with exposures effectively controlling for unobserved confounding. For instance, data on MZ pairs discordant for smoking are analyzed to study the direct effects of smoking on lung cancer (Hjelmborg et al., 2017). Further, as demonstrated in the article, the design allows for analyzing whether genetic effects moderate the influence of smoking and vice versa, that is, whether gene-by-smoking effects are influential for risk of lung cancer.

Main Findings from NorTwinCan I

Prior to the updated linkages, a number of studies were published based on the data available in NorTwinCan I. The main paper summarizing the heritability and familial risk of cancer was published in *JAMA* in 2016. The lifetime incidence of cancer in the cohort was estimated to be 32%, based on 27,156 cancers diagnosed among 23,980 twin individuals from same-sex pairs over a follow-up period of 32 years on average. The familial risk, that is, the cancer incidence in the co-twin given cancer in the first twin, was estimated at 37% in DZ pairs and 46% in MZ pairs. The familial risks were higher than the individual risks at all ages and were also greater among the MZ than the DZ pairs at all ages. This pattern indicates that a genetic liability to cancer is present throughout life (Mucci et al., 2016). We show elsewhere (Skytthe et al., 2019) that overall mortality and cancer incidence in the twins is comparable to the background populations, thus permitting results from the twin studies to be generalized to the population at large.

The first Nordic analysis of cancer (Lichtenstein et al., 2000) provided evidence for a genetic component for three of the most common cancers (breast, prostate, colon). This list was expanded based on NorTwinCan I analyses to also include melanoma and nonmelanoma skin cancer, corpus uteri, ovary and kidney cancers (Mucci et al., 2016). The NorTwinCan I analyses also considered colon and rectum as separate sites. A more detailed analysis of colorectal cancer (Graff et al., 2017) showed that the heritability estimate for colorectal cancer as a single entity was 40% (95% CI [33, 48]), the risks for colon cancer and rectal cancer specifically being smaller. Indeed, the *JAMA* paper (Mucci et al., 2016) highlights that much of familial cancer risk is across sites, with a minority of pairs where both twins were diagnosed with cancer having cancer at the same site. Even among MZ pairs, only 38% (522 out of 1383 cancer concordant pairs) had cancer at the same site, compared to 26% (496 out of 1933 cancer concordant pairs) of DZ pairs. Deeper analyses of these cross-site concordances were not examined and are currently the focus of analyses being conducted with the NorTwinCan II data. Heritability was estimated using a model that included additive genetic (A) and non-shared (E) environmental influences, common environment (C) was also included in the model when there were five or more concordant pairs. The point estimates for shared environment were greater than zero for seven sites, but significant for lung cancer only (24%, 307 95% CI [7, 40]) and breast (16%, 95% CI [10, 32]) (Mucci et al., 2016). Notably for overall cancer, the point estimate for shared environment was zero despite more than adequate power to detect such an effect — the heritability estimates for all cancers being 33% (95% CI [30, 37]).

In addition to colorectal cancer (Graff et al., 2017), we have examined familial risk and heritability of breast (Moller et al., 2016) and prostate cancers (Hjelmborg et al., 2014) in more detail. With a total of 3933 breast cancer cases in NorTwinCan I, we observed greatly increased familial risk in both DZ pairs (20%) and MZ pairs (28%) compared to the lifetime incidence among individuals, estimated at 8%. This yielded a heritability estimate of 31% overall, which was slightly lower when premenopausal (27%) and postmenopausal (22%) cancers were considered separately. From age 50 onward, the MZ familial risk was consistently higher than the DZ familial risk, and both were higher than the individual risk, and the heritability estimate did not vary by age (Moller et al., 2016). Notably, as individuals, MZ and DZ female twins do not differ in incidence. In a further analysis, twins from OSDZ pairs had the same breast cancer risk as those from SS pairs (Ahrenfeldt et al., 2015), which did not support the 'twin testosterone transfer' hypothesis of intrauterine hormonal influences on cancer development. Ahrenfeldt et al. (2015) showed similar results for other female (ovary, corpus uteri, cervix and other female genitals) and male (prostate, testis) cancers, further indicating lack of support for the testosterone hypothesis.

For the prostate cancer analysis (Hjelmborg et al., 2014), we had data on 4109 prostate cancer cases diagnosed in MZ, SSDZ and males from OSDZ NorTwinCan pairs through 2009, which is a slightly different analysis sample than reported in Mucci et al. (2016). The differences in prostate cancer screening practices between Denmark and the other Nordic countries result in a lower incidence in Denmark; therefore, some analyses contrasted Denmark with the other three countries. Thus, in addition to the lifetime risk for prostate cancer being expectedly lower in Denmark, the case-wise concordances were lower in Denmark. Despite that, the heritability estimate was somewhat higher (59%) in Denmark than for Finland, Norway and Sweden combined (52%). No evidence for shared environmental effects overall was found, but it may be present at younger ages (under 75 years). An analysis by age suggested that heritability was quite constant with age.

In the NorTwinCan results described earlier, specific exposures were not considered. One of the major risk factors for many cancers is smoking. Therefore, our first risk factor analysis considered the effect of smoking on lung cancer incidence and familial risk. For this purpose, we identified the twins with information on smoking and followed them up for lung cancer incidence. In the four Nordic cohorts, 115,407 twins (of whom 43,512 were MZ) had smoking data, and we found 1508 incident lung cancers (Hjelmborg, Korhonen et al., 2017). Expectedly, lung cancer risk was strongly dependent on smoking status at baseline, with no differences by zygosity and no differences by sex among never smokers. The overall nine-fold higher risk among current smokers was higher in men than in women. With only one concordant MZ pair for lung cancer among more than 18,000 pairs of never smokers, the heritability of lung cancer among never smokers could not be estimated.

Most lung cancer concordant pairs were observed among pairs in which both twins were current smokers at baseline. In such pairs, we observed a higher concordance in MZ than DZ pairs. Among current smoking pairs, the heritability of liability to lung cancer was 0.41 (95% CI [0.26, 0.56]) under an AE model, and 0.29 under an ACE model. Finally, we examined lung cancer risk in twin pairs in which one was a smoker at baseline and the other had never smoked. There were 35 such discordant MZ pairs with lung cancer incident in one twin; among them the lung cancer was diagnosed in the current smoker for 31 pairs, in contrast to 4 pairs in which the never smoker had lung cancer. The hazard ratio was 6.0 (95% CI [2.1, 17.3], $p = .001$). These results confirm the causal role of smoking in lung cancer independent of genetic factors (Hjelmborg, Korhonen et al., 2017). We are now continuing the analyses to other cancers that are considered to be tobacco-related, but for which the evidence for causality is less convincing, as described here.

Current Ongoing Work — Cross-Cancer Associations

Analyses of the cross-cancer associations on the NorTwinCan II data are ongoing. Model development was completed while awaiting the updated data and preliminary analyses were conducted using the NorTwinCan I data. Through a series of analyses, we mapped the cross-cancer occurrences across 40 cancer sites. The analyses of the within-pair associations took censoring and competing risk of death into account. To explore the nature of the cross-cancer associations we computed the following:

Cross-cancer concordances that represent the lifetime risk that a pair will develop a particular combination of cancers.

Cross-case-wise concordance that estimates the risk to one twin for a specific cancer given that the co-twin has a different specific cancer.

Relative recurrence risks that provide information about how likely the particular combination of cancer co-occurrences is in the pair compared to nonrelated individuals.

Co-heritabilities of cancers estimated as twice the difference in lifetime concordance risk to the joint variation in risk. These estimates reflect the importance of genetic influences for explaining the relationships between sets of cancers.

Average within-pair differences between cancer occurrences, using time to event analyses.

The preliminary results were presented at the 16th Congress of the International Society of Twin Studies (ISTS) in Madrid in 2017 (Harris et al., 2017). These indicated excess familial risk for a great number of cancer co-occurrences, with estimates of coheritability clearly stronger for specific cancer clusters. Furthermore, certain cancers, such as prostate, breast and lung, tend to co-occur with cancers at multiple sites.

The median difference in age at diagnosis was significantly greater for some cancer co-occurrences and significantly less for others. For example, the age difference at diagnosis for brain and prostate cancer was 6.83 years among MZ pairs and 15.8 years among DZ pairs. In contrast, age differences in diagnosis for the co-occurrence of prostate and larynx cancer was 11.6 years among MZ pairs and 3.09 years among DZ pairs. These findings suggest that genetic effects influence the timing of disease development for certain sets of cancer while unique environmental factors could influence the timing of diseases for other sets.

Our findings of co-heritability for many of the cross-cancer associations may corroborate results from molecular studies. Such studies provide an ever-growing number of genes identified to be pleiotropic across cancers at different sites (Cheng et al., 2014; Hoadley et al., 2014; Jiang et al., 2019; Lim et al., 2014; Sampson et al., 2015; Setiawan et al., 2014; Sivakumaran et al., 2011). A potential advantage of the twin design could derive from comparisons between the twin-based and the molecular-based results that can help elucidate the nature of the factors mediating the cross-cancer occurrences. Co-heritable effects that are not reflected in molecular findings could signal genes or clusters of genes that have not been identified. And cross-cancer associations that show familiarity, but not significant co-heritability, could signal common environmental influences that affect the risk of developing cancer at specific sites.

We have extended the work on lung cancer and smoking (Hjelmborg, Korhonen et al., 2017) to other tobacco-related cancers, as the evidence of the causal association of smoking with some cancer types is less well established than for lung cancer. We used the same data set as for the lung cancer analysis with never (n = 59,093), former (n = 21,168) or current (n = 47,314) smokers. The focus was on incident cancer from the following sites — bladder, esophagus, kidney, larynx, liver, oral cavity, pancreas and pharynx — that showed an increased risk among smokers in the NorTwinCan data base. Within-pair analyses of these individual sites using smoking discordant pairs mostly found increased estimates of risk in the smoking twin. But for several sites the number of pairs discordant for both smoking at baseline and for incident cancer was small and estimates were nominally nonsignificant. When combined as 'tobacco-related cancers', the within-pair association was strong and statistically significant even in MZ pairs. This indicates that a common exposure such as smoking can result in different cancers even when genetic and shared familial environments are controlled for. A paper based on these results is being finalized, and preliminary results have been presented (Korhonen et al., 2017).

Cross-cancer analysis of brain with other cancers is another area where our investigation of cancer concordance in twins is revealing new findings. Despite the large sample size and median of 32 years of follow-up in NorTwinCan I, concordance for cancers of the brain and central nervous system (CNS) was rare. Only four pairs (one MZ and three DZ pairs) were concordant for brain and CNS cancers, corresponding to what would be expected by chance. However, when we extended our analyses to cross-cancer occurrences of brain with other cancers, a different picture of genetic risk begins to emerge. For example, the risk of brain cancer is more than doubled among MZ twins whose co-twin had skin cancer. We also found an elevated concordance risk for brain cancer with cancers of the prostate, breast, colon, kidney and leukemia. These cross-cancer findings suggest that genetic effects are significant across age at onset (Hjelmborg, Kaprio et al., 2017), and will be compared to findings emerging regarding genetic pleiotropy across cancer sites.

Data Management, Ethics and Data Protection

In our experience with NorTwinCan, we find that moving and updating large datasets across different national boundaries and institutes is both complicated and cumbersome, requiring considerable effort and resources to ensure data security. Solutions for access can also be burdensome and restrict the analytical potential of the collaborating scientists. An ideal solution would bypass data transfer and exploit modern technologies that offer a single access point to the datasets at participating locations. It also needs to ensure data protection as is now formalized under the EU data protection legislation (EU-GDPR, 2016).

One option under development is to establish data access through the Nordic Tryggve collaboration project (https://neic.no/tryggve/). Tryggve is funded through the Nordic e-Infrastructure Collaboration NeIC (branch of NordForsk), and the ELIXIR research infrastructure nodes of the Nordic countries in NorTwinCan. The goals are to develop a data management platform for sensitive datasets that could satisfy the needs of research and researchers using the NorTwinCan database. Tryggve is coordinated by the Centre for Scientific Computing CSC, which is a Finnish governmental body under the Ministry of Education offering high-capacity data management and computation. Under Tryggve, a researcher would access the datasets stored remotely in each of the providing locations through a remote desktop and process them in ePouta, a secure cloud for sensitive data provided by CSC. The datasets themselves would be hosted by their respective owners and secure, read-only access would be provided by Secure Data Access Service (DAS). Of note is that the datasets would not be copied from one location to another, but rather be accessible as a streaming service, which retains the control of the datasets with their respective owners.

Furthermore, all the researchers of the project would use the same common virtual environment to access and process the data. At the moment, a pilot demonstration of the system exists, and we are working with Tryggve to implement the system. This model will help extend data sharing and access beyond the scope of any individual funded project and help build sustainable solutions for collaboration. It could also serve as a model for other twin and family dataset collaborations.

Discussion and Future Work

The NorTwinCan work summarized earlier extended the datasets and analyses reported for the Nordic countries by Lichtenstein et al. (2000) in several important ways. The follow-up time was extended for the Danish, Finnish and Swedish cohorts, and the Norwegian twins were added. Thus, the analysis of cancer in twins covers a large fraction of all twins in the Nordic countries. Furthermore, analyses of the NorTwinCan I data revealed that the cancer incidence and overall mortality of the twin cohorts is representative of the Nordic countries. The cumulative incidence of any cancer is about one in three persons and provides a reliable estimate of the population risk of cancer, while taking into

consideration deaths from competing causes. Because cancer is common, the patterns of cancer within families need to be viewed within this population context. There are multiple sites where cancer can occur, and different histologies and degrees of malignancy. Investigating these patterns in both MZ and DZ pairs provides insights into the relative contributions of genes and environment to variation in the liability to develop cancers overall and to the most common cancers. Thus, we expanded findings showing that the genetic component is significant across a greater number of cancer sites than previously revealed, and for the most common cancers (lung, breast, prostate and colorectal) we investigated particular features of their genetic epidemiology. The risk of cancer diagnosis before 100 years of age and the lifetime risk reported in Mucci et al. (2016) for 40 main sites fits those in the corresponding background population, assuring external validity of the twin model. This is not the case for corresponding lifetime risks obtained from previously applied methodology (Lichtenstein et al., 2000).

A key novel finding was that the majority of twin pairs concordant for cancer were discordant for the cancer sites. This held true even for MZ pairs, suggesting a general genetic proneness or resistance, but that the actual tissue that develops cancer may be more likely to be determined by nongenetic factors and stochastic effects. This has led us to investigate the cross-site correlations of cancer incidence, and to quantify to what degree reflect genetic factors shared by different cancers (as defined by site), that is, genetic pleiotropy or whether there are shared environmental determinants (e.g., smoking exposure). These analyses are now ongoing using the recent update, NorTwinCan II.

Finally, the data have been used to investigate whether hormone-dependent and reproductive cancers such as breast and prostate may be influenced by co-twin sex. The twin testosterone transfer hypothesis posits that female twins may be exposed in utero to testosterone and be masculinized, which would affect their risk of developing certain cancers. Using information on SS and OS twin pairs, Ahrenfeldt et al. (2015) found no evidence to support the hypothesis. This hypothesis continues to be of general interest in potentially explaining some sex differences, and we can test more specific hypotheses on selected cancers with the NorTwinCan II update.

Estimates of the contribution of inter-individual genetic differences to interindividual differences in risk of developing cancer have been generated using family and twin data over the past decades. Much more recently, there has been an explosion in the availability of measured genotypes using both array techniques (to measure and impute common variants) and next generation sequencing to measure rarer variants and whole genome sequences. Genome-wide association studies of germline DNA from cancer cases and controls have led to the identification of tens and hundreds of loci or single nucleotide polymorphisms (SNPs) associated with cancer risk and opened a window into the genetic variation underlying inherited risk to develop cancer. Furthermore, novel statistical approaches have permitted the estimation of the contribution of all measured and imputed genotypes to disease risk, a measure known as SNP heritability. Comparisons of SNP-heritabilities and twin heritabilities can give rise to hypotheses on the origins of familial aggregations and sources of genetic variation.

The Nordic countries have immediate potential to take precision medicine forward (Njølstad et al., 2019), and the Nordic Twin study on Cancer represents a major resource in the Nordic portfolio of genetic, environmental and medical registers and databases. Another set of resources are the biobanks in each country. For example, the pathology archives from the past 30 years for virtually all cancer cases in Finland are available through the network of national biobanks. This became possible through new legislation, permitting more direct access by both academic and commercial researchers to samples. Combining information on family relationship, lifestyle and environmental exposures with medical register data on medications and comorbid conditions will permit unparalleled investigations of the causes of cancer.

Financial support. JK has been supported by the Academy of Finland (grants 308248 and 312073), and the Sigrid Juselius Foundation. JRH and JBH have been supported (2017–2019) by grants from the Nordic Cancer Union, project entitled 'Genetic Epidemiology and Familial Risk of Cross-Cancer Associations: A Nordic Twin Study'. NorTwinCan was supported in part by the Ellison Family Foundation. JBH was supported by the AgeCare program of the Academy of Geriatric Cancer Research, OUH.

Conflicts of interest. None.

Ethical standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

References

Ahlbom, A., Lichtenstein, P., Malmstrom, H., Feychting, M., Hemminki, K., & Pedersen, N. L. (1997). Cancer in twins: Genetic and nongenetic familial risk factors. *Journal of the National Cancer Institute*, *89*, 287–293.

Ahrenfeldt, L. J., Skytthe, A., Moller, S., Czene, K., Adami, H. O., Mucci, L. A., … Lindahl-Jacobsen, R. (2015). Risk of sex-specific cancers in opposite-sex and same-sex twins in Denmark and Sweden. *Cancer Epidemiology, Biomarkers & Prevention*, *24*, 1622–1628.

Cheng, I., Kocarnik, J. M., Dumitrescu, L., Lindor, N. M., Chang-Claude, J., Avery, C. L., … Peters, U. (2014). Pleiotropic effects of genetic risk variants for other cancers on colorectal cancer risk: PAGE, GECCO and CCFR consortia. *Gut*, *63*, 800–807.

Easton, D. F. (1994). The inherited component of cancer. *British Medical Bulletin*, *50*, 527–535.

Engholm, G., Ferlay, J., Christensen, N., Bray, F., Gjerstorff, M. L., Klint, A., … Storm, H. H. (2010). NORDCAN — A Nordic tool for cancer information, planning, quality control and research. *Acta Oncologica*, *49*, 725–736.

EU-GDPR. (2016). The EU General Data Protection Regulation (GDPR). Retrieved from https://eur-lex.europa.eu/eli/reg/2016/679/oj

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, *52*, 399–433.

Graff, R. E., Moller, S., Passarelli, M. N., Witte, J. S., Skytthe, A., Christensen, K., … Hjelmborg, J. B. (2017). Familial risk and heritability of colorectal cancer in the Nordic Twin Study of Cancer. *Clinical Gastroenterology and Hepatology*, *15*, 1256–1264.

Harris, J. R., Kaprio, J., Czene, K., Mucci, L., Pukkala, E., Christensen, K., … NortwinCan. (2017, November). *Familial risk of cross-cancer associations: Findings from the Nordic Twin Study of Cancer (NorTwinCan).* Poster presented at the 16th Congress of the International Society of Twin Studies, Madrid, Spain.

Harvald, B., & Hauge, M. (1963). Heredity of cancer elucidated by a study of unselected twins. *JAMA*, *186*, 749–753.

Hjelmborg, J., Kaprio, J., Korhonen, T., Mucci, L., Christensen, K., Adami, H. O., … Harris, J. R. (2017, November). On familial risk of brain tumor occurrence: The Nordic Twin Cancer Study. Poster presented at the 16th Congress of the International Society of Twin Studies, Madrid, Spain.

Hjelmborg, J., Korhonen, T., Holst, K., Skytthe, A., Pukkala, E., Kutschke, J., … Nordic Twin Study of Cancer. (2017). Lung cancer, genetic predisposition and smoking: The Nordic Twin Study of Cancer. *Thorax*, *72*, 1021–1027.

Hjelmborg, J. B., Scheike, T., Holst, K., Skytthe, A., Penney, K. L., Graff, R. E., ... Mucci, L. A. (2014). The heritability of prostate cancer in the Nordic Twin Study of Cancer. *Cancer Detection and Prevention*, *23*, 2303–2310.

Hoadley, K. A., Yau, C., Wolf, D. M., Cherniack, A. D., Tamborero, D., Ng, S., ... Stuart, J. M. (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*, *158*, 929–944.

Holst, K. K., Scheike, T. H., & Hjelmborg, J. (2016). The liability threshold model for censored twin data. *Computational Statistics*, *93*, 324–335.

Jiang, X., Finucane, H. K., Schumacher, F. R., Schmit, S. L., Tyrer, J. P., Han, Y., ... Lindstrom, S. (2019). Shared heritability and functional enrichment across six solid cancers. *Nature Communications*, *10*, 431.

Korhonen, T., Hjelmborg, J., Bonat, W., Holst, K., Skytthe, A., Pukkala, E., ... NorTwinCan. (2017, November). Smoking and cancer: The Nordic Twin Study of Cancer. Poster presented at the 16th Congress of the International Society of Twin Studies, Madrid, Spain.

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., ... 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*, 78–85.

Lim, U., Kocarnik, J. M., Bush, W. S., Matise, T. C., Caberto, C., Park, S. L., ... Le Marchand, L. (2014). Pleiotropy of cancer susceptibility variants on the risk of non-Hodgkin lymphoma: The PAGE consortium. *PLoS One*, *9*, e89791.

Lynch, H. T., Fusaro, R. M., & Lynch, J. (1995). Hereditary cancer in adults. *Cancer Detection and Prevention*, *19*, 219–233.

Moller, S., Mucci, L. A., Harris, J. R., Scheike, T., Holst, K., Halekoh, U., ... Hjelmborg, J. B. (2016). The heritability of breast cancer among women in the Nordic Twin Study of Cancer. *Cancer Epidemiology, Biomarkers & Prevention*, *25*, 145–150.

Mucci, L. A., Hjelmborg, J. B., Harris, J. R., Czene, K., Havelick, D. J., Scheike, T., ... Nordic Twin Study of Cancer Collaboration (2016). Familial risk and heritability of cancer among twins in Nordic countries. *JAMA*, *315*, 68–76.

Njølstad, P. R., Andreassen, O. A., Brunak, S., Borglum, A. D., Dillner, J., Esko, T., ... Stefansson, K. (2019). Roadmap for a precision-medicine initiative in the Nordic region. *Nature Genetics*, *51,* 924–930.

Sampson, J. N., Wheeler, W. A., Yeager, M., Panagiotou, O., Wang, Z., Berndt, S. I., ... Chatterjee, N. (2015). Analysis of heritability and shared heritability based on genome-wide association studies for thirteen cancer types. *Journal of the National Cancer Institute*, *107*, djv279.

Scheike, T. H., Hjelmborg, J. B., & Holst, K. K. (2015). Estimating twin pair concordance for age of onset. *Behavior Genetics*, *45*, 573–580.

Scheike, T. H., Holst, K. K., & Hjelmborg, J. B. (2014). Estimating twin concordance for bivariate competing risks twin data. *Statistics in Medicine*, *33*, 1193–1204.

Setiawan, V. W., Schumacher, F., Prescott, J., Haessler, J., Malinowski, J., Wentzensen, N., ... Le Marchand, L. (2014). Cross-cancer pleiotropic analysis of endometrial cancer: PAGE and E2C2 consortia. *Carcinogenesis*, *35*, 2068–2073.

Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J. G., Zgaga, L., Manolio, T., ... Campbell, H. (2011). Abundant pleiotropy in human complex diseases and traits. *American Journal of Human Genetics*, *89*, 607–618.

Skytthe, A., Harris, J. R., Czene, K., Mucci, L., Adami, H. O., Christensen, K., ... Pukkala, E. (2019). Cancer incidence and mortality in 260,000 Nordic twins with 30,000 prospective cancers. *Twin Research and Human Genetics*, *22*, 99–107.