Received: 28 December 2021 | Accepted: 17 January 2022

DOI: 10.1002/ijc.33963

CANCER EPIDEMIOLOGY

IJC INTERNATIONAL JOURNAL of CANCER

Cancer in twin pairs discordant for smoking: The Nordic Twin Study of Cancer

Tellervo Korhonen[1] | Jacob Hjelmborg[2,3] | Jennifer R. Harris[4] | Signe Clemmensen[2,3] | Hans-Olov Adami[5,6] | Jaakko Kaprio[1] | On behalf of the Nordic Twin Study of Cancer (NorTwinCan) collaboration

[1]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland

[2]Department of Epidemiology, Biostatistics and Biodemography, University of Southern Denmark, Odense, Denmark

[3]The Danish Twin Registry, University of Southern Denmark, Odense, Denmark

[4]Center for Fertility and Health, The Norwegian Institute of Public Health, Oslo, Norway

[5]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[6]Clinical Effectiveness Group, Institute of Health, University of Oslo, Oslo, Norway

Correspondence
Tellervo Korhonen, University of Helsinki, Institute for Molecular Medicine Finland, PO Box 20, FIN-00014, Helsinki, Finland.
Email: tellervo.korhonen@helsinki.fi

Funding information
Academy of Geriatric Cancer Research, Odense University Hospital AgeCare program; Juselius Foundation; Nordic Cancer Union; Nordic Union of Cancer; Terveyden Tutkimuksen Toimikunta, Grant/Award Numbers: 129680, 213506, 263278, 265240, 312073; The Ministry for Higher Education, Grant/Award Number: Swedish Twin Registry

Abstract

The discordant twin pair study design is powerful to control for familial confounding. We employed this approach to investigate the associations of smoking with several cancers. The NorTwinCan study combines data from the Danish, Finnish, Norwegian and Swedish twin and cancer registries. Follow-up started when smoking status was determined and ended at cancer diagnosis confirmed by information in the cancer registry, death or end of follow-up. We classified the participants as never (n = 59 093), former (n = 21 168) or current (n = 47 314) smokers. We pooled data from twin pairs where one co-twin was diagnosed with any of the following tobacco-related cancers: esophagus, kidney, larynx, liver, oral cavity, pancreas, pharynx or urinary bladder, while their co-twin had none of those. Lung cancer was included in further analysis. We used Cox regression allowing for pair-specific baseline functions to estimate hazard ratios (HRs) with 95% confidence intervals (CIs). For tobacco-related cancer sites, we recorded 7379 cases during median 27 years of follow-up. The analyses based on individual twins showed that former (HR 1.31, 95% CI: 1.17-1.48) and current (HR 2.14 [1.95-2.34]) smokers are at increased risk to develop one of cancers listed above, compared to never smokers. Among 109 monozygotic twin pairs discordant for cancer and smoking, the HR was 1.85 (95% CI: 1.15-2.98) among current smokers and 1.69 (1.00-2.87) among former smokers when compared to their never smoking co-twin. Thus, associations of smoking with several cancers were replicated for discordant identical twin pairs. Analyses based on genetically informative data provide evidence consistent with smoking causing multiple cancers.

KEYWORDS
cancer, case-co-twin design, smoking, twins

Abbreviations: CI, confidence interval; DZ, dizygotic; GWAS, genome-wide association study; HR, hazard ratio; IARC, The International Agency for Research on Cancer; ICD, International Classification of Diseases; MR, Mendelian randomization; MZ, monozygotic; NorTwinCan, The Nordic Twin Study of Cancer; RCT, randomized controlled trial; UZ, unknown zygosity.

What's new?

Unlike lung cancer, a causal association between smoking and other cancer sites that are associated with smoking remains to be established. In this Nordic Twin Study of Cancer, the authors employed the discordant twin pair study design to study the associations of smoking with several cancers while controlling for familial confounding. The analysis of genetically informative twin data indicates that cigarette smoking is a risk factor for multiple cancers, independent of familial confounding. The causative role of cigarette smoking in cancer is thus not limited to lung cancer but appears to extend to several other cancer types.

1 | INTRODUCTION

Globally, the total deaths due to noncommunicable diseases, including several types of cancer, continue to increase and comprised the greatest proportion of deaths worldwide in 2017.[1] Among the risk factors ranked by risk-attributable disability-adjusted life-years, cigarette smoking comes in second place after high systolic blood pressure.[1,2] Cigarette smoking accounts for a fifth of all global cancer deaths and has been established as a risk factor for at least 16 types of cancer, dominated by lung cancer.[3] Many cancers are more common among men, and this is partially attributed to higher rates of smoking among men,[4] but the basis for sex differences is not fully known.[5]

The Nordic Twin Study of Cancer (NorTwinCan) includes population-based cohorts from the Danish, Finnish, Norwegian and Swedish twin registries.[6,7] Each person in the Nordic countries is assigned a unique national registration number, which enables the individual-level information in the twin registries to be linked to data in the cancer and mortality registries which can further be linked to complete follow-up information, drop-out being only due to death or emigration. Thus, NorTwinCan[8] is a unique and valuable data resource for investigating the relative contribution of genetic and environmental factors to the risk of developing cancer. Our earlier study based on the NorTwinCan data provided strong evidence that smoking significantly impacts the risk for lung cancer independent of genetic liability for smoking and lung cancer.[9]

For several other cancer sites, such as squamous cell esophageal cancer, larynx cancer and bladder cancer, support for causality is strong while some uncertainty prevails regarding cancers of colon, kidney, liver, nasal area, oral cavity, pancreas, pharynx, rectum, stomach, as well as the cervix uteri and ovary and uterine adnexa for women only.[3,10-14] To date, studies conducted within the NorTwinCan have not investigated smoking as a risk factor for other types of cancer and we are not aware of other twin studies that have examined the causal role of smoking in the development of these other tobacco-related cancers. To this end, we leverage the NorTwinCan data with information on smoking. Specifically, we estimate the relative risk of specific cancers among smokers and then examine the pair-wise risk among smoking-discordant pairs. This permits investigation of the causal contribution of smoking while controlling for shared genetic and environmental factors that influence the risk of developing cancer. Specifically, if the association with smoking is replicated among monozygotic pairs—who share their genes and childhood environment but are discordant for both smoking and cancer—then the association between cancer and smoking is likely unconfounded by shared genetic and environmental influences.[15]

To summarize, our aim was to test if the previously reported associations between site-specific cancers and smoking are independent of familial confounding. The great strength of the within-pair design is that it naturally controls for age, sex (if using same-sex pairs), year of birth and the familial factors (both genetic and nongenetic) that are shared within twin pairs and may affect both the exposure and the outcome.

2 | MATERIALS AND METHODS

2.1 | Twin cohorts

The NorTwinCan data include 54 431 complete twin pairs [19 450 monozygotic (MZ) and 34 981 dizygotic (DZ)] and 18 713 incomplete ("half") pairs (4616 MZ and 14 097 DZ). Our analyses include twin individuals of known zygosity with data on smoking status assessed prior to cancer diagnosis. Cancer diagnoses were obtained from the national cancer registries in the participating countries. Follow-up of cancer incidence started from the time of the baseline questionnaire study when smoking status was assessed in the respective country cohorts. For individuals who reported smoking behavior over time in multiple questionnaires, we used the earliest available information. In all country cohorts, zygosity—MZ or DZ—was determined at baseline by validated questionnaire methodology, which classifies more than 95% of twin pairs correctly.[8] Twins who did not reply to the questionnaires, as well as a minority providing inconsistent responses, were classified as unknown zygosity (UZ). We excluded individuals from opposite-sex DZ pairs due to an insufficient amount of data collected from these pairs.

Descriptive characteristics of the NorTwinCan data included in the analyses by cohort, zygosity, sex, smoking status and some follow-up details are shown in Table 1. The following birth cohorts were included: Denmark 1870 to 1982, Finland 1880 to 1957, Norway 1915 to 1960 and Sweden 1886 to 1958. The range of years of data collection—starting with the first year of assessment of

smoking when cancer incidence follow-up starts to the end of follow-up—was 1959 to 2010 in Denmark, 1975 to 2011 in Finland, 1980 to 2009 in Norway and 1961 to 2010 in Sweden. More details are provided in an earlier article[9] based on the same cohorts. For the present analyses, we were able to access data on an additional 12 164 DZ individuals with the same start and end of follow-up.

We classified the participants as never smokers (n = 59 093), former smokers (n = 21 168) and current smokers (n = 47 314) based on the survey items used to assess smoking status. Former smokers had smoked previously but had quit smoking by the baseline assessment. Current smokers were those who reported being smokers at the time of the assessment. Smoking data in the Danish cohort was derived from questionnaire surveys conducted repeatedly from 1959 to 2002. The survey items are described in earlier studies based on the Danish cohort.[16-18] In Finland, smoking data came primarily from the first questionnaire survey conducted in 1975, but some twins who had not replied in 1975 responded to the second survey in 1981. The survey items are described in earlier studies based on the Finnish cohort.[19-21] In Norway, smoking data was derived from three questionnaire surveys conducted in 1980 to 1982, 1990 to 1992 and 1998. The survey items are described in earlier studies based on the Norwegian cohort.[22,23] In the Swedish cohort, smoking data came from four questionnaire surveys that were conducted in 1961, 1967, 1970 and 1973. The survey items are described in earlier studies based on the Swedish cohort.[24,25]

2.2 | Tobacco-related cancers

We considered a cancer to be tobacco-related if previous studies report a relatively strong and robust association between smoking and cancer at that specific site.[3,10-14] Based on these findings, we conducted individual-based and pair-wise analyses for multiple cancer sites (bone marrow/acute myeloid leukemia, colon, esophagus, kidney, larynx, liver, lung, nasal cavity/sinus, oral cavity, pancreas, pharynx, rectum, stomach, bladder/other urinary organs) including two cancer sites (cervix uteri and ovary and uterine adnexa) relevant only for women. Site classification was based on the NordCan classification, which is The International Agency for Research on Cancer (IARC)-generated classification (available at https://www-dep.iarc.fr/nordcan/english/frame.asp). The "cancer dictionary" on the website provides the NordCan categories and International Classification of Diseases (ICD)-10 codes. Analyses were conducted for outcomes of site-specific cancers and on an index of whether the individual had received a positive diagnosis for any of the cancers under study.

2.3 | Pooled cancer sites

Pooling across cancer sites helps to increase statistical power, especially for the doubly discordant MZ pair analyses. Two sets of analyses, one based on prior literature and one essentially post hoc analysis, were conducted. First, based on prior literature[3,10-14] we pooled together tobacco-related cancer sites—13 excluding and 14 including lung cancer—as listed above. Here, we did not include the two cancers affecting women only (cervix uteri and ovary and uterine adnexa).

Second, for the post hoc analyses, we used the results of the individual-based and pair-wise analyses to identify those cancer sites where the association with smoking appeared independent of familial influences. This was based on the size of hazard ratio (HR) point estimates of pair-wise analyses being similar or higher compared to those of individual-based analyses. In other words, if the results of the individual-based analyses are replicated in the pair-wise analyses of

TABLE 1 Description of the NorTwinCan data included in the analyses by cohort, zygosity, sex, cancer incidence and smoking status

Country	Denmark	Finland	Norway	Sweden	Total
Birth cohort	1870-1982	1880-1957	1915-1960	1886-1958	
N individual twins	41 819	23 488	12 708	49 560	127 575
N (%) MZ twins	11 883 (28)	7361 (31)	5606 (44)	18 666 (38)	43 516 (34)
N (%) female twins	22 841 (55)	12 032 (51)	6863 (54)	26 767 (54)	68 503 (54)
Beginning of follow-up	1959	1975	1980	1961	
End of follow-up	2010	2011	2009	2010	
Median follow-up time (IQR), years	7.7 (7.6-7.7)	34.7 (26.5-34.7)	26.7 (23.5-27.3)	36.0 (25.9-36.0)	26.5 (7.7-34.7)
Median entry age (IQR), years	48.7 (36.1-58.8)	32.1 (24.4-45.5)	35.0 (28.4-47.1)	38.9 (27.5-47.4)	40.0 (28.8-52.2)
N incident cancers					
Any cancer site	4612	3574	1662	9729	19 577
Tobacco-related cancer sites	1820	1342	623	3594	7379
N (%) smoking status at baseline					
Current smoker	15 435 (37)	7571 (32)	5521 (43)	18 787 (38)	47 314 (37)
Former smoker	9134 (22)	3833 (16)	2630 (21)	5571 (11)	21 168 (17)
Never smoked	17 250 (41)	12 084 (51)	4557 (36)	25 202 (51)	59 093 (46)

data from pairs discordant for smoking and for cancer, this would indicate that smoking is a causal factor for cancer. In the present study, the term "causally tobacco-related" is used for brevity only and it means that the associations appear to be independent of shared familial influences. This interpretation follows from the co-twin control design because it provides a way to study the relationship between smoking and cancer and controls for familial factors that encompass shared genes (~50% for DZ pairs and 100% for MZ pairs) and shared environmental exposures (in MZ and DZ pairs) relevant for the smoking-cancer association.[15] Guided by the results of the first set of pooled analyses, the second set of pooled analyses included only the "causally tobacco-related" cancer sites.

2.4 | Concordance and discordance for cancer by smoking status

A twin pair is considered concordant for cancer if both twins shared the same specific cancer diagnosis or diagnostic grouping of pooled cancer sites in question. That holds for the specific tobacco-related cancers and the two groupings of pooled cancers, that is, all tobacco-related cancers and causally tobacco-related cancers. Cancers that were not tobacco-related were not considered as cases and were censored in survival analyses.

A twin pair is classified as discordant for cancer if one twin was diagnosed with a tobacco-related cancer and their co-twin was not diagnosed with that specific cancer. A corresponding classification was used when analyzing the groupings of tobacco-related cancers. Further, the pair is doubly discordant if, additionally, one twin is a current or former smoker while their co-twin is a never smoker. Analyses addressed two types of smoking-discordance, pairs that included former vs never smoker and pairs that were current vs never smoker. We assumed that former smokers might also have an elevated risk for cancer because of lifetime exposure to tobacco smoke, whereas the cancer risk among current smokers, who were still smoking at the time of the survey, was assumed to be even higher than among former smokers.

2.5 | Statistical analyses

The cumulative incidence of tobacco related cancers was estimated by smoking status using the Aalen-Johansen estimator and taking censoring and competing risk of death into account.[26] External validity of results from the twin cohort is reflected by the degree to which the cumulative incidence in the twin cohort represents that of the background population. To test the association of cancer occurrence with tobacco exposure adjusting for background variables (age at first interview, sex and country), the Cox proportional hazards regression model was applied with age as the time scale. Inference from this model was corrected for within-pair dependence. This analysis also examined the HR of cancers in women vs men after adjusting for smoking status and country cohort.

This approach for studying associations among the group of twins as individuals was extended to control for (unmeasured) shared confounding factors by comparing twins in pairs discordant for exposure and outcome. This was achieved by the stratified Cox regression model in which baseline hazard functions are pair specific.[27,28] Matching within MZ twin pairs controls to a certain extent for general genetic and early environmental (eg, family environments) effects. For DZ twins the matching is more open to interpretation given the association studied but matching for early environmental effects is presumable high. For this reason and that the matched design is vulnerable to nonshared confounding, the analysis is complemented by the univariate model described above in which the twins are treated as singletons (and inference corrected for within pair dependence). Thus, if an association between smoking and cancer is observed in both the individual-based and pairwise analyses, but particularly in discordant MZ pairs, this would provide strong evidence for an association between smoking and cancer.

To address sensitivity of results, we assessed the degree of unmeasured confounding effect needed to explain away the found association. For this we computed *E*-values for the observational, individual level associations[29] using an online calculator (https://www.evalue-calculator.com/). Analyses were carried out using the statistical software R.[30]

3 | RESULTS

3.1 | Specific cancer sites

3.1.1 | Results based on individuals

Table S1 shows the number of cancer cases and cumulative incidence rates by smoking status among men and women for each specific cancer site. With the exception of colon cancer and stomach cancer and ovarian cancer in women, the highest incidence rates were consistently observed for current smokers, followed by former smokers and were lowest for never smokers. The results of the time-to-event analyses based on individuals are shown first for each specific cancer site in Table S2. Comparisons between former and never smokers revealed fewer significantly elevated risks, these being for cancers of the lung (HR 4.18), nasal cavity/sinus (HR 2.89) as well as urinary bladder and other urinary organs (HR 1.79). Comparisons between current smokers and never smokers revealed significantly elevated HRs among current smokers for almost all cancer sites, except bone marrow/acute myeloid leukemia, colon and ovary and uterine adnexa. The greatest HRs were observed for lung (14.6), larynx (6.25) and nasal cavity/sinus (4.24) cancer.

3.1.2 | Results based on discordant pairs

The numbers of concordant and discordant pairs for each cancer site are shown in Tables S3A (pairs discordant for former smoking) and S3B (pairs discordant for current smoking). Notably, among MZ pairs there are very few or often no concordant pairs, that is, both co-twins having cancer.

Analysis of data from the doubly discordant pairs across the specific cancer sites yielded significantly elevated risk estimates for several cancers (Table S4). Comparisons between the twins who were a current smoker at the time of smoking assessment with their never smoking co-twins showed that smoking was most highly associated with cancer of the larynx (HR 9.31), lung (HR 7.60), oral cavity (HR 5.19) and pharynx (HR 4.46). Among the MZ pairs where one twin is a current smoker and the co-twin had never smoked, high and statistically significant estimates were found for lung (HR 13.8) and cervix uteri (HR 5.14). Furthermore, the risks for these two cancers were also elevated among the former smoking co-twins compared to the

never-smoking co-twins. As shown in Table S4, for other specific cancer sites, the MZ risk estimates did not reach statistical significance. Further, when comparing the risk estimates for current smokers between MZ and DZ pairs, the HR estimates are similar in magnitude for lung, esophagus, liver, pancreas and urinary bladder cancer.

3.2 | Pooled cancer sites

3.2.1 | Results based on individuals

Table 2 shows the number of cancer cases and cumulative incidence rates by smoking status for men and women for the tobacco-related and causally tobacco-related cancers pooled together. The results from the individual-based survival analysis showing the HR and their respective confidence intervals between the smoking groups are reported in Table 3 for the tobacco-related cancers pooled together, and for the causally tobacco-related cancers. When we restricted the

TABLE 2 Cases of first cancer diagnosis and cumulative incidence rate/1000 by pooled cancer sites, smoking status and sex

	Never smoker N = 59 093				Former smokerN = 21 168				Current smokerN = 47 314				
	Number of cases		Cumulative incidence rate/1000[a]		Number of cases		Cumulative incidence rate/1000[a]		Number of cases		Cumulative incidence rate/1000[a]		Total number of cases
Pooled cancer sites	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	
All tobacco-related[b] (including lung)	849	1677	70.4	55.4	912	316	98.3	67.4	2524	1101	142.5	107.1	7379
All tobacco-related[c] (excluding lung)	820	1578	68.0	52.0	782	259	83.6	54.4	1800	712	98.2	68.1	5951
Causally tobacco-related[d] (including lung)	352	658	28.9	20.7	501	146	55.6	33.5	1716	728	99.1	72.6	4101
Causally tobacco-related[e] (excluding lung)	323	559	26.5	17.2	371	89	40.9	20.4	992	339	54.7	33.6	2673

[a]Men: Time0 = 63.7 and Time1 = 77.2; Women: Time0 = 64.7 and Time1 = 80.
[b]Leukemia, colon, esophagus, kidney, larynx, liver, lung, nasal cavity/sinus, oral cavity, pancreas, pharynx, rectum, stomach, urinary bladder.
[c]Leukemia, colon, esophagus, kidney, larynx, liver, nasal cavity/sinus, oral cavity, pancreas, pharynx, rectum, stomach, urinary bladder.
[d]Esophagus, kidney, larynx, liver, lung, oral cavity, pancreas, pharynx, urinary bladder.
[e]Esophagus, kidney, larynx, liver, oral cavity, pancreas, pharynx, urinary bladder.

TABLE 3 Associations of smoking status with pooled sites of first cancer diagnosis: Individual-based survival analyses adjusted for sex and country-specific cohort

	Former smokers vs never smokers (HR = 1.00)		Current smokers vs never smokers (HR = 1.00)		
Pooled cancer sites	HR	95% CI	HR	95% CI	*E*-value[a]
Tobacco-related[b] (including lung)	1.33	1.23-1.43	2.18	2.06-2.30	3.78
Tobacco-related[c] (excluding lung)	1.17	1.09-1.27	1.56	1.46-1.65	2.49
Causally tobacco-related[d] (including lung)	1.69	1.52-1.87	3.63	3.36-3.93	6.72
Causally tobacco-related[e] (excluding lung)	1.31	1.17-1.48	2.14	1.95-2.34	3.70

[a]To assess the minimum degree of unmeasured confounding effect needed to explain away the found association, we computed *E*-values for the observational, individual level associations using an online calculator (https://www.evalue-calculator.com/).
[b]Leukemia, colon, esophagus, kidney, larynx, liver, lung, nasal cavity/sinus, oral cavity, pancreas, pharynx, rectum, stomach, urinary bladder.
[c]Leukemia, colon, esophagus, kidney, larynx, liver, nasal cavity/sinus, oral cavity, pancreas, pharynx, rectum, stomach, urinary bladder.
[d]Esophagus, kidney, larynx, liver, lung, oral cavity, pancreas, pharynx, urinary bladder.
[e]Esophagus, kidney, larynx, liver, oral cavity, pancreas, pharynx, urinary bladder.

individual-based analysis to the eight causally related pooled cancer sites (excluding lung cancer), the risk was elevated for both current smokers (HR 2.14; 95% confidence interval [CI]: 1.95-2.34) and former smokers (HR 1.31; 95% CI: 1.17-1.48). To assess the degree of unmeasured confounding, we computed *E*-values for the observational, individual-level associations.[29] These *E*-values range from 2.49 upward, which implies that an unknown confounder would have to have at least 2.5-fold associations with both the exposure and outcome to account for the observed association (Table 3).

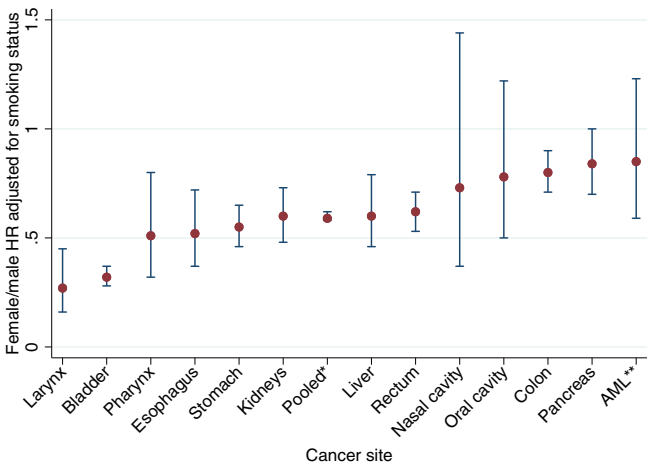

FIGURE 1 Hazard ratios (HR) of women for multiple cancer sites adjusted for smoking status. HR (women compared to men) is displayed on *Y* axis and cancer site is displayed on *X* axis. Cancer sites are ordered by HR from smallest to largest. *Cancer sites pooled together. **Bone marrow/acute myeloid leukemia

The hazard for tobacco-related cancers, adjusted for smoking status, was significantly lower among women (HR 0.50; 95% CI: 0.46-0.55) than men. The HR for women vs men for each cancer site, plus for pooled sites, was calculated adjusting for smoking status and country cohort. The HRs for women are shown in Figure 1 and in the Supportive table for Figure S1. All risk estimates were consistently lower in women than men, and significantly lower for esophagus, kidney, larynx, liver, pharynx, rectum, stomach, urinary bladder and for the eight cancer sites pooled together.

3.2.2 | Results based on discordant pairs

Table 4 shows the number of cancer-discordant and cancer-concordant pairs with one twin being a never and the other a former smoker (ie, smoking-discordant pairs) for the pooled cancer sites (ie, all tobacco-related cancers with and without lung cancer, plus causally tobacco-related cancers with and without lung cancer). Cancer-discordant pairs are further stratified by whether the former smoker or the never smoker had an incident cancer among the pooled ones. Table 5 shows the same information as Table 4, but with the smoking discordance reflecting pairs where one twin is a never and the other twin is a current smoker.

The results of the pooled analyses conducted for twin pairs discordant for smoking and cancer are shown in Table 6. The HRs reflect the association of one twin being a current or former smoker compared to their co-twin being a never smoker. Here, we focused on MZ twin pairs where one twin had none of the tobacco-related cancers, but their co-twin was diagnosed with any of the tobacco-related cancers (ie, 14 cancer sites including lung cancer and 13 sites excluding lung cancer). As a post hoc analysis, we also provide pair-wise results

TABLE 4 The numbers of pairs discordant and concordant for cancer at the end of follow-up for the smoking discordant pairs (one twin never smoker/the other former smoker) by pooled cancer sites and zygosity

	Monozygotic			Dizygotic		
	Number of pairs discordant for cancer		Number of cancer concordant pairs	Number of pairs discordant for cancer		Number of cancer concordant pairs
Pooled cancer sites	Cancer in the former-smoker twin	Cancer in the never-smoker twin	Cancer in both twins	Cancer in the former-smoker twin	Cancer in the never-smoker twin	Cancer in both twins
All tobacco-related[a] (including lung)	45	44	10	142	116	17
All tobacco-related[b] (excluding lung)	42	43	10	125	111	14
Causally tobacco-related[c] (including lung)	21	17	2	84	43	3
Causally tobacco-related[d] (excluding lung)	18	16	2	65	36	2

[a]Leukemia, colon, esophagus, kidney, larynx, liver, lung, nasal cavity/sinus, oral cavity, pancreas, pharynx, rectum, stomach, urinary bladder.
[b]Leukemia, colon, esophagus, kidney, larynx, liver, nasal cavity/sinus, oral cavity, pancreas, pharynx, rectum, stomach, urinary bladder.
[c]Esophagus, kidney, larynx, liver, lung, oral cavity, pancreas, pharynx, urinary bladder.
[d]Esophagus, kidney, larynx, liver, oral cavity, pancreas, pharynx, urinary bladder.

TABLE 5 The numbers of pairs discordant and concordant for cancer at the end of follow-up for the smoking discordant pairs (one twin never smoker/the other twin current smoker) by pooled cancer sites and zygosity

	Monozygotic			Dizygotic		
	Number of pairs discordant for cancer		Number of cancer concordant pairs	Number of pairs discordant for cancer		Number of cancer concordant pairs
Pooled cancer sites	Cancer in the current-smoker twin	Cancer in the never-smoker twin	Cancer in both twins	Cancer in the current-smoker twin	Cancer in the never-smoker twin	Cancer in both twins
All tobacco-related[a] (including lung)	130	90	15	401	243	29
All tobacco-related[b] (excluding lung)	105	91	12	292	231	23
Causally tobacco-related[c] (including lung)	69	33	5	252	96	4
Causally tobacco-related[d] (excluding lung)	42	33	3	139	80	2

[a]Leukemia, colon, esophagus, kidney, larynx, liver, lung, nasal cavity/sinus, oral cavity, pancreas, pharynx, rectum, stomach, urinary bladder.
[b]Leukemia, colon, esophagus, kidney, larynx, liver, nasal cavity/sinus, oral cavity, pancreas, pharynx, rectum, stomach, urinary bladder.
[c]Esophagus, kidney, larynx, liver, lung, oral cavity, pancreas, pharynx, urinary bladder.
[d]Esophagus, kidney, larynx, liver, oral cavity, pancreas, pharynx, urinary bladder.

TABLE 6 Associations of smoking status with pooled sites of first cancer diagnosis: Discordant pair analyses of all pairs, DZ pairs and MZ pairs

	All pairs				DZ pairs				MZ pairs			
	Former smokers vs never smokers		Current smokers vs never smokers		Former smokers vs never smokers		Current smokers vs never smokers		Former smokers vs never smokers		Current smokers vs never smokers	
Pooled cancer sites	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
All tobacco-related[a] (including lung)	1.27	1.09-1.48	1.71	1.50-1.96	1.22	1.01-1.46	1.75	1.50-2.05	1.39	1.03-1.86	1.61	1.25-2.09
All tobacco-related[b] (excluding lung)	1.20	1.02-1.41	1.35	1.17-1.56	1.15	0.94-1.40	1.36	1.15-1.62	1.32	0.97-1.80	1.33	1.02-1.75
Causally tobacco-related[c] (including lung)	1.79	1.42-2.26	3.09	2.50-3.82	1.72	1.31-2.25	3.23	2.54-4.13	1.96	1.22-3.16	2.71	1.76-4.18
Causally tobacco-related[d] (excluding lung)	1.60	1.23-2.10	2.16	1.69-2.76	1.57	1.12-2.14	2.29	1.72-3.04	1.69	1.00-2.87	1.85	1.15-2.98

[a]Leukemia, colon, esophagus, kidney, larynx, liver, lung, nasal cavity/sinus, oral cavity, pancreas, pharynx, rectum, stomach, urinary bladder.
[b]Leukemia, colon, esophagus, kidney, larynx, liver, nasal cavity/sinus, oral cavity, pancreas, pharynx, rectum, stomach, urinary bladder.
[c]Esophagus, kidney, larynx, liver, lung, oral cavity, pancreas, pharynx, urinary bladder.
[d]Esophagus, kidney, larynx, liver, oral cavity, pancreas, pharynx, urinary bladder.

for the occurrence of cancer at the nine (including lung) and eight (excluding lung) causally tobacco-related cancer sites, for which we found evidence consistent with a causal effect, that is, independent of familial factors. When we restricted the analysis to those eight cancer sites (excluding lung cancer) pooled together, the MZ discordant pair data provided evidence for an increased cancer risk among the twins with a smoking history compared to their never-smoking co-twins. The values were HR 1.85 (95% CI: 1.15-2.98) for the current smokers and HR 1.69 (95% CI: 1.00-2.87) for the former smokers. Results based on data from the DZ pairs were slightly higher in magnitude for current smokers (HR 2.29; 95% CI: 1.72-3.04), and slightly lower for former smokers (HR 1.57; 95% CI: 1.12-2.14), but not significantly different from the respective estimates for MZ pairs, as indicated by the overlapping confidence intervals.

4 | DISCUSSION

This large twin study investigated the associations of smoking with tobacco-related cancers while controlling for genetic and shared

environmental confounding. Our findings support a causal association between smoking and cancer at several sites. Many epidemiological studies have insufficient numbers of cancer cases to examine each cancer site separately. Thus, combining cases across cancer sites for which a causal association with smoking has been shown is a useful strategy to boost statistical power in order to detect significant associations. We excluded lung cancer from some of the analyses because its causal association with smoking has been established[14] and recently reported in studies from NorTwinCan,[9] Nordic population data[31] and Finnish population-based cohort data.[32] Therefore, we extended the causality analyses to the study of other cancer sites that are associated with smoking, but for which causality between smoking and cancer has not been deeply investigated with genetically informative twin data.

Since twin data allows for controlling for genetic and environmental influences shared by smoking and cancer, we consider the co-twin control design, especially the doubly discordant identical (MZ) twin design, to be among the most powerful approaches for testing causality of the association between smoking and cancer.[15] However, in some cases, we still did not have enough statistical power for discordant twin pair analysis because of the small number of cancer cases. A strength of our analysis was the use of *E*-values to supplement the observed HRs. Indeed, the *E*-values in the analyses of individuals were high, indicating that an unmeasured confounder would have to have a very strong association with both smoking and the cancer. It is unlikely that such confounders would not be known, and we could not identify such strong confounders from prior research. While our twin design excludes familial confounding, by itself it remains prone to uncontrolled nonfamilial confounding. In combination with our *E*-value analyses, we believe nonfamilial confounding is less likely to account for our results and the most parsimonious explanation is for a causal role of smoking.

In recent years, several additional designs have arisen to test hypotheses of causality and to triangulate the evidence for or against causality. The Mendelian randomization (MR) design can provide robust evidence when a strong genetic instrument is available. Typically, such an instrument is a risk score based on multiple variants known to be associated with smoking in large meta-analyses based on genome-wide association studies (GWAS). A recent large MR study of lung cancer showed that smoking was by far the strongest risk factor for lung cancer among multiple factors that were examined.[33] The investigators used only one genetic variant as the genetic instrument for smoking and thus did not make use of the currently available multiple variants known to be associated with smoking behavior.[34]

Gormley et al[35] used MR to show a causal effect of smoking on oral/oropharyngeal cancer. These findings are consistent with our observations for oral cavity and pharynx cancers in the within-pair analyses for all pairs and DZ pairs. The number of discordant MZ pairs was too small for these two sites. A large MR study of colorectal cancer[36] found associations of amount smoked with colorectal cancer, but no effect of smoking status (having ever smoked regularly), consistent with our findings for colon cancer and rectal cancer. When Larsson et al[37] used data from large cancer GWAS consortia and the UK Biobank with 367 643 adults to conduct MR analyses for smoking with multiple sites, they found evidence for causal associations for cancers of the lung, esophagus and cervix. Statistically nonsignificant associations of a similar magnitude were seen for head and neck cancer as well as for stomach cancer. Thus, twin and MR study designs provide convergent evidence for a causal role of smoking in multiple cancers.

Molecular genetic studies have identified genes known to be associated with smoking that are also associated with cancer. One early study of smoking and lung cancer showed that variation in a region of 15q25.1 containing nicotinic acetylcholine receptor genes contributed to lung cancer risk.[38] Since these genes were also associated with smoking, two possible explanations for the association arose. These could be genes that independently influence smoking behavior and lung development, that is, pleiotropic genes effects. Alternatively, as discussed by Hjelmborg et al,[9] smoking can act as a mediator when the identified genes increase smoking quantity, which in turn increases cancer risk. Since these genes do not increase lung cancer risk in nonsmokers,[39] the mediation mechanism is more likely. Hence, reducing smoking leads to less cancer risk. Needless to say, using the most powerful and robust designs to test causality, the randomized controlled trial (RCT) design would not be ethical in investigating the causal association of smoking with cancer. Examination of simultaneous trends in smoking and cancer incidence provide additional evidence as well. The increase and subsequent decline in smoking in the United States and many other countries has been mirrored in a similar development of lung cancer with a delay of a few decades, further supporting a causal association. This observed linkage of temporal trends occurs because smoking is by far the main risk factor for lung cancer. When the relative risk is smaller, temporal trends may be obscured by changes in other risk factors of the cancer in question.

Due to low numbers of cancer cases for most individual cancer sites in our data, we pooled eight cancer sites together to enable a more powerful statistical analysis using the discordant twin pair design. In designing the pooled analyses, we first identified the set of total tobacco-related cancers based on prior observational studies. These cancers can be viewed as a mix of causally related and noncausally related cancers, the latter being due to known and unknown confounders. We then did a post hoc analysis by narrowing down the set of cancers to those eight sites where the within-pair risk estimates were elevated and consistent with the between pair estimates. Since this was performed post hoc, ideally our results would need to be replicated in other data sets. However, currently, we are not aware of any twin cohort of sufficient size with data on smoking and cancer incidence.

We also estimated cancer risk for women compared to men while adjusting for smoking status and country cohort effect. Our results are consistent with the well-established findings showing that women have a lower risk of cancer across many cancer sites,[5] which may suggest sex interactions with environmental exposure or other sex-specific modifiers or mediators such as hormones or anthropometrics.[40] Our finding that sex differences exist across cancer sites even after adjusting for smoking is congruent with earlier evidence.[5,41,42]

However, we acknowledge that only adjusting for smoking status may not be enough, as cumulative exposure to smoking, such as "pack years," may explain more of the difference between men and women. Unfortunately, we did not have such detailed smoking data available for all country cohorts in this large multicountry dataset.

Strengths of our study include its large sample size and prospective, population-based and multicountry design with virtually complete long-term follow-up with smoking data collected before cancer diagnosis. There is evidence that the cancer incidence among twins in NorTwinCan reflects that in the background population,[43] which greatly supports the generalizability of our findings beyond twin populations. Even if there were minor differences between twins and singletons in the incidence of cancer and in the prevalence of risk factors such as smoking, the strength of the association between risk factors and outcome in twins and singletons is unlikely to be consequential.

Among the study limitations, we acknowledge that smoking status was self-reported and did not systematically provide data on the amount and duration of smoking (eg, pack-years) for current and former smokers. While this most likely did not affect our main conclusions, more detailed data would have permitted analyses to test dose-response relationships between smoking and cancer. Another weakness is a potential misclassification of smoking status because we used the status on first assessment, and it was not repeatedly updated. Thus, if individuals changed their smoking status during the follow-up this was not captured in our data. However, nearly all surveys were conducted on adults and most initiation of smoking takes place during adolescence. Given these secular trends in smoking, it is probably more likely that some current smokers had quit rather than never smokers initiating smoking or former smokers relapsing back to smoking during follow-up. Hence, some individuals classified as current smokers might indeed have become former smokers which would lead to underestimation of the differences in risk between these two categories. Further, although response rates have been high in these cohorts, it is well recognized that in general smokers are less likely to participate in health-related surveys compared to nonsmokers. Additionally, surveys on smoking habits were conducted at different times in the different cohorts.

Information on the histology of cancers was not available, therefore, we could not consider histological subtypes of any cancer diagnosis. Thus, we could not specify the histology of, for example, esophageal squamous cell cancers, which are known to be causally tobacco-related, while adenocarcinoma is less associated with smoking.[44] Another limitation is that the dataset did not allow for multivariable analyses adjusted for potential confounders such as alcohol use and other lifestyle factors or sociodemographic factors. Finally, these cohorts recorded the onset of cancer over a long time-period, but we do not have clinical information, including changes in diagnostic tools or treatment in the dataset.

5 | CONCLUSIONS

Findings from the individual-based analyses showing significant associations between current smoking and several types of cancer were replicated through the doubly discordant genetically identical twin pairs. By pooling the specific cancer sites together, we optimized the power for detecting causal associations. In other words, smoking seems to affect the risk of multiple cancers independent of shared familial influences. Given that tobacco contains known carcinogens, the twin analyses and *E*-value analyses provide very strong convergent evidence to support a causal role of smoking in the cancers we have studied here. Thus, the causative role of cigarette smoking in cancer is not limited to lung cancer but appears to extend to several other cancer types. Regarding clinical implications, there should be even more emphasis on recording and acting on smoking status in screening activities. Naturally, smoking cessation should be strongly encouraged in the prevention of all types of cancer.

ACKNOWLEDGEMENTS

This work was supported by funding from the Nordic Union of Cancer (PI J. Kaprio). The Finnish Twin Cohort was supported by the Academy of Finland (grant numbers 213506, 129680, 265240, 263278 & 312073) and the Sigrid Juselius Foundation (to J Kaprio). The Ministry for Higher Education financially supports the Swedish Twin Registry. The Danish Twin Cohort was supported by the Odense University Hospital AgeCare program of the Academy of Geriatric Cancer Research. Jennifer R Harris and Jacob Hjelmborg have received funding (2017-2019) by grants from the Nordic Cancer Union, project entitled "Genetic Epidemiology and Familial Risk of Cross-Cancer Associations: A Nordic Twin Study." We acknowledge the contribution of all the researchers and staff of the twin and cancer registries contributing to the NorTwinCan project. We thank Alyce Whipp from University of Helsinki Language Services for language editing.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Tellervo Korhonen: Conceptualization, investigation, writing-original draft, writing—review & editing. **Jacob Hjelmborg:** Conceptualization, data curation, formal analysis, investigation, methodology, resources, software, validation, visualization, writing—review & editing. **Jennifer R. Harris:** Conceptualization, funding acquisition, resources, writing—review & editing. **Signe Clemmensen:** Data curation, writing—review and editing. **Hans-Olov Adami:** Conceptualization, funding acquisition, resources, writing—review & editing. **Jaakko Kaprio:** Conceptualization, data curation, funding acquisition, investigation, project administration, resources, supervision, validation, visualization, writing—review & editing.

DATA AVAILABILITY STATEMENT

The data for the present analysis were compiled with the agreement of the participating twin cohorts and national cancer registries. Requests to access additional data should be directed to the individual national cohorts and registers who are responsible for the datasets (http://nortwincan.org/). Further information is available from the corresponding author upon request.

ETHICS STATEMENT

The study was approved by the ethical committees in each country. The participants provided their informed consent by returning the baseline questionnaire.

ORCID

Tellervo Korhonen https://orcid.org/0000-0003-2838-3085

TWITTER

Tellervo Korhonen @TellervoKorhone

REFERENCES

1. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392:1736-1788.
2. GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392:1923-1994.
3. Jemal A, Vineis P, Bray F, Torre L, Forman D, eds. *The Cancer Atlas*. 2nd ed. Atlanta, GA: American Cancer Society; 2014. https://publications.iarc.fr/Non-Series-Publications/Other-Non-Series-Publications/The-Cancer-Atlas-2nd-Edition-2014
4. Dorak MT, Karpuzoglu E. Gender differences in cancer susceptibility: an inadequately addressed issue. *Front Genet*. 2012;3:268.
5. Edgren G, Liang L, Adami HO, Chang ET. Enigmatic sex disparities in cancer incidence. *Eur J Epidemiol*. 2012;27:187-196.
6. Hjelmborg JB, Scheike T, Holst K, et al. The heritability of prostate cancer in the Nordic Twin Study of Cancer. *Cancer Epidemiol Biomarkers Prev*. 2014;23:2303-2310.
7. Harris JR, Hjelmborg J, Adami HO, et al. The Nordic Twin Study on Cancer—NorTwinCan. *Twin Res Hum Genet*. 2019;22:817-823.
8. Mucci LA, Hjelmborg JB, Harris JR, et al. Familial risk and heritability of cancer among twins in Nordic countries. *JAMA*. 2016;315:68-76.
9. Hjelmborg J, Korhonen T, Holst K, et al. Lung cancer, genetic predisposition and smoking: the Nordic Twin Study of Cancer. *Thorax*. 2017;72:1021-1027.
10. Office of the Surgeon General (US); Office on Smoking and Health (US). *The Health Consequences of Smoking: A Report of the Surgeon General*. Atlanta (GA): Centers for Disease Control and Prevention (US); 2004. PMID: 20669512.
11. Cogliano VJ, Baan R, Straif K, et al. Preventable exposures associated with human cancers. *J Natl Cancer Inst*. 2011;103:1827-1839.
12. Ezzati M, Henley SJ, Lopez AD, Thun MJ. Role of smoking in global and regional cancer epidemiology: current patterns and data needs. *Int J Cancer*. 2005;116:963-971.
13. Vineis P, Alavanja M, Buffler P, et al. Tobacco and cancer: Recent epidemiological evidence. *J Natl Cancer Inst*. 2004;96:99-106.
14. National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General*. Atlanta (GA): Centers for Disease Control and Prevention (US); 2014. PMID: 24455788.
15. McGue M, Osler M, Christensen K. Causal inference and observational research: the utility of twins. *Perspect Psychol Sci*. 2010;5:546-556.
16. Wienke A, Herskind AM, Christensen K, Skytthe A, Yashin AI. The heritability of CHD mortality in Danish twins after controlling for smoking and BMI. *Twin Res Hum Genet*. 2005;8:53-59.
17. Osler M, McGue M, Christensen K. Socioeconomic position and twins' health: a life-course analysis of 1266 pairs of middle-aged Danish twins. *Int J Epidemiol*. 2007;36:77-83.
18. Johnson W, Kyvik KO, Mortensen EL, Skytthe A, Batty GD, Deary IJ. Does education confer a culture of healthy behavior? Smoking and drinking patterns in Danish twins. *Am J Epidemiol*. 2011;173:55-63.
19. Kaprio J, Koskenvuo M. A prospective study of psychological and socioeconomic characteristics, health behavior and morbidity in cigarette smokers prior to quitting compared to persistent smokers and non-smokers. *J Clin Epidemiol*. 1988;41:139-150.
20. Kaprio J, Koskenvuo M. Genetic and environmental factors in complex diseases: the older Finnish Twin Cohort. *Twin Res*. 2002;5:358-365.
21. Kaprio J, Bollepalli S, Buchwald J, et al. The older Finnish Twin Cohort: 45 years of follow-up. *Twin Res Hum Genet*. 2019;22:240-254.
22. Nilsen TS, Brandt I, Magnus P, Harris JR. The Norwegian Twin Registry. *Twin Res Hum Genet*. 2012;15:775-780.
23. Harris JR, Magnus P, Tambs K. The Norwegian Institute of Public Health twin program of research: an update. *Twin Res Hum Genet*. 2006;9:858-864.
24. Lichtenstein P, De Faire U, Floderus B, Svartengren M, Svedberg P, Pedersen NL. The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. *J Intern Med*. 2002;252:184-205.
25. Pedersen NL, Lichtenstein P, Svedberg P. The Swedish Twin Registry in the third millennium. *Twin Res*. 2002;5:427-432.
26. Scheike TH, Holst KK, Hjelmborg JB. Estimating heritability for cause-specific mortality based on twin studies. *Lifetime Data Anal*. 2014;20:210-233.
27. Gerster M, Madsen M, Andersen PK. Matched survival data in a co-twin control design. *Lifetime Data Anal*. 2014;20:38-50.
28. Nicholas PJ. *Statistics for Epidemiology*. Boca Raton: Chapman & Hall/CRC, 2004, pp. 352, ISBN: 1-58488-433-9.
29. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med*. 2017;167:268-274.
30. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2019. https://www.R-project.org/
31. Hemminki K, Forsti A, Hemminki A, Ljungberg B, Hemminki O. Incidence trends in lung and bladder cancers in the Nordic countries before and after the smoking epidemic. *Eur J Cancer Prev*. 2021. https://doi.org/10.1097/CEJ.0000000000000694. Online ahead of print.
32. Pitkäniemi J, Heikkinen S, Seppä K, et al. Pooling of Finnish population-based health studies: lifestyle risk factors of colorectal and lung cancer. *Acta Oncol*. 2020;59:1338-1342. https://doi.org/10.1080/0284186X.2020.1789214.
33. Shen J, Zhou H, Liu J, et al. A modifiable risk factors atlas of lung cancer: a Mendelian randomization study. *Cancer Med*. 2021;10:4587-4603.
34. Liu M, Jiang Y, Wedow R, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet*. 2019;51:237-244.
35. Gormley M, Dudding T, Sanderson E, et al. A multivariable Mendelian randomization analysis investigating smoking and alcohol consumption in oral and oropharyngeal cancer. *Nat Commun*. 2020;11:6071. https://doi.org/10.1038/s41467-020-19822-6.
36. Dimou N, Yarmolinsky J, Bouras E, et al. Causal effects of lifetime smoking on breast and colorectal cancer risk: Mendelian randomization study. *Cancer Epidemiol Biomarkers Prev*. 2021;30:953-964.
37. Larsson SC, Carter P, Kar S, et al. Smoking, alcohol consumption, and cancer: a mendelian randomisation study in UK Biobank and international genetic consortia participants. *PLoS Med*. 2020;17:e1003178. doi:10.1371/journal.pmed.1003178
38. Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet*. 2008;40:616-622.

39. McKay JD, Hung RJ, Han Y, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet*. 2017; 49:1126-1132.
40. Fu BC, Song M, Li X, et al. Height as a mediator of sex differences in cancer risk. *Ann Oncol*. 2020;31:634-640.
41. Zhu Y, Shao X, Wang X, Liu L, Liang H. Sex disparities in cancer. *Cancer Lett*. 2019;466:35-38.
42. Kim HI, Lim H, Moon A. Sex differences in cancer: epidemiology, genetics and therapy. *Biomol Ther (Seoul)*. 2018;26:335-342.
43. Skytthe A, Harris JR, Czene K, et al. Cancer incidence and mortality in 260,000 Nordic twins with 30,000 prospective cancers. *Twin Res Hum Genet*. 2019;22:99-107.
44. Abnet C, Nyrén O, Adami H. Esophageal cancer. In: Adami H, Hunter D, Lagiou P, Mucci L, eds. *Textbook of Cancer Epidemiology*. 3rd ed. New York: Oxford University Press; 2018:183-212. https://oxford.universitypressscholarship.com/view/10.1093/oso/9780190676827.001.0001/oso-9780190676827-chapter-9

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Korhonen T, Hjelmborg J, Harris JR, et al. Cancer in twin pairs discordant for smoking: The Nordic Twin Study of Cancer. *Int. J. Cancer*. 2022;151(1):33-43. doi:10.1002/ijc.33963