

Paper III

Existing data sources in clinical epidemiology: The Scandinavian Thrombosis and Cancer (STAC) Cohort

Hilde Jensvoll,^{1,2} Marianne T. Severinsen,^{3,4} Jens Hammerstrøm,⁵ Sigrid K. Brækkan,^{1,2} Søren R. Kristensen,^{4,6} Suzanne C. Cannegieter,⁷ Kristine Blix,^{1,2} Anne Tjønneland,⁸ Frits R. Rosendaal,^{1,7} Olga Dziewiecka,¹ Kim Overvad,^{9,10} Inger Anne Næss¹¹ and John-Bjarne Hansen^{1,2}

¹K.G. Jebsen – Thrombosis Research and Expertise Center (TREC), Department of Clinical Medicine, UiT - The Arctic University of Norway, Tromsø, Norway

²Division of Internal Medicine, University Hospital of North Norway, Tromsø, Norway

³Department of Hematology, Aalborg University Hospital, Aalborg, Denmark

⁴Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

⁵Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway

⁶Department of Clinical Biochemistry, Aalborg University Hospital, Aalborg, Denmark

⁷Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, Netherlands

⁸Diet, Genes and Environment, Danish Cancer Society Research Center, Copenhagen, Denmark

⁹Department of Cardiology, Aalborg University Hospital, Aalborg, Denmark.

¹⁰Department of Public Health, Section for Epidemiology, Aarhus University, Aarhus, Denmark

¹¹Department of Hematology, Trondheim University Hospital, Trondheim, Norway

Running header: The Scandinavian Thrombosis and Cancer Cohort

Word count abstract: 239

Word count main text (excluding tables and references): 3509

Number of tables: 4

Number of figures: 4

Number of references: 23

Correspondence to: Hilde Jensvoll, K.G. Jebsen -Thrombosis Research and Expertise Center,

Department of Clinical Medicine, UiT - The Arctic University of Norway, Tromsø, Norway

E-mail: hilde.jensvoll@uit.no, telephone: +47 77 62 09 92

Abstract

Background: Although venous thromboembolism (VTE) is a known common complication in cancer patients, there is limited knowledge on patient-related and cancer-specific risk factors in the general population. The Scandinavian Thrombosis and Cancer (STAC) Cohort was established by merging individual data from three large Scandinavian cohorts (The Tromsø Study, the Nord-Trøndelag Health (HUNT) Study and the Danish Diet, Cancer and Health (DCH) Study). Here, we present the profile of the STAC Cohort and provide age-specific incidence rates of VTE and cancer.

Methods: The STAC Cohort includes 144952 subjects aged 19-101 years without previous VTE or cancer. Baseline information collected in 1993-1997 included physical examination, self-administered questionnaires and blood samples. Validated VTE events and cancer diagnoses were registered up to 2007-2012.

Results: There were 2444 VTE events (1.4 per 1000 person-years) during follow-up, and the incidence increased exponentially from 0.3 per 1000 person-years in subjects aged 20-29 years to 6.4 per 1000 person-years in subjects aged 80+. Overall, 51% of the VTE events were provoked, and cancer was the most common provoking factor (19%), followed by immobilization and surgery (both 15%). In total, 19757 subjects developed cancer during follow-up (9.8 per 1000 person-years), and the 5-year age-specific incidence rates of cancer were concordant with corresponding rates from the Norwegian Cancer Registry.

Conclusion: The STAC Cohort will provide a unique opportunity to explore the epidemiology and impact of genetic and environmental patient-related and cancer-specific risk factors of VTE in the general population.

Introduction

Venous thromboembolism (VTE), including deep venous thrombosis and pulmonary embolism, is a common complication and a predictor of poor survival in cancer patients.¹ The reported annual incidence proportions of VTE in cancer patients varies substantially between 0.5% and 20%,² and cancer is associated with 20-25% of all incident VTE events.³⁻⁵ Thromboprophylaxis is not recommended to all cancer patients due to the heterogeneity of VTE risk in various cancer types.⁶ These considerations emphasize the need to identify risk factors for cancer-related VTE in order to recognize high-risk patients who may benefit from thromboprophylaxis.

Discharge registries may represent large and valuable databases when exploring the risk of VTE in cancer patients. However, a recent Danish study reported that the positive predictive value of VTE discharge diagnoses was only 75%, indicating that registry data without any further VTE validation should be used with caution in medical research.⁷ We have established the Scandinavian Thrombosis and Cancer (STAC) Cohort, which is a large population-based cohort study including individual data from three well-established prospective cohort studies in Norway and Denmark. These cohorts include thorough validation of symptomatic VTE events and the information regarding cancer exposure is obtained from complete national cancer registries.

The overall aim of the STAC Cohort is to investigate the impact of cancer on the risk of VTE by exploring genetic and environmental patient-related and cancer-specific risk factors in order to identify high-risk patients who may benefit from thromboprophylaxis. Age-specific absolute and relative risks of VTE and related complications (mortality and recurrent VTE) in cancer patients will be calculated and compared with the general population. The risk will be assessed for overt and occult cancer, within various cancer sites, across cancer stages,

across histologic grades and at different time intervals after cancer diagnosis. The impact of established and novel genetic variants will be investigated by targeted gene sequencing in a large case-control study derived from STAC.

Here we present the overall design and methods of the STAC Cohort. We will also provide age-specific incidence rates of VTE and cancer in the three cohorts and in the STAC Cohort. The age-specific incidence rates of cancer in the STAC Cohort will be compared with expected incidence rates based on data from the Cancer Registry of Norway.

Methods

Study population

The STAC Cohort includes individual data from three large population-based cohort studies with enrollment in 1993-1997: the fourth survey of the Tromsø Study (Tromsø 4) in Norway, the second Nord-Trøndelag Health (HUNT 2) Study in Norway, and the Diet, Cancer and Health (DCH) Study in Denmark.

The Tromsø Study⁸ consists of six health surveys conducted in Tromsø, which is the largest city in Northern Norway (about 70000 inhabitants). The Tromsø Study was established to investigate the epidemiology of cardiovascular diseases and a broad spectrum of other chronic diseases. In the Tromsø 4 Study, conducted in 1994-1995, all inhabitants in the municipality of Tromsø ≥ 25 years were invited and 27158 attended (77% of the eligible population).

The HUNT Study⁹ has conducted three health surveys in the Nord-Trøndelag County, which is located geographically in the middle of Norway and includes six small towns (between 12000 and 21000 inhabitants) combined with sparsely populated rural areas. The

HUNT 2 Study in 1995-1997 was also designed to investigate cardiovascular disease and other chronic diseases in accordance with national health priorities. All residents of Nord-Trøndelag County ≥ 20 years were invited, and 65237 participated (69% of the eligible population).

The DCH Study¹⁰ was initiated in 1993-1997 and designed to investigate the associations between diet, lifestyle factors, cancer and other chronic diseases. In the DCH Study, all inhabitants aged 50-64 years who were born in Denmark, living in the urban areas of Copenhagen and Aarhus and without a previous diagnosis of cancer registered in the Danish Cancer Registry were invited, and 57054 participated (35% of those invited).

In the two Norwegian cohorts, subjects with a cancer diagnosis (registered in the Norwegian Cancer Registry) prior to enrollment were excluded before merging the three cohorts. Moreover, subjects with known VTE prior to baseline were excluded in all cohorts prior to merging. In the DCH Study, discharge diagnoses were used as basis for exclusions of subjects with prior VTE, whereas in the two Norwegian cohorts, subjects with prior VTE were detected through the medical records review. After merging, the study population counted 144952 subjects aged 19-101 years who were followed from date of inclusion (1993-97) to end of follow up (2007-2012). Figure 1 summarizes the numbers of invited and attending subjects in the three cohorts and the total number of participants in the STAC Cohort.

All included study participants gave their informed written consent, and respective regional committees for research ethics in Norway and Denmark approved the cohort studies and the collaboration study.

Baseline measurements

Baseline information was obtained by physical examination, self-administered questionnaires and non-fasting blood samples. Height and weight were measured with light clothing and without shoes, and body mass index (BMI) was calculated as weight in kilograms divided by the height in meters squared (kg/m^2). Trained personnel recorded blood pressure using an automatic device. Three measurements were performed after 2 minutes at rest in the Tromsø 4 Study and the HUNT 2 Study, and the average of the two last readings were chosen. In the DCH Study, blood pressure was measured twice after 5 minutes at rest, and the lowest reading was used.

Information regarding education level, smoking habits (never/former/current and duration in years), alcohol consumption (units per week), physical activity, diabetes mellitus and use of antihypertensives was collected from the self-administered questionnaires. The level of education was categorized into basic schooling (7-10 years), upper secondary education (high school/vocational school), and tertiary education (college/university). In the STAC Cohort, hard physical activity was defined as equal to or more than one hour of sports/physical activity per week that caused sweating or breathlessness. Detailed information about the baseline variables in each cohort is described elsewhere.⁸⁻¹⁰

Identification and validation of cancer diagnosis

Cancer registration is mandatory by law in both Norway and Denmark. Individual data from the three cohorts was linked to the Norwegian and Danish cancer registries by use of the unique national civil registration number, which is assigned to every Norwegian and Danish citizen at birth. The obtained information included date of cancer diagnosis, location of the disease (ICD10 codes C00-96), histological grade (ICO-3) and cancer stage (localized,

regional, distant or unknown). Subjects with non-melanoma skin cancers (ICD10 C44) were regarded as cancer-free. Both cancer registries are considered complete and valid, and evaluations have reported 98.8% completeness in Norway and 95-98% in Denmark, with 94% and 93% microscopically confirmed diagnoses, respectively.^{11,12}

Identification and validation of venous thromboembolic events

Only first time, objectively confirmed symptomatic VTE events have been included in the STAC Cohort. Trained personnel have reviewed the medical records for each potential case of VTE in all three included cohorts, and both inpatients and outpatients have been included.

In the Tromsø 4 Study,¹³ VTE events during follow-up were identified by searching the hospital discharge diagnosis registry, the radiology procedure registry and the autopsy registry at the University Hospital of North Norway, which is the only hospital serving the municipality of Tromsø. The four validation criteria were symptomatic VTE combined with objective confirmatory tests (ultrasound, venography, computed tomography (CT) scan, perfusion-ventilation scan, pulmonary angiography, or autopsy) which resulted in a VTE diagnosis that required treatment.

In the HUNT 2 Study, a VTE endpoint registry, the TROL Study was established.¹⁴ VTE events during follow-up were identified by use of the hospital discharge diagnosis registry and the radiology procedure registry at the two local hospitals (Levanger and Namsos) and at the tertiary-care center of the region, St Olav University Hospital. The validation criteria included symptomatic VTE events that required treatment and confirmation by objective diagnostic tests (ultrasound, venography, CT scan or perfusion-ventilation scan).

In the DCH Study,¹⁵ incident VTE events were found by use of the participants' civil registration numbers and linkage to the Danish National Patient Registry and the Danish National Death Registry. A verified VTE diagnosis required typical clinical symptoms combined with a confirmatory diagnostic test (ultrasound, venography, CT-scan, perfusion-ventilation scan, echocardiography or autopsy). Detailed information on the identification and validation of the VTE events in each cohort study has been previously described.¹³⁻¹⁵

The VTE events in all three cohorts were classified as deep venous thrombosis (DVT) or pulmonary embolism (PE), and concurrent DVT and PE was registered as PE. The VTE events were further classified as provoked and unprovoked VTE. A provoked event in the STAC Cohort required that one or more of the following provoking factors was present: overt cancer at the time of the VTE event, recent trauma or surgery, marked immobilization, travel, acute medical condition or other risk factors described by the physician in the medical record such as intravascular catheters. There were some differences in definitions of provoking factors in the three included cohorts.^{7,13,14} Provoking factors were registered up to 8 weeks prior to the VTE event in the Tromsø Study, while registered up to 3 months ahead of the event in the other two studies. Additionally, the minimum duration of the travel-variable defined to be a provoking factor varied between 4 and 8 hours in the three cohorts. Furthermore, the HUNT 2 Study did not include acute medical conditions as individual factors, but instead registered prolonged bed rest (immobility) due to acute medical conditions. The acute medical conditions included in both the Tromsø 4 Study and the DCH Study were stroke, acute myocardial infarction and infection, while the DCH Study also included exacerbation of chronic lung disease and activity in collagenous disease.

Study design

In our prospective cohort, all participants were followed from date of inclusion in 1993-97 to date of death, migration or end of follow-up for VTE and cancer registration in the respective cohorts. Incident VTE events were identified until Dec 31, 2012 in Tromsø 4, Dec 31, 2007 in HUNT 2, and April 30, 2008 in DCH. Cancer diagnoses were registered up to Dec 31, 2012 in Tromsø 4 and DCH and to Dec 31, 2009 in HUNT 2. Information regarding date of migration and death were obtained by linkage of the national civil registration numbers to the Central Population Registers in Norway and Denmark.

When comparing studies regarding risk of VTE in cancer patients, heterogeneity in the follow-up time is a methodological challenge, and it is recommended that follow-up should preferably start at the time of cancer diagnosis.¹ In the STAC Cohort, subjects with cancer or VTE prior to enrollment were excluded. Consequently, all individuals started to accrue non-exposed person-time in the beginning of the study period. Individuals who developed cancer during follow-up changed their exposure status, and started to accrue cancer-exposed observational time from the date of cancer diagnosis. Thus, individuals who developed cancer contributed with both non-exposed and cancer-exposed person-years at risk during the study period. The design of the STAC Cohort also facilitated an observational time with occult cancer exposure, which is particularly interesting as VTE-risk during this period is not influenced by cancer treatment, a well-known risk factor for cancer-related VTE.¹ Figure 2 illustrates the potential observational times.

Statistical analyses

Descriptive characteristics of the study population and the VTE events were presented within each cohort as means with 5-95% ranges for normally distributed continuous variables and as

medians with 25-75% ranges for continuous variables with skewed distribution. Categorical variables were given as percentages and absolute numbers. Ten-year age-specific IRs of VTE and cancer with 95% confidence intervals (CIs) were calculated as numbers of events per 1000 person-years at risk. In these analyses, the subjects contributed with person-years into different age groups, as they were aging during follow-up.

For overall cancer in the STAC Cohort, we calculated 5-year age-specific incidence rates per 1000 person-years in women and men for the period 1998-2009 and compared these figures graphically with corresponding incidence rates in the Norwegian Cancer Registry. This period was chosen because it included complete follow-up of cancer for all the three included cohorts. Statistical analyses were performed with STATA version 13.0 (Stata Corporation, College Station, Texas, USA).

Results

The distribution of baseline characteristics in the three included cohorts and in the merged STAC Cohort is presented in Table 1. Median age at inclusion was highest in the DCH Study, followed by the HUNT 2 Study and the Tromsø 4 Study (56, 48 and 45 years, respectively). The proportion of women was slightly higher in the STAC Cohort (52.6% versus 47.4%, respectively) as well as in the three cohorts separately. In the DCH Study, the proportion of subjects with low education (basic school) was lower than in the Tromsø 4 Study and HUNT 2 Study (14.9%, 33.9% and 34.2%, respectively). It has been reported that subjects with basic school as highest completed education are underrepresented in the DCH Study.¹⁰ However, subjects with basic school as highest completed education represent 26.7% of the participants in the STAC Cohort due to the higher proportion of subjects with low education in the Norwegian studies. In concordance with reported differences in alcohol consumption

between Norway and Denmark,¹⁶ participants in the DCH Study had substantially higher weekly alcohol consumption than the Norwegian participants. Participants in the HUNT 2 Study had the highest proportion of non-smokers. The systolic and diastolic blood pressure and use of antihypertensives increased with advancing median age in the cohorts.

Characteristics of patients with VTE by study origin are shown in Table 2. The proportion of provoked VTE was slightly higher in the Tromsø 4 Study (55.1%) than in the DCH Study and HUNT 2 Study (50.1% and 48.0% respectively). In the STAC Cohort, 51% of the VTE events were provoked while 49% were unprovoked. Cancer was the most common provoking factor in the STAC Cohort and associated with 18.9% of all the VTE events, followed by immobilization (15.1%) and surgery (14.6%).

Because the age-distributions differed between the included cohorts, we estimated 10-year age-specific incidence rates of VTE per 1000 person-years in order to enable comparisons between the cohorts (Table 3). There were 2444 VTE events during a median follow-up of 11.7 years, and the incidence increased with age for all cohorts. In the STAC Cohort, the overall incidence of VTE was 1.4 per 1000 person-years and the incidence increased from 0.3 per 1000 person-years in subjects aged 20-29 years to 6.4 per 1000 person-years in subjects above 80 years. In general, overall and age-specific incidence rates of VTE were highest in the Tromsø 4 Study, followed by the HUNT 2 Study and the DCH Study.

Table 4 presents 10-year age-specific incidence rates of cancer per 1000 person-years. In total, 19757 incident cancers were detected during a median follow-up of 14.1 years (9.8 per 1000 person-years), and the incidence increased with age, as expected, for all cohorts. The age-specific cancer rates were approximately similar in the three included cohorts, except for a higher incidence among subjects aged 50-59 years in the DCH Study.

The distribution of selected cancer sites within each cohort is shown in Figure 3. Prostate, breast, colorectal and lung were the most common cancers in all cohorts, accounting for 53-61% of the cancer diagnoses. Breast cancer constituted 17.8% of all cancer diagnoses in the DCH Study, while the corresponding percentages in the Tromsø 4 Study and the HUNT 2 Study were 12.2% and 11.6%, respectively. The 10-year age-specific incidence rates of breast cancer were higher in the DCH Study than in the Norwegian studies (data not shown).

Figure 4 displays 5-year age-specific incidence rates of overall cancer per 1000 person-years for women and men in the STAC Cohort between 1998 and 2009 compared with the corresponding rates given by the Norwegian Cancer Registry in the same time period. The age-specific incidence rates of cancers in women and men in the STAC Cohort are similar to national figures.

Discussion

We have established a large population-based cohort study with validated VTE and cancer events to investigate the risk of VTE in cancer patients. In accordance with previously reported incidence rates in adult populations in developed countries^{4,17,18} and the reported impact of age,^{19,20} we found an overall incidence rate of VTE in the STAC Cohort of 1.4 per 1000 person-years and an exponential increase at higher ages. The overall and 10-year age-specific IRs of VTE were highest in the Tromsø 4 Study, followed by the HUNT 2 Study and lastly the DCH Study. This may be explained by the 5-year longer follow-up time of VTE in the Tromsø Study, along with the most comprehensive procedure to identify VTE events.

Definitions of provoking factors varied somewhat between the three cohorts, but in spite of this, the proportion of provoked VTE events was approximately 50% in all studies, which is in accordance with previous population-based studies.^{4,5} The proportion of provoked

VTEs in a given study will clearly depend on the definition of included provoking factors, and a definite definition does not exist. For instance, the Tromsø 4 Study had a slightly higher proportion of provoked VTEs (55%), presumably because both acute medical conditions and trauma were registered as provoking factors. Conversely, the HUNT 2 Study did not include acute medical conditions and the DCH Study registered trauma-related VTEs, but found no events. Nevertheless, since the overall proportion of provoked VTEs was quite similar in the included cohorts, we regard the STAC Cohort suitable for future analyses differentiating between provoked and unprovoked VTE events.

The incidence of cancer increased similarly with age in all cohorts. The proportion of breast cancer, however, was higher in the DCH Study than in the Norwegian cohorts. The incidence rate of breast cancer is higher in Denmark than in Norway, and Copenhagen, where the majority of DCH participants are recruited from, has particularly high incidence rates.²¹ The DCH population has higher alcohol consumption and a lower proportion of subjects with basic school as their highest education level than the Norwegian cohorts. Interestingly, higher risk of breast cancer in women from the Copenhagen area was recently reported to be related to higher education level and suggested to be attributed to differences in alcohol consumption, reproductive patterns and use of hormone therapy.²² Similar results have been found in a Norwegian study including 93638 women.²³

Age-specific incidence rates of cancer in the STAC Cohort were similar to corresponding incidence rates from the Cancer Registry of Norway. The attendance rates in the Tromsø 4 Study and HUNT 2 Study were high (77% and 69% respectively), while the DCH had a low rate of 35%. Even though it has been reported that subjects with lower socioeconomic positions were under-represented in the DCH Study, the age-standardized incidence rates of cancer were not significantly different from the general Danish population.¹⁰

The main strengths of the STAC Cohort are the prospective design with data from three large, well-established cohort studies, characterized by wide age distribution, long follow-up and detailed information on cancer exposure, VTE events and confounders. The cancer events were obtained by linkage to complete cancer registries, and trained personnel confirmed all the VTE events according to objective criteria after a comprehensive identification process where different patient registries were examined. Some limitations merit consideration. We did not have verified information on previous history of VTE among all participants in the Norwegian cohorts. In these cohorts, individuals that were excluded due to prior VTE were detected through the medical records review, and the majority of these had a recurrent episode during the study period. Hence, some of the subjects who were treated as healthy participants during follow-up could be prevalent VTE cases who should have been excluded. In the DCH Study, discharge diagnoses were used as basis for exclusions of subjects with prior VTE, and may have led to some misclassifications. However, this would only lead to a small change in overall number of person-years in the STAC Cohort, and thus would presumably have a negligible influence on the incidence rates of VTE. The DCH Study did not search the radiology procedure registry for VTE events, which may have slightly underestimated the incidence of VTE in the DCH and hence in the STAC Cohort. Differences in follow-up time may also have influenced the overall risk of VTE and cancer in the three included cohorts. Harmonization of baseline variables, such as physical activity and education had some methodological challenges due to different questionnaires in Norway and Denmark, which may have resulted in some imprecisions in these variables. Another limitation is that we do not have access to measured biomarkers in blood (hemoglobin, platelet count, leukocyte count etc.) from all cohorts at baseline or during follow up. Information about cancer treatment (i.e. radiation and chemotherapy) is also limited and restricted to data from the national cancer registries. Lastly, information on lifestyle factors was only assessed at

baseline and may have changed over time and hence resulted in some non-differential misclassifications.

We have established a large Scandinavian cohort including individual data from three established population-based cohorts, the STAC Cohort, which will provide a unique opportunity to explore the impact of genetic and environmental risk factors, gene-environmental interactions, and cancer-specific characteristics on the risk of cancer-related VTE in the general population.

Acknowledgements

The study has used data from the Cancer Registry of Norway and the Danish Cancer Registry. The interpretation and reporting of these data are the sole responsibility of the authors, and no endorsement by these registries is intended nor should be inferred.

The authors declare that they have no conflicts of interests.

Table 1. Baseline characteristics obtained in 1993-1997 within the separate cohorts and in the merged STAC Cohort

	Tromsø 4	HUNT 2	DCH	STAC
Subjects, (n)	26093	62846	56013	144952
Age (years) ^a	45 (35-57)	48 (36-63)	56 (53-60)	53 (43-61)
Sex				
Men	47.7 (12438)	47.0 (29538)	47.6 (26674)	47.4 (68650)
Women	52.3 (13655)	53.0 (33308)	52.4 (29339)	52.6 (76302)
Questionnaire data				
Education				
Basic school	33.9 (8852)	34.2 (21471)	14.9 (8318)	26.7 (38641)
High school	35.9 (9366)	41.6 (26153)	62.7 (35130)	48.7 (70649)
College/University	29.8 (7780)	19.1 (11995)	22.1 (12375)	22.2 (32150)
Missing	0.4 (95)	5.1 (3227)	0.3 (190)	2.4 (3512)
Hard physical activity				
≥ 1 hour per week	30.5 (7948)	24.5 (15389)	31.7 (17736)	28.3 (41073)
Missing	< 0.1 (22)	10.3 (6462)	0.1 (75)	4.5 (6559)
Smoking				
Never	36.0 (9399)	44.6 (28009)	35.1 (19633)	39.4 (57041)
Former	26.6 (6947)	20.4 (12827)	28.8 (16116)	24.8 (35890)
Current < 15 cigarettes	24.5 (6401)	20.4 (12825)	13.1 (7342)	18.3 (26568)
Current 15-25 cigarettes	10.7 (2789)	6.7 (4231)	16.1 (9030)	11.1 (16050)
Current 25+ cigarettes	2.0 (509)	2.2 (1357)	6.9 (3857)	3.9 (5723)
Missing	0.2 (48)	5.7 (3597)	0.6 (35)	2.5 (3680)
Smoke time (years) ^{ab}	20 (12-30)	20 (11-30)	32 (20-39)	25 (14-35)
Missing	0.5 (88)	11.1 (3483)	0.7 (272)	4.4 (3843)
Alcohol units/week ^a	1 (0-3)	1 (0-2)	8 (3-18)	2 (1-7)
Missing	0.8 (197)	12.8 (8027)	<0.1 (3)	5.7 (8227)
Self-reported diabetes	1.7 (452)	2.9 (1794)	2.1 (1175)	2.4 (3421)
Use of antihypertensives	6.0 (1575)	10.7 (6718)	12.3 (6868)	10.5 (15161)
Health examination				
Systolic blood pressure (mmHg)	134 (109-175)	137 (109-179)	140 (109-176)	137 (109-177)
Diastolic blood pressure (mmHg)	78 (60-100)	80 (62-102)	83 (67-101)	81 (63-101)
Height (cm)	170 (155-186)	170 (156-186)	170 (156-185)	170 (156-185)
Weight (kg)	73 (53-97)	76 (56-101)	76 (55-100)	75 (55-100)
Body mass index (kg/m ²)	25.2 (19.9-32)	25.8 (20.6-33.7)	26.0 (20.4-33.3)	26.0 (20.4-33.3)

Values are percentages with absolute numbers in parentheses or means with 5-95% ranges except for ^a

^a Values are medians with 25-75% ranges due to skewed distribution of data.

^b Smoke time is reported for smokers only (previous and current), and the missing proportion is among smokers. The proportion of missing is not reported for variables that contain less than 2% missing.

Table 2. Characteristics of patients with venous thromboembolism (VTE) events at the time of VTE diagnosis within the separate cohorts and in the merged STAC Cohort

	Tromsø 4	HUNT 2	DCH	STAC
Subjects with VTE	664	989	791	2444
Age at VTE diagnosis	69 (44-88)	69 (41-88)	65 (56-74)	68 (45-87)
Sex (men)	46.7 (310)	47.9 (474)	56.4 (446)	50.3 (1230)
VTE type				
PE ^a	42.8 (284)	38.7 (383)	41.2 (326)	40.6 (993)
DVT ^a	57.2 (380)	61.3 (606)	54.5 (431)	58.0 (1417)
Unprovoked VTE ^b	44.9 (298)	52.0 (514)	49.9 (367)	49.4 (1179)
Provoked VTE ^b	55.1 (366)	48.0 (475)	50.1 (369)	50.6 (1210)
Provoking factors^b				
Overt cancer	22.9 (152)	15.7 (155)	19.7 (156)	18.9 (463)
Surgery	14.3 (95)	17.1 (169)	11.4 (84)	14.6 (348)
Trauma	7.7 (51)	10.4 (103)	0	6.4 (154)
Acute medical condition	14.3 (95)	-	10.2 (75)	7.1 (170)
Immobilization	17.6 (117)	16.3 (161)	11.1 (82)	15.1 (360)
Travel	0.9 (6)	2.5 (25)	5.7 (42)	3.1 (73)
Other factor	5.1 (34)	5.4 (53)	5.7 (42)	4.1 (129)

Values are percentages with absolute numbers in parentheses or mean with 5-95% ranges in parentheses.

^aThere are 34 subjects in the DCH Study with missing information for PE and DVT.

^bSeveral subjects have more than one provoking factor. The given percentages of unprovoked VTE, provoked VTE and each provoking factor are among total VTE events in each study with exclusion of 55 events with missing information on provoking factors in the DCH Study.

Table 3. Age-specific incidence rates (IRs) of venous thromboembolism per 1000 person-years (PY) with 95% confidence intervals within the separate cohorts and in the merged STAC Cohort

Age	Tromsø 4			HUNT 2			DCH			STAC		
	PY	Events	IR	PY	Events	IR	PY	Events	IR	PY*	Events	IR
20-29	6648	1	0.2 (0.02-1.1)	39515	12	0.3 (0.2-0.5)	-	-	-	46163	13	0.3 (0.2-0.5)
30-39	55087	17	0.3 (0.2-0.5)	105397	35	0.3 (0.2-0.5)	-	-	-	160484	52	0.3 (0.2-0.4)
40-49	99147	46	0.5 (0.3-0.6)	137592	78	0.6 (0.5-0.7)	-	-	-	236739	124	0.5 (0.4-0.6)
50-59	96481	118	1.2 (1.0-1.5)	141081	127	0.9 (0.8-1.1)	221503	153	0.7 (0.6-0.8)	459064	398	0.9 (0.8-1.0)
60-69	65749	139	2.1 (1.8-2.5)	101471	176	1.7 (1.5-2.0)	359362	480	1.3 (1.2-1.5)	526582	795	1.5 (1.4-1.6)
70-79	40241	179	4.4 (3.8-5.2)	83773	299	3.6 (3.2-4.0)	69888	158	2.3 (1.9-2.6)	193903	636	3.3 (3.0-3.5)
80+	20342	164	8.1 (6.9-9.4)	45779	262	5.7 (5.1-6.5)	-	-	-	66121	426	6.4 (5.9-7.1)
Total	383695	664	1.7 (1.6-1.9)	654609	989	1.5 (1.4-1.6)	650753	791	1.2 (1.1-1.3)	1689056	2444	1.4 (1.4-1.5)

Subjects contributed person-years within different age groups while aging during follow-up. Age at the time of VTE diagnosis was registered for the VTE events.

Table 4. Age-specific incidence rates (IRs) of cancer per 1000 person-years (PY) with 95% confidence intervals within the separate cohorts and in the merged STAC Cohort

Age	Tromsø 4			HUNT 2			DCH			STAC		
	PY	Events	IR	PY	Events	IR	PY	Events	IR	PY	Events	IR
20-29	7163	4	0.6 (0.2-1.5)	44183	30	0.7 (0.5-1.0)	-	-	-	51345	34	0.7 (0.5-0.9)
30-39	64668	63	1.0 (0.8-1.2)	127307	127	1.0 (0.8-1.2)	-	-	-	191975	190	1.0 (0.9-1.1)
40-49	114732	287	2.5 (2.2-2.8)	162592	359	2.2 (2.0-2.4)	-	-	-	277324	646	2.3 (2.2-2.5)
50-59	104188	551	5.3 (4.9-5.8)	166570	915	5.5 (5.1-5.9)	218949	1386	6.3 (6.0-6.7)	489707	2852	5.8 (5.6-6.0)
60-69	68363	903	13.2 (12.4-14.1)	120889	1417	11.7 (11.1-12.3)	447916	5752	12.8 (12.5-13.2)	637167	8072	12.7 (12.4-12.9)
70-79	39989	853	21.3 (20.0-22.8)	93720	1814	19.4 (18.5-20.3)	165723	3448	20.8 (20.1-21.5)	299432	6115	20.4 (19.9-20.9)
80+	19783	452	22.8 (20.8-25.1)	53227	1282	24.1 (22.8-25.4)	4004	114	28.5 (23.7-34.2)	77015	1848	24.0 (22.9-25.1)
Total	418886	3113	7.4 (7.2-7.7)	768487	5944	7.7 (7.5-7.9)	836592	10700	12.8 (12.6-13.0)	2023965	19757	9.8 (9.6-9.9)

Subjects contributed person-years within different age groups while aging during follow-up. Age at the time of cancer diagnosis was registered for the cancer events.

Figure 1. The STAC Cohort

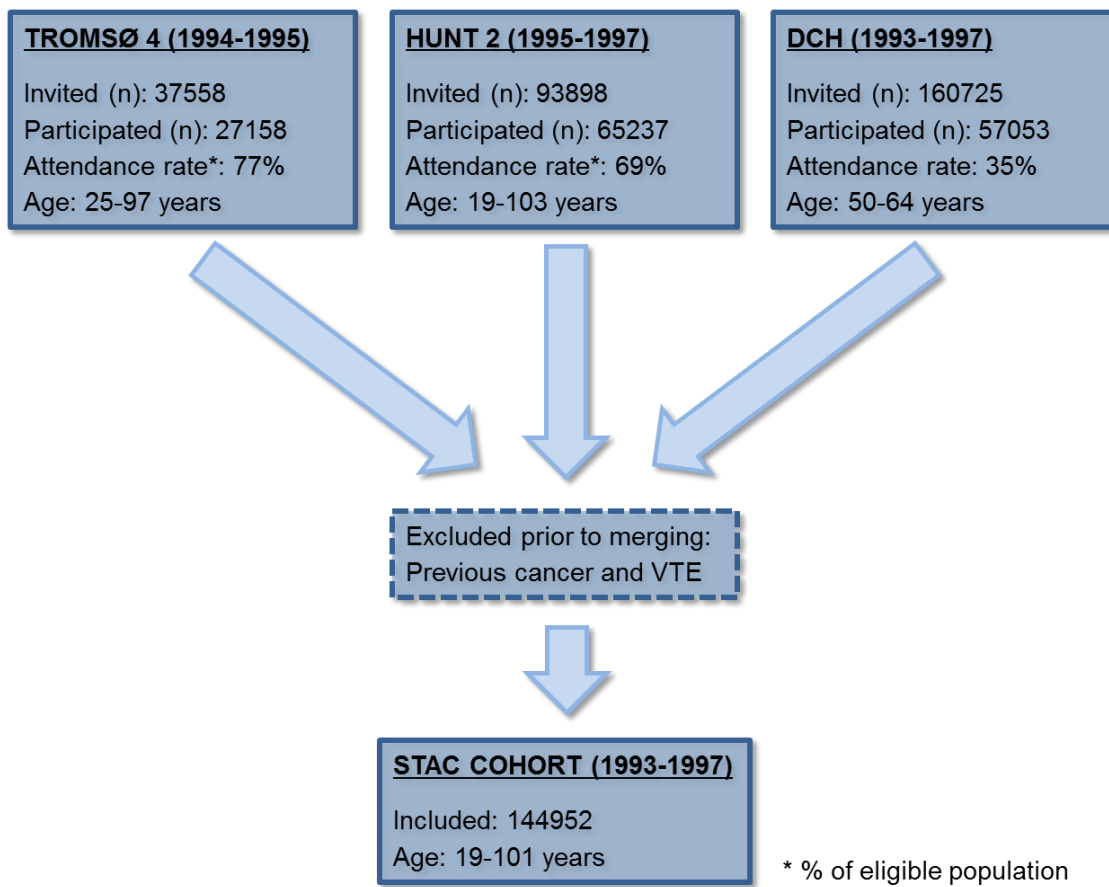
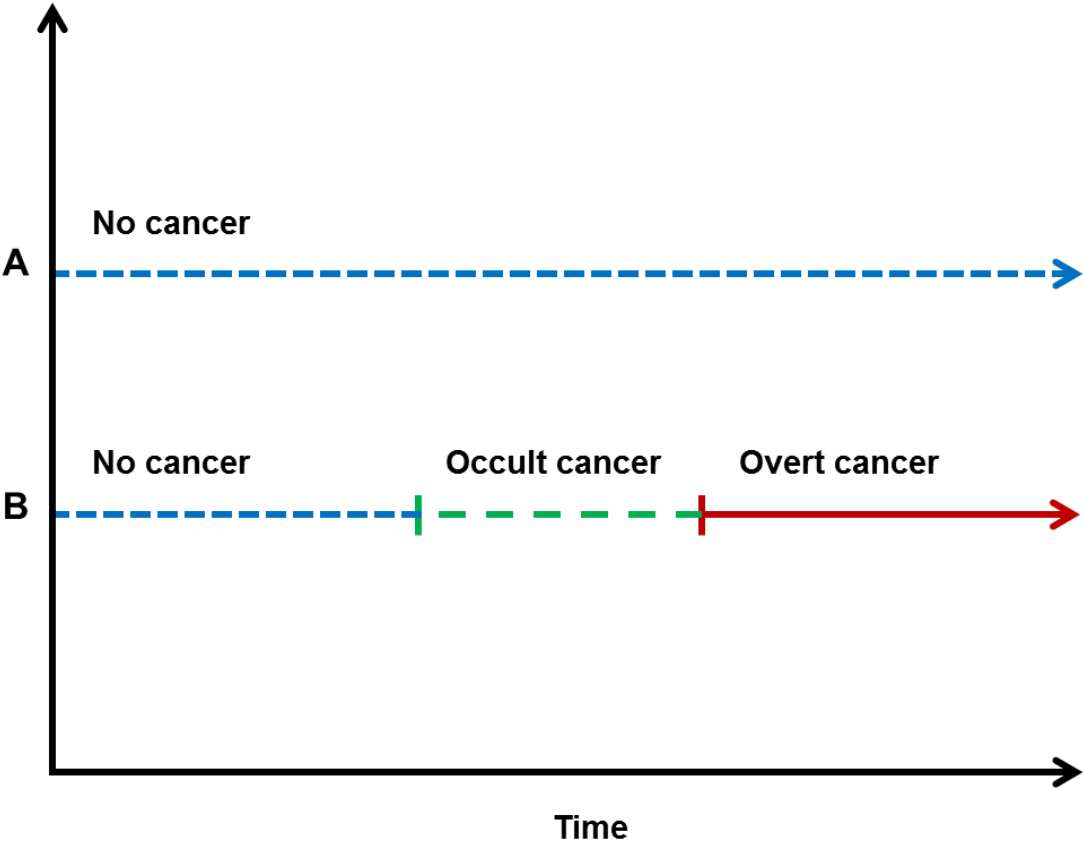


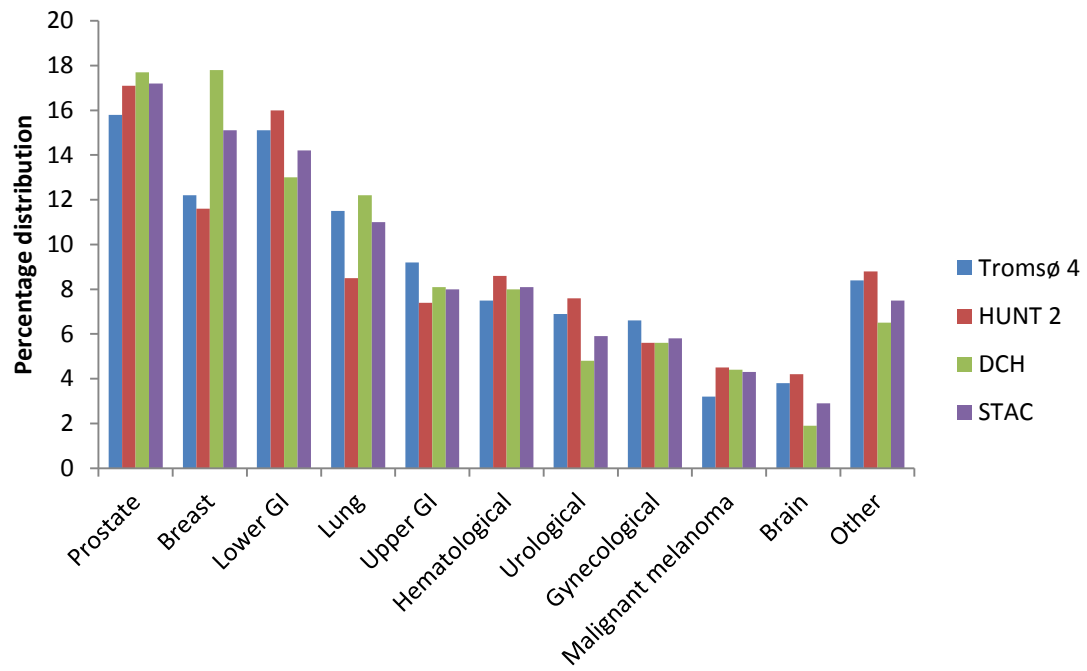
Figure 2. Potential observational periods in the STAC Cohort



A: A person who did not develop cancer during follow-up contributed person time at risk of VTE in the non-cancer exposed category during the entire study period.

B: A person who developed cancer during follow-up would contribute with person time at risk of VTE in three different categories during the study period: (i) the non-cancer exposed, (ii) the occult cancer exposed (starting one year before the cancer was diagnosed) and (iii) the overt cancer exposed starting from the date of cancer diagnosis.

Figure 3. Distribution of cancer sites within the separate cohorts and in the merged STAC Cohort

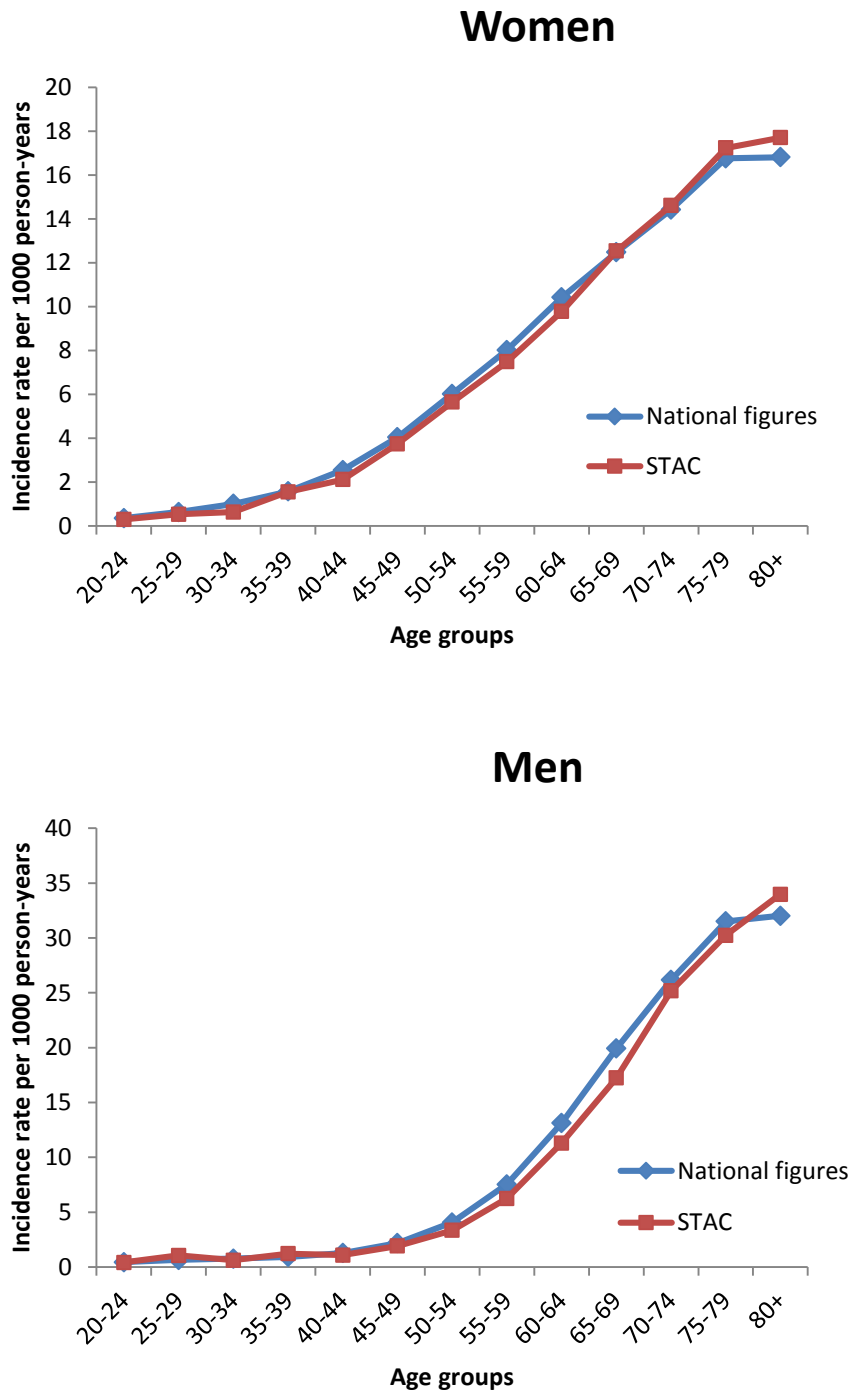


GI: Gastrointestinal

Lower: Colorectal and anal cancer

Upper: Esophagus, stomach, small intestine, hepatobiliary and pancreatic cancer

Figure 4. Age-specific incidence rates for all cancers in women and men in the STAC Cohort compared with national figures from the period 1998-2009



References

1. Timp JF, Braekkan SK, Versteeg HH, Cannegieter SC. Epidemiology of cancer-associated venous thrombosis. *Blood*. 2013;122(10):1712-1723.
2. Horsted F, West J, Grainge MJ. Risk of venous thromboembolism in patients with cancer: a systematic review and meta-analysis. *PLoS medicine*. 2012;9(7):e1001275.
3. Heit JA, O'Fallon WM, Petterson TM, et al. Relative Impact of Risk Factors for Deep Vein Thrombosis and Pulmonary Embolism: A Population-Based Study. *Arch Intern Med*. 2002;162(11):1245-1248.
4. Cushman M, Tsai AW, White RH, et al. Deep vein thrombosis and pulmonary embolism in two cohorts: the longitudinal investigation of thromboembolism etiology. *The American Journal of Medicine*. 2004;117(1):19-25.
5. White RH. The epidemiology of venous thromboembolism. *Circulation*. 2003;107(23 Suppl 1):I4-8.
6. Lyman GH, Khorana AA, Kuderer NM, et al. Venous thromboembolism prophylaxis and treatment in patients with cancer: American Society of Clinical Oncology clinical practice guideline update. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2013;31(17):2189-2204.
7. Severinsen MT, Kristensen SR, Overvad K, Dethlefsen C, Tjonneland A, Johnsen SP. Venous thromboembolism discharge diagnoses in the Danish National Patient Registry should be used with caution. *Journal of clinical epidemiology*. 2010;63(2):223-228.
8. Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njolstad I. Cohort profile: the Tromso Study. *International journal of epidemiology*. 2012;41(4):961-967.
9. Krokstad S, Langhammer A, Hveem K, et al. Cohort Profile: The HUNT Study, Norway. *International journal of epidemiology*. 2013;42(4):968-977.
10. Tjonneland A, Olsen A, Boll K, et al. Study design, exposure variables, and socioeconomic determinants of participation in Diet, Cancer and Health: a population-based prospective cohort study of 57,053 men and women in Denmark. *Scandinavian journal of public health*. 2007;35(4):432-441.
11. Larsen IK, Smastuen M, Johannesen TB, et al. Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. *Eur J Cancer*. 2009;45(7):1218-1231.
12. Storm HH, Michelsen EV, Clemmensen IH, Pihl J. The Danish Cancer Registry--history, content, quality and use. *Danish medical bulletin*. 1997;44(5):535-539.
13. Braekkan SK, Mathiesen EB, Njolstad I, Wilsgaard T, Stormer J, Hansen JB. Mean platelet volume is a risk factor for venous thromboembolism: the Tromso Study, Tromso, Norway. *Journal of thrombosis and haemostasis : JTH*. 2010;8(1):157-162.
14. Naess IA, Christiansen SC, Romundstad P, Cannegieter SC, Rosendaal FR, Hammerstrom J. Incidence and mortality of venous thrombosis: a population-based study. *Journal of thrombosis and haemostasis : JTH*. 2007;5(4):692-699.
15. Severinsen MT, Kristensen SR, Johnsen SP, Dethlefsen C, Tjonneland A, Overvad K. Smoking and venous thromboembolism: a Danish follow-up study. *Journal of thrombosis and haemostasis : JTH*. 2009;7(8):1297-1303.
16. Storm HH, Kejs AM, Engholm G, et al. Trends in the overall survival of cancer patients diagnosed 1964-2003 in the Nordic countries followed up to the end of 2006: the importance of case-mix. *Acta Oncol*. 2010;49(5):713-724.
17. Heit JA. Venous thromboembolism: disease burden, outcomes and risk factors. *Journal of thrombosis and haemostasis : JTH*. 2005;3(8):1611-1617.
18. Rosendaal FR. Venous thrombosis: a multicausal disease. *Lancet*. 1999;353(9159):1167-1173.

19. Tsai AW, Cushman M, Rosamond WD, Heckbert SR, Polak JF, Folsom AR. Cardiovascular Risk Factors and Venous Thromboembolism Incidence: The Longitudinal Investigation of Thromboembolism Etiology. *Arch Intern Med.* 2002;162(10):1182-1189.
20. Silverstein MD, Heit JA, Mohr DN, Petterson TM, O'Fallon WM, Melton LJ, 3rd. Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population-based study. *Arch Intern Med.* 1998;158(6):585-593.
21. Engholm G FJ, Christensen N, Johannesen TB, Khan S., Køtlum JE, Milter MC, Ólafsdóttir E, Pukkala E, Storm HH. . NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 6.1 (25.04.2014). Association of the Nordic Cancer Registries. Danish Cancer Society. Available from <http://www.ancr.nu>. Accessed 20.10.2014.
22. Hvidtfeldt UA, Lange T, Andersen I, et al. Educational differences in postmenopausal breast cancer--quantifying indirect effects through health behaviors, body mass index and reproductive patterns. *PloS one.* 2013;8(10):e78690.
23. Braaten T, Weiderpass E, Kumle M, Lund E. Explaining the socioeconomic variation in cancer risk in the Norwegian Women and Cancer Study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* 2005;14(11 Pt 1):2591-2597.