Lung cancer, genetic predisposition and smoking: the Nordic Twin Study of Cancer

Jacob Hjelmborg*, Ph.D., Tellervo Korhonen*, Ph.D., Klaus Holst, Ph.D., Axel Skytthe, Ph.D., Eero Pukkala, Ph.D., Julia Kutschke, Ph.D., Jennifer R. Harris, Ph.D., Lorelei A. Mucci, Sc.D., Kaare Christensen, M.D., Ph.D., Kamila Czene, Ph.D., Hans-Olov Adami, M.D., Ph.D., Thomas Scheike, Ph.D., and Jaakko Kaprio, M.D, Ph.D., *on behalf of the Nordic Twin Study of Cancer (NorTwinCan) collaboration.*

*Drs. Hjelmborg and Korhonen contributed equally to this article.

Author affiliations:

Department of Epidemiology, Biostatistics and Biodemography, University of Southern Denmark, Denmark (J.H., A.S., K.C.); The Danish Twin Registry, University of Southern Denmark, Denmark (J.H., A.S., K.C.); Department of Biostatistics, University of Copenhagen, Denmark (K.H., T.S.); Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston MA; (L.A.M); Centre for Public Health Sciences, University of Iceland, Reykjavik, Iceland (L.A.M); Finnish Cancer Registry, Institute for Statistical and Epidemiological Cancer Research, Helsinki, Finland (E.P): School of Health Sciences, University of Tampere, Tampere, Finland (E.P.); Division of Epidemiology, The Norwegian Institute of Public Health, Oslo, Norway (J.Ku., J.R.H.); Department of Public Health, University of Helsinki, Finland and Department of Health, National Institute for Health & Welfare, Helsinki, Finland (J.Ka., T.K); Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden (K.Cz., H-O.A.); and Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA (L.A.M., H-O.A); Institute for Molecular Medicine, University of Helsinki, Finland (J.Ka.); Institute of Public Health and Clinical Nutrition, University of Eastern Finland (T.K.)

Corresponding author:

Jaakko Kaprio, Department of Public Health, PO Box 41, FI-00014 University of Helsinki, Finland; Phone: +358 50 3715419; Email: Jaakko.kaprio@helsinki.fi

Key messages

What is the key question?

Is there a significant genetic component to the occurrence of lung cancer and is the genetic influence modified by smoking and age?

What is the bottom line?

The interplay between genes and tobacco smoking in the etiology of lung cancer has remained controversial, and we disentangle genetic and environmental causes in cancer while taking smoking status into account.

Why read on?

Our study shows that tobacco exposure causes lung cancer even when adjusting for genetic factors. Interactions between genes and environmental exposure in the development of lung cancer are not supported from the largest twin cohort study with longest follow-up ever. Familial effects have decreased influence with increasing age.

Word count: Abstract 250; text 3336

Abstract

Background

We aimed to disentangle genetic and environmental causes in lung cancer while considering smoking status.

Methods

Four Nordic Twin Cohorts (43,512 monozygotic (MZ) and 71,895 same sex dizygotic (DZ) twin individuals) had smoking data before cancer diagnosis. We used time-to-event analyses accounting for censoring and competing risk of death to estimate incidence, concordance risk and heritability of liability to develop lung cancer by smoking status.

Results

During a median of 28.5 years of follow-up we recorded 1,508 incident lung cancers. Of the 30 MZ and 28 DZ pairs concordant for lung cancer, nearly all were current smokers at baseline and only one concordant pair was seen among never smokers. Among ever smokers the case-wise concordance of lung cancer, that is the risk before a certain age conditional on lung cancer in the co-twin before that age was significantly increased compared with the cumulative incidence for both MZ and DZ pairs. This ratio, the relative recurrence risk, significantly decreased by age for MZ, but was constant for DZ pairs. Heritability of lung cancer was 0.41 (95%CI 0.26–0.56) for currently smoking and 0.37 (95%CI 0.25–0.49) for ever smoking pairs. Among smoking discordant pairs, the pairwise hazard ratio for lung cancer of the ever smoker twin compared to the never smoker cot-win was 5.4 (95%CI 2.1–14.0) in MZ pairs and 5.0 (95%CI 3.2–7.9) in DZ pairs.

Conclusions

The contribution of familial effects appears to decrease by age. The discordant pair analysis confirms that smoking causes lung cancer.

Introduction

Smoking is the primary cause of lung cancer globally, though several other environmental exposures play a role.[1] The estimated heritable genetic contribution to variation in risk to lung cancer overall has been modest in family (heritability estimate of 0.08)[2] and twin (0.26[3] and 0.18[4]) studies. Genome-wide association (GWA) studies further suggest that some gene loci are associated with lung cancer in both smokers and non-smokers, while other variants, such as the functional D398N (rs16969968) variant in CHRNA5, are associated with lung cancer only among smokers.[5,6] Thus, the heritability of lung cancer may vary as a function of smoking, but the differential effect of smoking on genetic variation underlying development of lung cancer has not been quantified.

To this end, our aim is to estimate the heritability of liability to lung cancer based on the largest twin cohort to date, the Nordic Twin Study of Cancer (NorTwinCan)[4], which extends the Lichtenstein (2000)[3] study with longer follow-up and new birth cohorts and refined methodology. We sought to estimate the heritability in the liability to lung cancer and whether it is modified by smoking or age.

Methods

Material

NorTwinCan includes population-based cohorts from the Danish, Finnish, Norwegian, and Swedish twin registries.[7] Each twin has an individually unique national registration number, allowing for linkage to the national cancer and mortality registries with complete follow-up, drop-out being only due to death or emigration. Lung cancer occurrence was obtained from the national cancer registries and computed from the baseline when smoking status was determined until the end of follow-up (Table 1). In all cohorts, zygosity - monozygotic (MZ) or dizygotic (DZ) - was determined at baseline by validated questionnaire methodology, which classifies more than 95% of twin pairs correctly.[3] Twins, who have not replied to the questionnaires, as well as a minority providing inconsistent responses, are classified as unknown zygosity (UZ). The ethics committees for each country approved the study.

Given the major role of smoking in the etiology of lung cancer, our analysis includes twin individuals of known zygosity from the Danish, Finnish, Norwegian, and Swedish registries, where data on smoking status was available prior to lung cancer diagnosis. We excluded individuals from opposite-sex DZ pairs as data from them have not been as comprehensively collected. For individuals who reported smoking behavior on more than one questionnaire, we used the earlier information.

Characteristics of the four national twin cohorts included in the analyses are summarized in Table 1. We classified the participants as never smokers, ever smokers (former or current at time of questionnaire) and current smokers based on the survey items used to assess smoking status. Smoking data in the Danish cohort came from the eight questionnaire surveys conducted from 1959 to 2002.[8–10] In Finland smoking data came primarily from the first questionnaire survey in 1975, but some twins who had not replied in 1975 responded to a questionnaire survey in 1981.[11,12] In the Norwegian cohort smoking data came from three questionnaire surveys in 1980–1982 & 1990–92 & 1998.[13,14] In the Swedish cohort smoking data came from questionnaire surveys in 1961, 1967, 1970, and 1973.[15,16]

We included individuals with histologically confirmed lung cancer. Among those with smoking data, we recorded a total of 1,508 incident lung cancers with a mean follow-up time of 25.2 years (21.0 years in lung cancer patients).

Statistical analysis

After defining cohort-specific dates of entry and follow-up, we accounted for left-truncation from variable initiation of cancer registration and right-censoring among those censored at the end of follow-up, and lost to follow-up due to emigration (<2%). We examined the individual risk of lung cancer diagnosis by age by estimating cumulative lung cancer, incidence[17] and lifetime risk as the cumulative incidence (the probability of lung cancer) by age 80 years. We modeled potential competing deaths[18,19] which allows estimation of lung cancer risk in a twin given the occurrence of other disease in his/her co-twin. We obtained the case-wise concordances by age[18,19] (see supplementary material for details) as well as relative recurrence risks in MZ and DZ pairs and the multilocus index.[20,21]

We extended standard biometrical modelling methods to address issues of censoring at follow-up[7,22] . Results would agree with those obtained from standard models for twin data[18,23,24] if no censoring were present. Quantitative models were analyzed to estimate the magnitude of variation explained by genetic and environmental influences[18] underlying the liability to develop lung cancer by smoking status. The relative magnitude of genetic influences on variation in liability to lung cancer is thus estimated among pairs in which neither had ever smoked, among pairs where both co-twins are ever (former or current) smokers and among pairs in which both co-twins are current smokers.

We use information on lung cancer incidence in MZ and DZ pairs to decompose variation into additive genetic effects (A), dominant genetic effects (which represent deviations of the heterozygote genotype from the mean of the homozygote genotype) (D), common environmental effects (C), and individually unique environmental effects (E). Within-pair covariance of liability is expressed as κ var(A) + γ var(D) +var(C), where $\kappa = \gamma = 1$ for MZ pairs and $\kappa = 1/2$ and $\gamma = 1/4$ for DZ pairs.[18] We tested a series of models sequentially to assess the significance of specific parameters. We estimated measurement error in E which is the component of variance that does not contribute to within-pair resemblance. Dominance effects are, typically, biologically implausible in the absence of additive effects. The primary models are thus the ACE and ADE models, as well as their sub-models AE, CE, and E. We assessed the fit of the sub-models by the Akaike information criterion[22].

We tested for equal thresholds (i.e., normal quantiles of prevalence) between MZ and DZ twins, which is equivalent to assuming that the risk of disease does not differ by zygosity. We tested for constant relative recurrence risk (RRR) over age by grouping into five-year interval from age 65 to 90 years of age for MZ and DZ pairs. To correct for possible bias due to censoring, individuals were assigned weights obtained by calculating the inverse probability of being censored at time of follow-up[7,18,19,22] Estimates have not been adjusted for the effect of left-truncation that would cause an upwards bias, which is not yet feasible for the approach.

For gene and smoking status interaction the magnitude on liability scale could not be estimated due to having one concordant pair among all never-never and never-ever smoking pairs. The presence of genetic interaction with smoking status was therefore investigated by comparing observed concordance in strata of smoking status to the expected when assuming same variance components on the liability scale as in ever-ever pairs but using smoking-status specific cumulative incidence by age as well as follow-up time of the specific pairs in the cohort. This procedure leads to an approximate test, which we later refer to as the binomial test, and takes into account the smoking-status specific cumulative incidence by age, as well as follow-up time of the specific pairs in the cohort and we then computed the probability that a randomly selected pairs were concordant using the dependence parameters of the liability threshold model for the ever-ever pairs.

Among pairs in which one twin was a smoker and the other was not, we computed within pair hazard ratios for the association of smoking with lung cancer using a Cox model with pair-specific baseline hazard functions. Given that MZ pairs share their genomic sequence, an association of smoking with lung cancer risk within such pairs is independent of genetic liability. This hypothesis has historically competed with the hypothesis[25] of shared genes underlying both smoking and lung cancer. The statistical program R was used for all analyses with the package *mets*.[26]

Results

Among those with smoking data, we recorded 1,508 incident lung cancers among a total of 115,407 twin (43,512 MZ and 71,895 DZ) individuals. Forty-seven percent were never smokers (n=54238), 16% former smokers (n=18,231) and 37% current smokers (n=42,938) at baseline. Figure 1 shows the cumulative incidence of lung cancer by smoking status (never, former, current) and sex. The risk of lung cancer diagnosis before 80 years of age is estimated at 0.6% (95% CI 0.5%–0.7%) among never smokers, 2.0% (1.7%–2.3%) among former and 5.7% (5.4%–6.0%) among current smokers adjusting for censoring and competing risk of death. The only sex difference is seen among smokers. There was no difference in risk between MZ and DZ twin individuals.

The numbers of pairs concordant and discordant for lung cancer incidence are presented in Table 2 for those with smoking data (n=50,595 pairs with smoking status on both twins) overall and further classified by smoking status.

Among twin pairs where both are ever smokers, the risk of lung cancer in a twin before a given age given that his or her co-twin also has lung cancer before that age, the case-wise concordance by age is depicted in Figure 2 in both MZ and DZ pairs, as well as the cumulative incidence of lung cancer by age in individuals. The case-wise concordance risk was larger in MZ twins than the individual cumulative incidence risk, testing for a difference from the cumulative incidence across the five year age intervals (chisq=22.1, df=6, p=0.001). For the DZ twins we found that the case-wise concordances were borderline significantly different from the cumulative incidence (chisq=13.4, df=6, p=0.04). The estimated case-wise concordance at 90 years of age was 0.20 (0.13-0.27) for MZ pairs and 0.13 (0.08-0.17) for DZ pairs.

This excess risk of MZ and DZ pairs of the case-wise concordance relative to the population based individual cumulative incidence of lung cancer, the relative recurrence risk (also known as the lambda value) is depicted in Figure 3 and demonstrates the presence of familial effects at all ages. The RRR is higher at younger ages, in fact the lung cancer risk is increased 10.2 -fold (3.2-17.2) at 65 years of age and decreases significantly to a 3.6 (2.3-4.9) -fold increase at 90 years of age if a MZ co-twin is diagnosed (p-value = 0.04, test for trend). The RRR is suggested to be constant by age for DZ twins (p-value = 0.25, test for trend) (Figure 3). (A table of relative risks by age-group is provided in supplemental Table 1.) We tested if the absolute differences of the MZ and DZ curves at each five-year interval from age 65 to age 90 years of age were significantly different, which there was no sign of (p-value=0.21). Our results are thus consistent with the hypothesis of rather strong familial influences that do not increase across age. We hypothesize that the genetic part of the familial influence may become weaker by age.

We then examined evidence for genetic factors in the liability to develop lung cancer by smoking status. Among pairs in which neither had ever smoked (7,871 MZ pairs and 10,768 DZ pairs), there was one lung cancer concordant MZ pair with 43 MZ and 59 DZ lung cancer discordant pairs. Heritability could not be estimated. However, the dependence in the never-never and never-ever pairs was not significantly different from the dependence among the ever-ever pairs (p=0.28, binomial test of observing more than one concordant pair of lung cancer).

The overall estimate of familial aggregation (genetic variance and shared environment component) for lung cancer liability is 44% with 38% (0.05- 0.72) of variability attributed to genetic effects. When adjusted for smoking status, effects of country and sex, variability attributed to genetic effects was 34% (0.00-0.70) (Table 3). A comparison of the MZ and DZ tetrachoric within-pair correlations in liability to develop lung cancer (Table 3) adjusting for age, sex, country and smoking, and further adjustment for censoring hypothesizing equal correlations, gave a p-value of 0.07 (Wald test). Among the pairs where both twins are ever (current or former) smokers, the heritability estimates ranged from 28% (0.00-0.66) to 37% (0.25-0.49), depending on the assumptions of the genetic model (Table 4). A pure environmental model did not fit the data. Among current smokers, the heritability was estimated at 29% (0.00-0.74) or 41% (0.26-0.56), depending on genetic assumptions (Table 4).

Finally, for smoking discordant pairs, we examined whether smoking status was associated with future lung cancer. In the ever smoking discordant pairs (3,274 MZ pairs and 8,350 DZ pairs), 40 MZ pairs were discordant for lung cancer (Table 5). Of these 35 cases were among ever smokers (with their non-smoking co-twin being unaffected) and only five in the never-smokers (while their smoking co-twin was unaffected), yielding a paired analysis hazard ratio (HR) of 5.4. Results for DZ pairs and for current-smoking *versus* never smoking discordant pairs are shown in Table 5. Most discordant pairs arose from pairs in which the smoker still smoked at baseline. None of the smoking discordant pairs were concordant for lung cancer.

Discussion

In the largest study of lung cancer in twins to date, we found that genetic effects account for a significant amount of the variation in the liability to develop lung cancer, and the magnitude of this estimate is independent of smoking status. The largest estimate of heritability in the liability to lung cancer was found in pairs where both were current smokers at baseline. Among twin pairs where both twins were never smokers, only one concordant lung cancer pair was seen and a formal estimate of heritability could not be derived. A test of gene by smoking interaction was not significant suggesting that the relative contribution of genetics does not vary by smoking status. Furthermore, testing suggests that the contribution of familial effects does not increase by age. Our pairwise analysis of smoking discordant pairs confirmed that smoking causes lung cancer independent of genetic liability either to smoking or to lung cancer.

Twin pairs discordant for both lung cancer and smoking status at baseline are informative for causal analyses. In the lung cancer and smoking doubly discordant pairs, the pairwise relative risk for lung cancer was 5.4 among ever smokers in MZ pairs. It is of historical interest that after the landmark papers of Doll and Hill[27] and Wynder and Graham[28] in the early 1950s, the causality of the relationship between smoking and lung cancer was soon challenged by the great statistician Ronald Fisher.[25] He pointed out the greater similarity of MZ vs. DZ pairs for smoking, and indicated genetics as a potential confounder. MZ pairs discordant for smoking would help to resolve the issue of causality. Following up on prior twin studies of smoking discordant pairs,[29,30] we can now finally put this issue to rest, an issue debated for many years because of tobacco industry's prolonged refusal to acknowledge publicly that smoking causes lung cancer.

Smoking is the most important cause of lung cancer. Taking smoking into account permits us to test for the dependence of genetic effects on smoking status. The overall estimate of familial aggregation (genetic variance and shared environment component) for lung cancer liability is 44%, with most variability attributed to genetic effects (38%), higher but still consistent with the estimate 26% (95%CI 0%–49%) by Lichtenstein et al.[3] also unadjusted for smoking and for censoring, but based on a smaller number of affected pairs. We recently reported on the heritability for liability to lung cancer in the entire NorTwinCan data, with an overall estimate of familial aggregation of 42%.[4] The present analysis extends these estimates by accounting for the effect of smoking status prior to disease occurrence and examines heritability among the smoking pairs.

In our analysis, adjustment for smoking eliminates the estimates for shared environmental effects. Shared environmental effects (i.e. exposure to smokers in the childhood home, and among peers in adolescence) are of importance for the initiation of smoking[31] so it is not surprising that adjustment for smoking controls for this source of variation. The highest estimates of heritability and recurrence risks were seen among current smoking pairs. Among never smokers, we cannot estimate the heritability of lung cancer.

Prior family[2] and twin[3,4] studies of lung cancer have demonstrated familial aggregation and provided very modest estimates for the role of genes. The Swedish multi-generational register family study[2] estimated the heritability of lung cancer to be 8% (95% CI 5%–9%), without information on smoking in the families. The American World War II veterans' study [32] followed 12,938 male twin pairs for 44 years for mortality. Among pairs with at least one lung cancer death, only 10 of 269 MZ pairs and 21 of 373 DZ pairs were concordant, and no heritability estimate was provided. Smoking information was not used in the analysis, but smoking-related cancers showed less MZ – DZ differences in similarity than other cancers. Despite the large number of pairs in our present study, the final number of concordant pairs with smoking information was limited. Thus, we could not examine heritability of lung cancer risk in relation to time trends in lung cancer or histological subtypes of lung cancer. Nor did we have information on smoking amount, duration or changes in smoking status comprehensively and comparably assessed in all the twin cohorts.

Since detailed smoking information was not available, it should be acknowledged as a potential limitation that there might be residual confounding that remains in the estimates of heritability estimation. Because MZ twins, who are smokers, are also more similar than DZ pairs in age of smoking initiation, amount smoked and duration of smoking[31], the heritability of lung cancer among smokers may still contain residuals effects of genetics on smoking, and thus on lung cancer risk.

The overall genetic contribution to lung cancer as a function of smoking status is relevant for gene discovery. Since 2007, 21 lung cancer genome-wide analysis (GWA) and genome-wide meta-analysis studies[33] (www.genome.gov/gwastudies) have found the strongest association to the CHRNA5 functional D398N (rs16969968) variant. The functional changes[34,35] in nicotinic acetylcholine receptor activity are linked to increased risk for nicotine dependence, higher amount smoked[36-39] and higher cotinine levels.[40,41] Thus, those with a risk allele smoke more, are more tobacco-dependent and are less likely to quit, and therefore at higher risk of developing lung cancer. However, D398N is not a risk factor for lung cancer in non-smokers, based on a GWA meta-analysis of 14,900 lung cancer cases and 29,485 controls[6] and among 56,037 individuals from the HUNT population study in Norway.[5] This variant requires exposure to smoking to affect lung cancer risk and thus contributes to the heritability seen among current smokers. In contrast to D398N, associations with other loci found to be significant for lung cancer such as those in 5p15 (TERT and CLPTM1L genes) and 6p21 (BAG6/BAT3) are found also in non-smokers.[33,6] The existence of a modest familial liability to lung cancer independent of smoking status was also observed in the analysis of Utah genealogical data.[42] An increased risk of lung cancer was seen even in distant relatives; the high proportion of non-smoking lung cancer cases (31%) and a large proportion of missing data on smoking status (which was assessed through the death certificate and not prospectively) calls for replication in other populations. A recent large meta-analysis yielded an array-based heritability estimate for lung cancer of 21% (95% CI 14-27%).[43] This is somewhat smaller than our overall twin estimates suggesting that much of the genetic liability to lung cancer is attributable to common variants, but other genetic effects may exist. The same study estimated that 24% of the heritability of lung cancer is accounted for by genetic determinants of smoking behavior.

In conclusion, our study extends earlier studies to examine the heritability in liability to lung cancer by smoking status and age. We find no formal evidence for a gene by environmental exposure interaction in lung cancer; more detailed environmental exposures and larger sample sizes may be required. We hypothesize that a genetic part of the rather strong familial influence demonstrated may become weaker by age. Studies of genetic factors and hence molecular mechanisms in cancer would benefit by carefully taking into account known environmental risk factors and identifying the population groups at highest genetic risk using environmental stratification. However, the discordant pair analysis conclusively demonstrates that tobacco exposure causes lung cancer even when adjusting for genetic factors.

Contributions

Jacob Hjelmborg (J.H.) designed the study, contributed to developing the statistical methodology, conducted the data analysis, interpreted the data, and wrote the methods section of the manuscript.

Tellervo Korhonen (T.K.) contributed to the design and wrote the manuscript together with J.H. and J.K.

(Drs. Hjelmborg and Korhonen contributed equally to this article.)

Klaus Holst made central contributions to developing the statistical methodology, took part in conducting the statistical analysis as well as in revising the manuscript.

Axel Skytthe was responsible for quality assurance of the combined data set, conducted the data analysis, reviewed and commented the manuscript.

Eero Pukkala contributed to quality assurance of the combined data set reviewed, commented and edited the manuscript.

Julia Kutschke (nee Isaeva) helped to prepare the Norwegian data.

Jennifer R. Harris helped in the drafting and providing critical comments to manuscript.

Lorelei A. Mucci reviewed, commented and edited the manuscript.

Kaare Christensen reviewed, commented and edited the manuscript.

Hans-Olov Adami was involved in initiating, designing and funding the study as well as in interpreting the results and editing the manuscript.

Thomas Scheike contributed to statistics and took part in revising the manuscript.

Jaakko Kaprio (J.K.) designed the study, contributed to data interpretation, and wrote the manuscript together with J.H. and T.K.

Funding

This work was supported by funding from the Ellison Foundation (PI L.A. Mucci) and the Nordic Union of Cancer (PI J. Kaprio). L.A. Mucci is supported by the Prostate Cancer Foundation. The Finnish Twin Cohort was supported by the Academy of Finland (grant numbers 213506, 129680, 265240, 263278) and Karolinska Institutet Distinguished Professor Award to Prof. H-O Adami (Dnr: 2368/10-221). The Ministry for Higher Education financially supports the Swedish Twin Registry.

Acknowledgments

We acknowledge the contribution of all the researchers and staff of the twin and cancer registries contributing to the NorTwinCan project.

Conflict of interest statement:

Tellervo Korhonen and Jaakko Kaprio have consulted for Pfizer on nicotine dependence from 2012 to 2015. Other authors declare no conflict of interest.

References

1. Boffetta P, Trichopoulos D. Cancer of the lung, larynx, and pleura. In (Adami H, Hunter DJ, Trichopoulos D, eds). *Textbook of cancer epidemiology*. Oxford; New York: Oxford University Press, 2008.

2. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer* 2002;99:260–6.

3. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;343:78–85.

4. Mucci LA, Hjelmborg JB, Harris JR, et al. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA.* 2016;315(1):68-76.

5. Gabrielsen ME, Romundstad P, Langhammer A, Krokan HE, Skorpen F. Association between a 15q25 gene variant, nicotine-related habits, lung cancer and COPD among 56,307 individuals from the HUNT study in Norway. *Eur J Hum Genet* 2013; 21:1293–1299.

6. Timofeeva MN, Hung RJ, Rafnar T, et al. Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. *Hum Mol Genet* 2012;21:4980–95.

7. 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;23:2303–10.

8. 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;8:53–9.

9. 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.

10. 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.

11. 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;41:139–50.

12. Kaprio J, Koskenvuo M. Genetic and environmental factors in complex diseases: the older Finnish Twin Cohort. *Twin Res* 2002;5:358–65.

13. Nilsen TS, Brandt I, Magnus P, Harris JR. The Norwegian Twin Registry. *Twin Res Hum Genet*. 2012;15:775–80.

14. Harris JR, Magnus P, Tambs K. The Norwegian Institute of Public Health twin program of research: an update. *Twin Res Hum Genet.* 2006; 9:858–64.

15. 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.

16. Pedersen NL, Lichtenstein P, Svedberg P. The Swedish Twin Registry in the third millennium. *Twin Res* 2002;5:427–32.

17. Allignol A, Schumacher M, Beyersmann J. Empirical Transition Matrix of Multi-State Models: The etm Package. *J Stat Softw* 2011;38.

18. Scheike TH, Holst KK, Hjelmborg JB. Estimating twin concordance for bivariate competing risks twin data. *Stat Med* 2014;33:1193–204.

19. Scheike TH, Holst KK, Hjelmborg JB. Estimating heritability for cause specific mortality based on twin studies. *Lifetime Data Anal* 2014;20:210–33.

20. Risch N. Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* 1990;46:222–8.

21. Risch N. The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiol Biomarkers Prev* 2001;10:733–741.

22. Holst KK, Scheike T, Hjelmborg JB. The liability threshold model for censored twin data [published online ahead of print January 2015]. *Computational Statistics & Data Analysis* 2015; doi: 10.1016/j.csda.2015.01.014.

23. Neale MC, Cardon LR, North Atlantic Treaty Organization. Scientific Affairs Division. *Methodology for genetic studies of twins and families*. Dordrecht ; Boston: Kluwer Academic Publishers, 1992.

24. Sham P. *Statistics in human genetics.* London; New York: Arnold; John Wiley & Sons, Inc., 1998.

25. Fisher RA. Cancer and smoking. *Nature* 1958;182:596.

26. Holst K, Scheike, T. H. mets: Analysis of Multivariate Event Times, R package version 0.2.8.1, http://lava.r-forge.r-project.org/

27. Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. *Br Med J* 1952; 2:1271–1286.

28. Wynder EL, Graham EA. Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma: a study of 684 proved cases. *J Am Med Assoc.* 1950; 143:329–336.

29. Floderus B, Cederlof R, Friberg L. Smoking and mortality: a 21-year follow-up based on the Swedish Twin Registry. *Int J Epidemiol* 1988;17:332–340.

30. Kaprio J, Koskenvuo M. Cigarette smoking as a cause of lung cancer and coronary heart disease. A study of smoking-discordant twin pairs. *Acta Genet Med Gemellol* (Roma) 1990;39:25–34.

31. Rose RJ, Broms U, Korhonen T, Dick DM, Kaprio J. Genetics of Smoking behavior. In: Kim YK, ed. *Handbook of Behavior Genetics*. New York: Springer, 2009:411–432.

32. Braun MM, Caporaso NE, Page WF, Hoover RN. A cohort study of twins and cancer. *Cancer Epidemiol Biomarkers Prev* 1995;4:469–73.

33. Yang IA, Holloway JW, Fong KM. Genetic susceptibility to lung cancer and co-morbidities. *J Thorac Dis* 2013;5:S454-62.

34. Bierut LJ, Stitzel JA, Wang JC, et al. Variants in nicotinic receptors and risk for nicotine dependence. *Am J Psychiatry* 2008;165:1163–1171.

35. Fowler CD, Lu Q, Johnson PM, Marks MJ, Kenny PJ. Habenular alpha5 nicotinic receptor subunit signalling controls nicotine intake. *Nature* 2011;471:597–601.

36. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 2008;452:638–642.

37. Thorgeirsson TE, Gudbjartsson DF, Surakka I, et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet* 2010;42:448–53.

38. Liu JZ, Tozzi F, Waterworth DM, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet* 2010;42:436–440.

39. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 2010;42:441–447.

40. Keskitalo K, Broms U, Heliovaara M, et al. Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (CHRNA3/CHRNA5/CHRNB4) on chromosome 15. *Hum Mol Genet* 2009;18:4007–40.

41. Munafo MR, Timofeeva MN, Morris RW, et al. Association between genetic variants on chromosome 15q25 locus and objective measures of tobacco exposure. *J Natl Cancer Inst* 2012;104:740–748.

42. Carr SR, Akerley W, Hashibe M, Cannon-Albright LA. Evidence for a genetical contribution to non-smoking-related lung cancer. *Thorax* 2015; doi:10.1136/thoraxjnl-2014-206584.

43. Sampson JN, Wheeler WA, Yeager M et al. Analysis of Heritability and Shared Heritability Based on Genome-Wide Association Studies for Thirteen Cancer Types. *J Natl Cancer Inst*. 2015 Oct 12;107(12):djv279. doi: 10.1093/jnci/djv279. Print 2015 Dec.

Table 1. Characteristics of the twin cohorts included in the analyses by zygosity and sex (individuals with smoking data), NorTwinCan

Sex and zygosity of twin individuals	**Denmark**	**Finland**	**Norway**	**Sweden**	**Total**
Males					
MZ	5,309	3,421	2,532	8,525	19,787
DZ	8,263	8,035	3,313	14,262	33,873
UZ	480	1,247	-	1,131	2,858
All males	14,052	12,703	5,845	23,918	56,519
Females					
MZ	6,570	3,940	3,074	10,141	23,725
DZ	9,525	8,092	3,788	16,617	38,022
UZ	473	1,049	-	996	2,518
All females	16,568	13,081	6,862	27,754	64,265
Birth cohort included	1870–1982	1880–1957	1915–1960	1886–1958	
1st Year of assessment of smoking and start of lung cancer occurrence follow-up	1959	1975	1980	1961	
End of follow-up for lung cancer occurrence	2010	2011	2009	2010	
Number of incident lung cancers	354	341	152	661	1508
Mean age at baseline (years)	49.0	36.2	38.3	38.9	
Mean follow-up time (years)	10.2*	30.1	24.6	32.1	

Note: **The 5,376 twins with unknown zygosity are included in the table but are excluded from pairwise analysis.**

*In Denmark, smoking data came from eight surveys conducted from 1959 to 2002.

Table 2. The numbers of pairs concordant and discordant for lung cancer at the end of follow-up by baseline pairwise smoking status and zygosity.

	Pairwise lung cancer status					
	Monozygotic			**Dizygotic**		
Baseline pairwise smoking status	Number of Concordant Pairs		Number of Discordant Pairs	Number of Concordant Pairs		Number of Discordant Pairs
Concordant pairs for smoking	Neither affected	Both affected	One twin in the pair affected	Neither affected	Both affected	One twin in the pair affected
Never / Never	7827	1	43	10709	0	59
Ever / Ever	7942	29	332	11474	28	527
Current / Current#	4741	24	241	6341	24	356
Discordant pairs for smoking						
Never / Ever	3234	0	40	8177	0	173
Never / Current##	1982	0	35	5511	0	144

# Current/current pairs are a subset of ever/ever pairs

## Never/current pairs are a subset of the never/ever pairs.

Table 3. Heritability estimates for lung cancer in the NorTwinCan cohort among those in the present analysis with smoking data, with and without adjustment for smoking status (n=1508 cases). All estimates adjusted for country and sex.

Number of complete MZ/DZ pairs	**Casewise concordance rates 95% Confidence Intervals**		**Adjustment for smoking**	**Variance component estimates 95% Confidence Intervals**		
	MZ	**DZ**		**A**	**C**	**E**
5299 9359	0.22 0.15 to 0.29	0.13 0.09 to 0.17	No	0.38 0.05 to 0.72	0.06 0.00 to 0.31	0.55 0.43 to 0.68
			Yes	0.34 0.00 to 0.70	0.02 0.00 to 0.29	0.64 0.50 to 0.78

Note: Variance components are: A: additive genetic effects, C: common environmental effects, and E: individually unique environmental effects estimated from biometrical twin model taking into account censoring (see methods in the online supplement).

Table 4. Pairwise correlations in liability, heritability estimates and model fit parameters for liability to incident lung cancer among ever smoking and current smoking concordant twin pairs from the NorTwinCan study. Estimates of genetic (A), shared environmental (C), and unshared environmental (E) variance are presented for the ACE, AE, and CE models.

Model	Correlation (95% CI)		A	C	E	AIC	p-value
	MZ	DZ	Estimate (95%CI)	Estimate (95%CI)	Estimate (95%CI)		
Ever smokers							
ACE	0.35 (0.21–0.49)	0.21 (0.09–0.33)	0.28 (0.0–0.66)	0.07 (0.0–0.36)	0.65 (0.50–0.79)	38759.12	0.01[1]
AE			0.37 (0.25–0.49)	0 -	0.63 (0.51–0.75)	38757.92	0.35
CE			0	0.28 (0.19–0.37)	0.72 (0.63–0.81)	38764.19	0
Current smokers							
ACE	0.39 (0.20–0.55)	0.24 (0.10–0.38)	0.29 (0.0–0.74)	0.10 (0.0–0.44)	0.62 (0.44–0.79)	30484.27	0.12[1]
AE			0.41 (0.26–0.56)	0 -	0.59 (0.44–0.74)	30483.46	0.27
CE			0	0.31 (0.20–0.42)	0.69 (0.58–0.80)	30488.49	0.01

[1] Compared to saturated model, the other models are compared to ACE model.

[2] 95%CI for C effect here could not be estimated reliably

Table 5. Lung cancer in twin pairs discordant for smoking at baseline by zygosity and smoking status

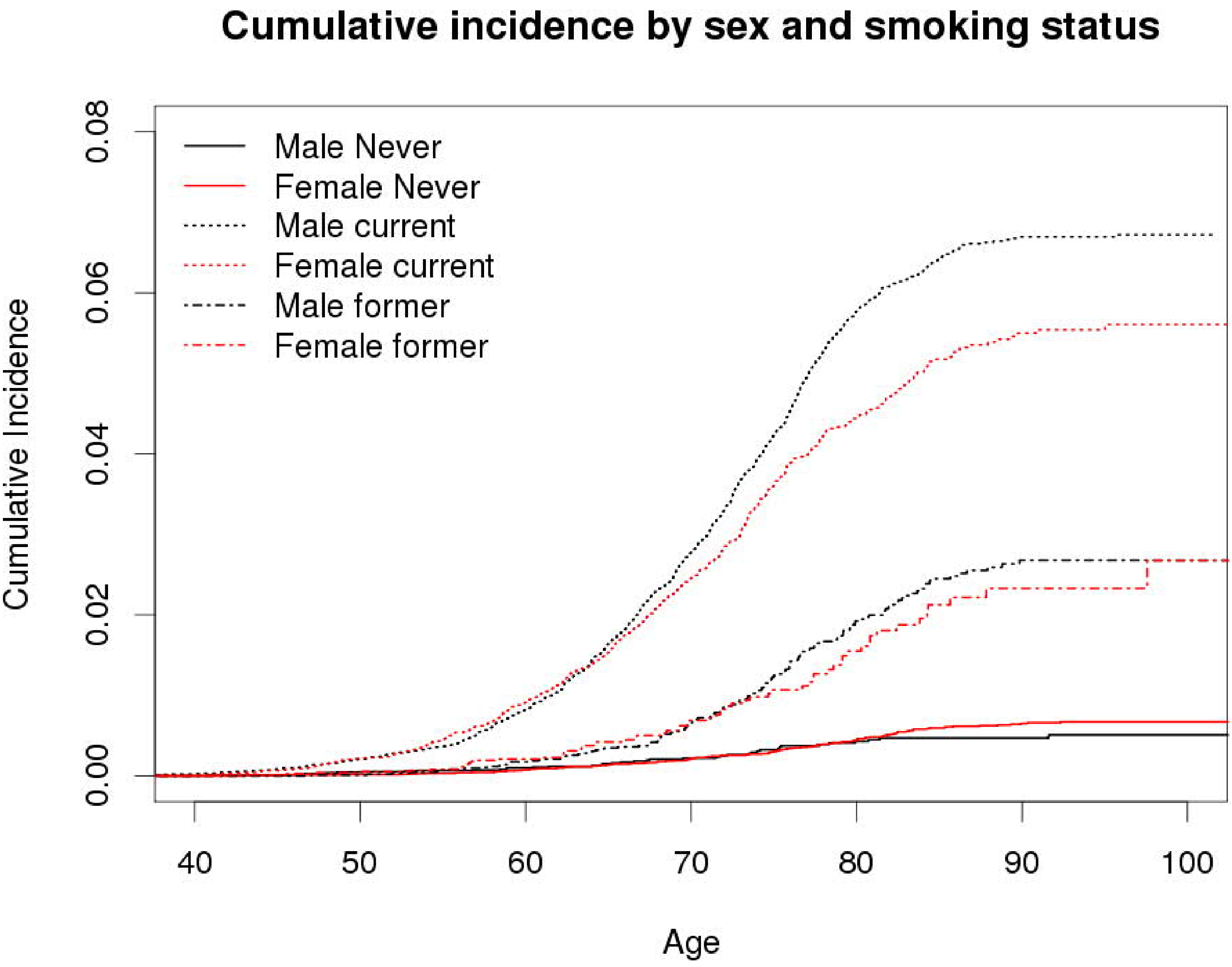
Smoking discordance	Zygosity	Pairs in which smoker had lung cancer and the non-smoking cotwin did not	Pairs in which non- smoker had lung cancer and the smoking cotwin did not	Hazard ratios (95% CI) and p-value
Ever/never	MZ	35	5	5.4 (2.1–14.0); p=0.0005
	DZ	145	28	5.0 (3.2–7.9); p=1.4e-12
Current/never	MZ	31	4	6.0 (2.1-17.3) p=0.001
	DZ	124	20	5.9 (3.5-9.8) p=1.4e-11

Figure legends

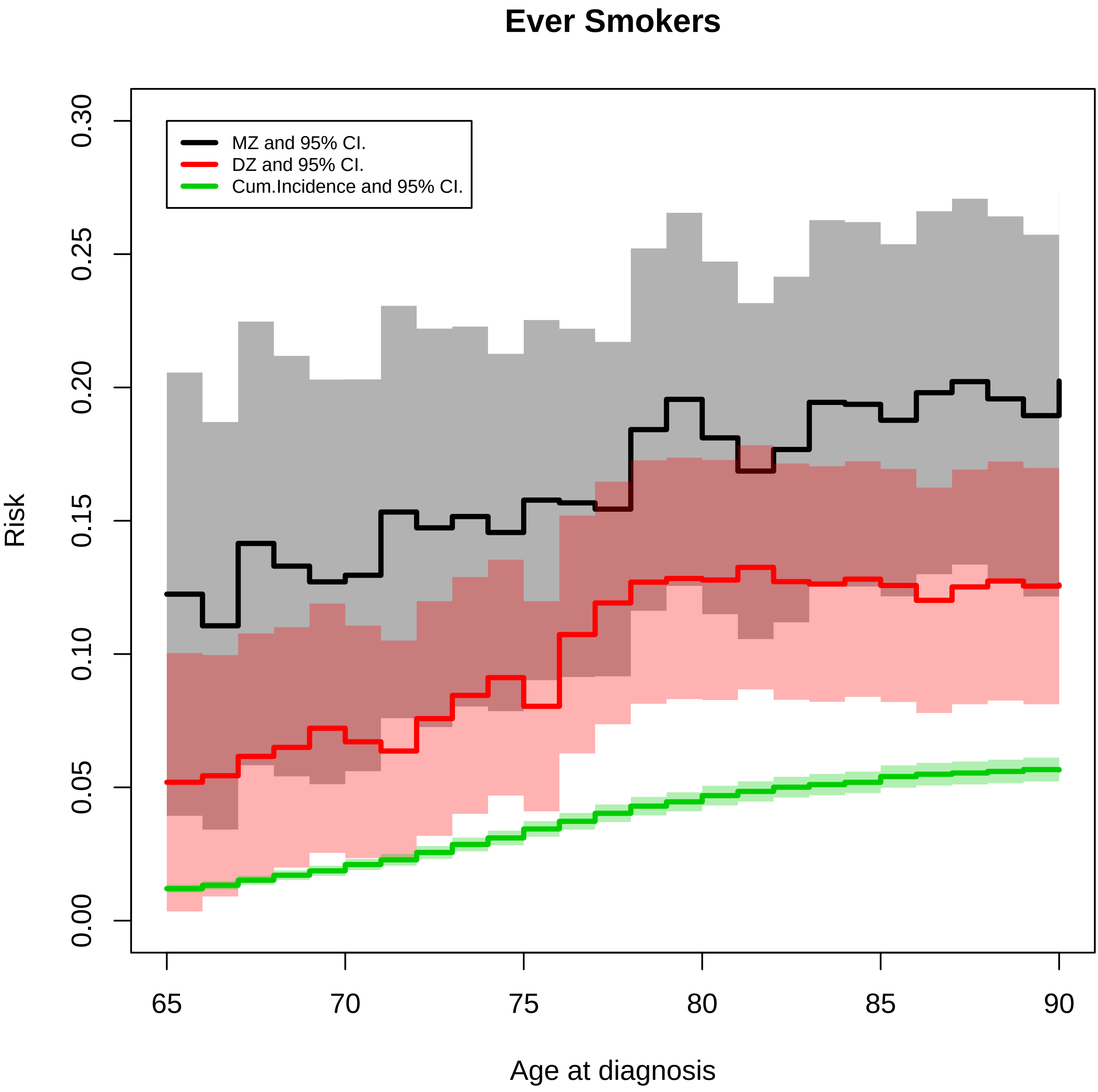
Figure 1. Cumulative incidence of lung cancer by smoking status (never, former, current) and sex (male, female). Cumulative incidence curves are adjusted for censoring, delayed entry to cancer registration, and competing risk of death. (Continuous lines are for never smokers, dashed lines for former smokers and dotted lines for current smokers; black for males and red for females).

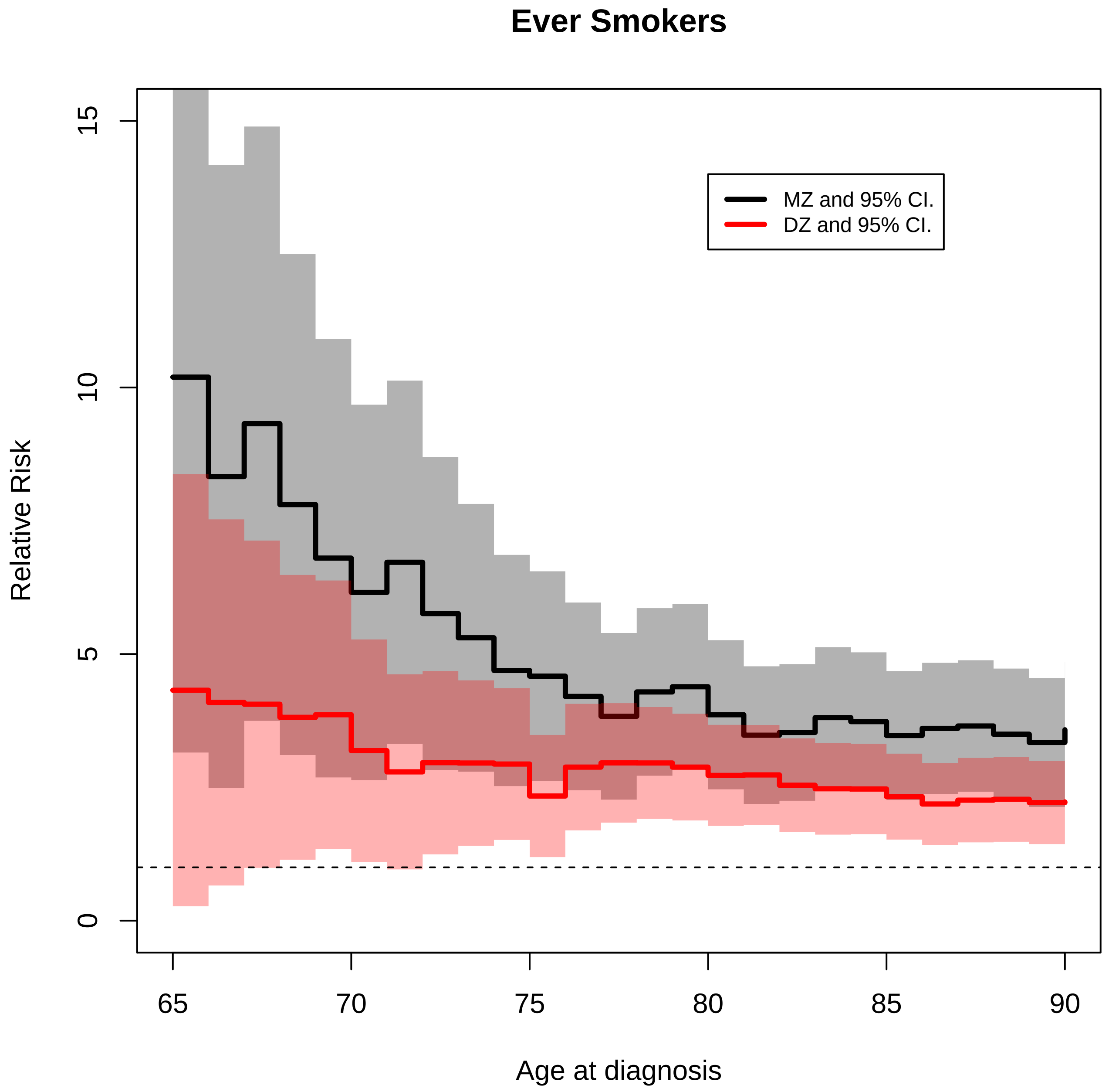
Figure 2. Case-wise concordance risk of lung cancer in MZ and DZ pairs compared to population risk among ever smokers, by age at diagnosis.

Figure 3. Relative recurrence risk ratio of lung cancer in MZ and DZ pairs compared to population risk among ever smokers, by age at diagnosis.



Cumulative incidence by sex and smoking status


Ever Smokers
MZ and 95% CI.
DZ and 95% CI.
Cum.Incidence and 95% CI.
Risk
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Age at diagnosis
65
70
75
80
85
90


Ever Smokers
MZ and 95% CI.
DZ and 95% CI.
Relative Risk
0
5
10
15
65
70
75
80
85
90
Age at diagnosis