UiT The Arctic University of Norway

Faculty of Health Science, Department of Community Medicine

Adult Attained Height and Breast Cancer incidence (NOWAC)

A NOWAC Cohort study

Student: Henrik Hansen Wallumrød

Master's thesis in MPH - HEL-3950 - May 2023

Supervisors: Marko Lukic, Karina Standahl Olsen and Andrew D. D. Mashchak

Table of Contents

List of Tables

List of Figures

Acknowledgments

Many thanks to my supervisors, Karina, Marko, and Andrew, for their extensive knowledge, feedback, and patience throughout this process.

I am appreciative of my family and friends for their continuous support and love. Delivering a thesis would not have been possible had it not been for them.

A special thanks to all that contributed to proofreading: Andrew, Sarita (love you mom), Thea and Joe.

Henrik H. Wallumrød

Abstract:

Introduction: Adult attained height has been associated with breast cancer incidence consistently in the literature. The inclusion of covariates to adjust the regression models is widely used as a tool for controlling for error in models and reducing bias. However, several epidemiological studies regarding height have not accounted for confounding variables in their analysis.

Material and Methods: The data was collected from the Norwegian Women and Cancer (NOWAC) cohort study (1991-2003). To investigate the association between adult attained heigh and breast cancer risk we used Cox Proportional Hazard regression to estimate the Hazard Ratio (HR) for height and overall breast cancer, stratified by menopausal status. To handle missing data multiple imputation was performed. Cubic splines were fitted to the main exposure to investigate potential non-linearity. The regression-model was constructed using a causal framework (DAG), including variables such as age at baseline, age at menarche, body shape in childhood, breast cancer in mother and years of education. Sensitivity testing was performed to determine the overall influence each covariate had on the effect estimates.

Results: The imputed multivariable model indicated a positive association between increased height (per 10 cm increase) and overall breast cancer incidence (HR: 1.21, 95% CI: 1.17-1.25). The association was also observed both in premenopausal women (HR = 1.20, 95% CI 1.14-1.26) and postmenopausal women (HR = 1.22, 95% CI 1.16-1.28). Plots with restricted cubic splines indicated a linear association between height and breast cancer risk. Sensitivity analysis suggested that none of the included variables affected the model to any great extent except year of education.

Conclusion: The analysis found a positive association between increased height in Norwegian women and incidence of breast cancer.

Keyword: NOWAC, Adult Attained Height, Menopausal status, Breast Cancer, Direct Acyclic Graphs

Abbreviations:

DAG – Directed Acyclic Graphs

EPIC – European Prospective Investigation into Cancer and Nutrition.

GH – Growth Hormones

IGF-1 – Insulin like Growth Factors

NA – Not Available, referring to missing value(s).

NOWAC – Norwegian Women and Cancer Study

1 Background

1.1 Introduction - Height and Breast Cancer

The average standing height among people have increased in industrialized countries in tandem with improved living standards, income and education levels (NCD, 2016). Adult attained height is often measured as self-reported standing height. Adult stature is result of a complex interplay between the inherited biology in dynamic interaction with their environment from gestation until the end of puberty (Cole, 2003; Little, 2020).

Incidence of cancers is increasing, where the developing countries suffer the greatest increase (Soerjomataram & Bray, 2021). Height has consistently been associated with cancer incidence across nationalities, ethnicities, sex and age-cohorts (Aune et al., 2015; Choi et al., 2019; Green et al., 2011; Jing et al., 2015; Krieg et al., 2022; Lahmann et al., 2004; Lerro et al., 2010; Shrestha et al., 2021; Wang et al., 2017; WCRF, 2018; Zhang et al., 2015; Zhou et al., 2022).

The etiology of carcinogenesis is complex and multifactorial, breast cancer being no exception. The literature regarding potential causal relationship between height and breast cancer is divided. Some of this is reflected in the heterogeneity in model composition found in the literature. Several studies adjust for covariables that does not meet the criteria of confounding (Groenwold et al., 2021; Wysocki et al., 2022), thus risking bias by controlling for variables that could be mediators, colliders or ancestors of the exposures (Richiardi et al., 2013; Rudolph & Lau, 2021).

1.2 Breast Anatomy and Cancer Pathology

1.2.1 Breast development

Development of the mammary gland occurs during gestation, puberty, and pregnancy. In each developmental stage, sex hormones play an central role in the development of the mammary glands; primarily through increased cell proliferation and differentiation (Loda et al., 2016).

During embryonic period mammary gland structure is highly dependent on maternal and placental steroid hormones (mainly testosterone). After birth the infant is cut off from maternal and placental derived hormones, causing the glands to shrink, and development of breast structures halts (Loda et al., 2016). At menarche the onset of secretion of estrogen and progesterone restarts the development of the mammary gland, marked by thelarche (Loda et

al., 2016; Sherwood, 2016). Pregnancy is the only point where the mammary gland reaches full maturation. Much like in puberty, high levels of estrogen stimulate growth for ducts and the progesterone activates extensive alveolar-lobular growth (Loda et al., 2016).

1.2.2 Breast Anatomy

The breast consists of a network of ducts that branch out from the nipple and end up in the lobules surrounded by adipose tissue. The lobules consist of a clustering of alveoli, resembling sacks, epithelial lined alveoli have the potential to produce milk which is secreted from the alveolar epithelial cells into the lumen, while the myoepithelial cells causes contractions for transportation of milk into the ducts (Sherwood, 2016).

1.2.3 Hallmarks of Cancer - Biological Perspective

A defining trait of cancer development is the uncontrolled growth of pathological tissue(neoplasm). Normal tissue carefully regulates cell proliferation, meticulously maintaining cell numbers and structure of the tissue (Hanahan & Weinberg, 2011). This is through regulating the start and progression of the cell-cycle. The immune system normally controls and regulates cell proliferation through tumor suppressors genes. Tumor suppressor genes generally functions as regulators of the cell-cycle, halting progress and determining if the cell in questions is eligible for further proliferation or apoptosis. When the cell detects physiological stress induced by hyperproliferation and/or DNA damage, it activates increased apoptotic activity to regulate cell prevalence in the tissue. Cells have a replicative limit reducing the potential for unlimited growth. Extensive proliferation and forming of tumors require large amounts of sustenance making it challenging for uncontrolled growth to sustain itself over time (Hanahan & Weinberg, 2011).

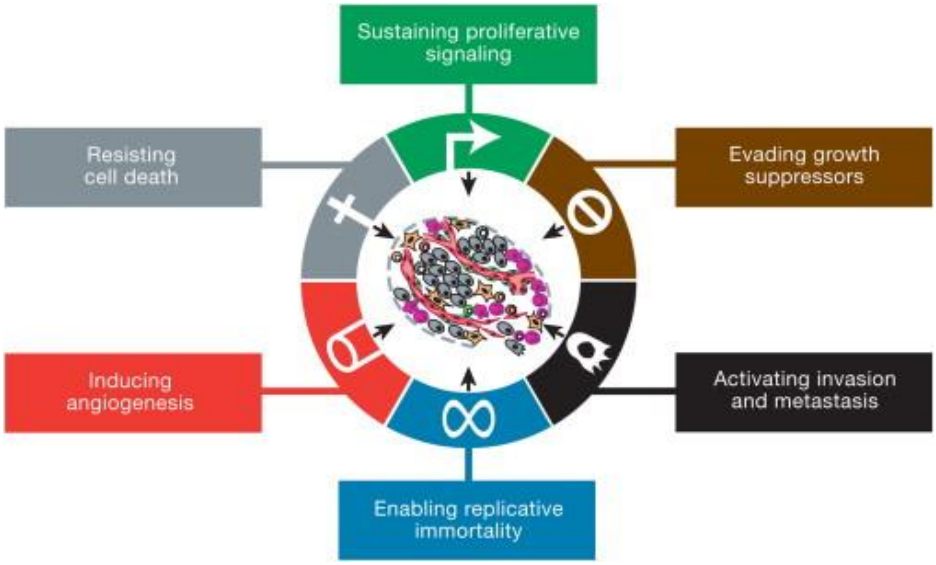


Figure 1 – Hallmarks of Cancer (Hanahan & Weinberg, 2011)

Neoplastic growth requires alteration to the genetic code of the cells. This alteration has been suggested to be enabled by two core characteristics: genomic instability and tumor-associated inflammation (premalignant and malignant). Genomic instability refers the number of mutations within a given genome, more mutation corresponds to increased instability with increased risk of mutations (Hanahan & Weinberg, 2011). Inflammations in tumor lesions is driven by immune cells combating the inflammation but unable to detect the neoplastic growth. The immune cell activity will both support biological traits of neoplastic growth and increased risk of mutation of the tissue (Hanahan & Weinberg, 2011). We can summarize that both characteristics influence tumor development through either influencing genomic change (e.g., DNA damage) or increased frequency of mutation.

Both characters increase the risk of mutations and furthering traits that are referred to as hallmarks of cancer. Hanahan and Weinberg (2011) define these traits as: sustaining proliferative signaling, evasion of suppressors, resisting cell death, inducing angiogenesis, enabling replicative immortality, and activation of invasion leading to metastasis (figure 2).

1.2.4 Classification:

Breast cancer specifically develops in the terminal duct-lobular unit TDLU, in the ducts (ductal neoplasms), the lobules (lobular neoplasms) or papillary (papillary neoplasm).

Besides the physical location where the tumor originates, there are several ways of categorizing tumors. Firstly, you have the four molecular subtypes of tumors: luminal A, Luminal B, basal-like and triple-negative, and human epidermal growth factor receptor 2. These are defined by their histology and correspond to different prognosis and outcome. Secondly, a tumor can be classified by its size, where larger sizes indicate a worse long-term prognosis (Loda et al., 2016).

1.2.5 Screening and diagnosis:

Breast cancer can be detected by observing symptoms like changes to the nipple or the discovery of palpable tumors. However, most detections of breast cancer are done by mammographic screening. Norway offers for any women aged 50-69 years for a mammographic screening through the BreastScreen Norway program every other year (CRN, 2023). Upon suspected neoplasm, either by mammography or clinical examination, the patient needs a biopsy to be able to determine the type of neoplastic growth and set a corresponding diagnosis (Loda et al., 2016).

1.2.6 Treatment:

Early detection and more refined methods of treatment allows for increased usage of breast conserving surgery with post-surgical radiation therapy. Chemotherapy or treatment with an antiestrogen can also be utilized to treat breast cancer (Loda et al., 2016).

1.3 Breast Cancer Epidemiology

1.3.1 Prevalence and Incidence

In 2020, 2,26 million new cases of breast cancer worldwide were reported, making it the most common cancer diagnosis with 11.7% of all cancers (GLOBOCAN, 2020). Breast cancer caused 685 000 deaths globally, being 6.9% of all cancer deaths (GLOBOCAN, 2020). Currently, the highest incidence is found in high income countries (GLOBOCAN, 2020; Huang et al., 2021). It is estimated that breast cancer incidence will increase with 131% by 2070 (Soerjomataram & Bray, 2021). Women in low-income countries are disproportionately affected by this disease (Huang et al., 2021), and the strongest reduction of both mortality and

comorbidities is by early detection and effective treatment (Huang et al., 2021; Soerjomataram & Bray, 2021).

Overall cancer incidence in Norway has increased steadily amounting to over 6000 more annual cases when comparing 2011 with 2021 (CRN, 2021). Incidence rates of all cancer have decreased 3.8% for males the last 5 years (2017-2021), and increased 2.1% for females (CRN, 2021). In total, 36 998 new cases of all cancers in 2021, of which 3 991 cases where breast cancer; making it the second largest cancer type in Norway surpassed only by prostate cancer (n = 5 188)(CRN, 2021).

1.3.2 Risk factors:

Breast cancer risk is not determined by any single risk factor. The risk of developing breast cancer is linked to the hallmarks of cancer through the enabling characteristics mentioned previously or a combination of both. Breast cancer is affecting women disproportionally, as it rarely that develops in men.

Age at menarche, and age at menopause have been linked to increased risk of breast cancer. Here an increased interval between age at menarche and age at menopause is associated with increased risk of breast cancer, the suggested mechanism is that increased reproductive span implies more menstrual cycles, thus more cumulative hormonal exposure (Loda et al., 2016).

Nulliparous women tend to be more at risk for breast cancer than parous women, suggesting that the more children born indicate a decreased risk of breast cancer. This association is dependent on age at first birth. As there is is an negative association between age at first birth and breast cancer risk (Loda et al., 2016). However, this association has a dual effect, as within ten years of pregnancy and birth will increase risk of breast cancer, but by 15 years the risk decreases (Loda et al., 2016). This might be due to increased cell proliferation in breast cells during the pregnancy, while the differentiation of breast cell after cessation of lactation might reduce the risk long term (Loda et al., 2016).

Breastfeeding over 6 months is associated with decreased risk of breast cancer, where the main mechanism is suggested to be delay of ovulatory cycles and also the terminal differentiation of breast cells (Loda et al., 2016).

Oral contraceptives consist of sex hormones (estrogen and progestogen), increasing breast tissue to exposure to hormones that increases cell proliferation that again increase the risk of breast cancer. The association tends towards being for recent use, where long term or ever use seems not to be associated with breast cancer risk (Loda et al., 2016).

Table 1 – *Breast Cancer Risk Factors*

Risk factor	Comparison		Effect estimation	I^2	Reference
Age at menarche	1 year delay	All	RR: 0.95	-	Loda et al. (2016)
Parity	Nulliparous vs parous	All	RR: 1.2-1.7	-	Loda et al. (2016)
Breastfeeding	Each year a woman breastfeeds	All	RR: 0.96 (CI 95% 0.93-0.96)	-	Loda et al. (2016)
	Per 5-month duration	All	RR: 0.98 (CI 95% 0.97-0.99)	-	WCRF (2017)
Oral contraceptives	Longer duration	All	RR: 1.24-1.54	-	Loda et al. (2016)
Menopausal hormone therapy	Lunger duration (5+ years vs never)	All	RR 1.30-1.47	-	Loda et al. (2016)
BMI Early-life adiposity (<25y)	Per 5 kg/m^2	All	RR: 0.84 (95% CI: 0.81-0.87)	63,4%	Byun et al. (2022)
BMI (18y-30y)	Per 5 kg/m^2 increase	Premenopausal	RR: 0.82 (95% CI 0.76-0.89)	14.9%	WCRF (2017)
		Postmenopausal	RR: 0.82 (95% CI 0.76-0.88)	43,5%	
BMI (30<)	Per 5kg/m^2 increase	Premenopausal	RR: 0.93 (95% CI 0.90-0.97)	54,5%	WCRF (2017)
		Postmenopausal	RR: 1.12 (95% CI1.09-1.15)	73,6%	
Weight change since 18	Per 5 kg increase	Premenopausal	RR: 0.99 (95% CI 0.96-1.03)	13%	WCRF (2017)
		Postmenopausal	RR: 1.06 (95% CI 1.05-1.08)	38%	
Physical activity	Highest activity vs lowest			35.6%	Chen et al. (2019)
	Premenopausal		RR: 0.83 (95% CI 0.79-0.87)	23.3%	
	Postmenopausal		RR: 0.91 (95% CI 0.85.-0.97)		
Family history	Increasing number of affected relatives (1, 2 or 3+)		RR: 1.5-3.90	-	Loda et al. (2016)
Alcohol	Per 10 g/day	Premenopausal	RR: 1.05 (CI 95% 1.02-1.08)	-	WCRF (2017)
		Postmenopausal	RR: 1.09 (CI 95% 1.07-1.12)	-	
Age at menopause	Increase in risk per increase in year		RR: 1.03	-	Loda et al. (2016)
Mammographic density	BI-RADS D vs BI-RADS B All		OR: 2.11(95% CI: 1.84-2.42)	48%	Bodewes et al. (2022)
Birthweight	Per 500 gr	Premenopausal	RR: 1.05 (CI 95% 1.02-1.09)	-	WCRF (2017)
		Postmenopausal	RR: 1.00 (CI 95% 0.98-1.02)	-	

I^2 = Percentage of heterogeneity observed in the meta-analysis
RR = Risk ratios
OR = Odds ratios

Menopausal hormone therapy (MHT) and its association with breast cancer is dependent on duration of usage, and recency. Usage over time increased the risk of breast cancer, but the risk rapidly declines after subjects stop using MHT and evens out with never users after five years' time (Loda et al., 2016).

Premenopausal adiposity is associated with decreased risk for premenopausal cancer and postmenopausal adiposity with increased risk of breast cancer (Loda et al., 2016). With early life adiposity being associated with reduced risk of both premenopausal and postmenopausal breast cancer. The mechanism behind adiposity and premenopausal association is not well

known, it is however suggested that it might be related to fewer and more irregular menstrual cycles (Loda et al., 2016). Adipose tissue is the main producer of estrogen for postmenopausal women suggesting that increased adipose tissue, more circulation estrogen (Loda et al., 2016).

1.4 Determinants of Height:

As with most aspects of human development, growth is an intricate and complex process. These complexities make it challenging to study and predict adult stature on an individual level. However, on the scale of populations it is possible to observe trends of growth (Silventoinen, 2003).

Genetic factors is a main determinant of adult attained height in any given individual, leaving adult stature to highly heritable (Silventoinen, 2003). The main mechanic behind human growth has been greatly dependent on the prevalence of growth hormones (GH) and insulin like growth factors (IGF) in all stages of early life growth-periods (Rosenfeld, 2003). GH is believed to affect genes facilitating production of IGF-1, where increased levels of a specific IGF-1 has been linked to prenatal, childhood and adolescent growth periods (Rosenfeld, 2003) IGF is linked to increased mitosis and reducing apoptosis in tissue, facilitating growth.

Including genetics there are environmental factors that influence growth. These affect growth periods throughout prenatal stages until the end of puberty (Silventoinen, 2003). This influence is referred to as gene-environment interaction, where the genes respond to the stimuli from the environment. Environmental determinants of heights, refers to socioeconomic and living conditions in the population (Silventoinen, 2003). It has been observed differences in living conditions form one generation to the next, where the children grow to a taller than their parents (Silventoinen, 2003). Some of these variations have been linked to parental socioeconomic status, where increased socioeconomic status increased likelihood of access to necessary maternal healthcare, early-life nutrition, safe environment, and quality healthcare (Silventoinen, 2003).

The influenced observed by environmental factors is observed to be lower when observed in setting where children have greater food security and less prone to infectious diseases throughout their developmental stages (Silventoinen, 2003). Typically, the environmental determinants are nutrition, and infectious diseases, often linked to socioeconomic status.

1.5 Causal framework – Direct Acyclic Graphs.

DAGs, or Direct Acyclic Graphs is a causal framework that allows for the representation of the data generating process and illustrating the causal assumptions set by the investigator (Cinelli et al., 2022). Looking to figure 3, each node denotes a variable (exposure and outcome), and the arrow illustrates the causal path between exposure and outcome. This association is often referred to as a direct effect, meaning there are no third variable mediating or masking the effect the exposure has on the outcome.

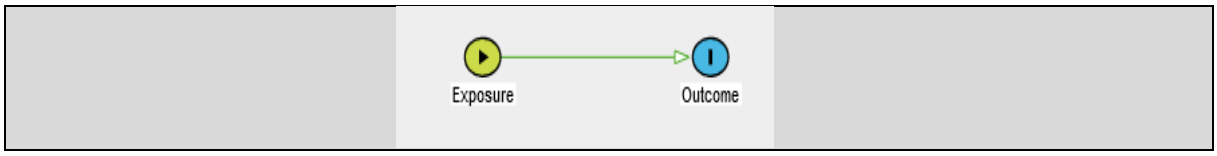


Figure 2 – Exposure-Outcome relationship. Illustrating direct effect. Created in Dagitty.

Cinelli et al. (2022) describes three central forms of association that further allows investigators to illustrate causal pathways and detected biasing patterns. One of these biasing patterns is confounders. Confounders can be described, as being causal of both exposure and outcome. Confounders are not descendant of either exposure or outcome, meaning they serve an independent influence on both variables. If this variable is not accounted for through study design or regression strategies, an association between exposure and outcome can occur, biasing the effect estimates.

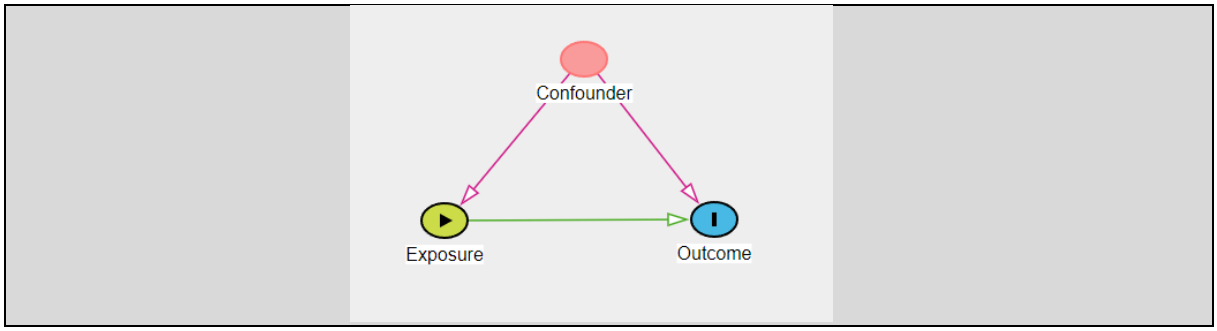


Figure 3 – Exposure-Confounder-Outcome relationship. Illustrating confounding. Created in Dagitty. Pink arrows indicate biasing pathways, green arrows indicate unbiased pathways.

Secondly, mediation refers to the causal relationship between an exposure and the outcome; that can be partly or fully explained by the mediator (figure 4). In contrast to a confounder, mediation is still part of the causal path. Not accounting for this, does not incur bias, but by

including mediator in regression analysis one needs to consider variables related to the mediator and exposures and outcome respectively.

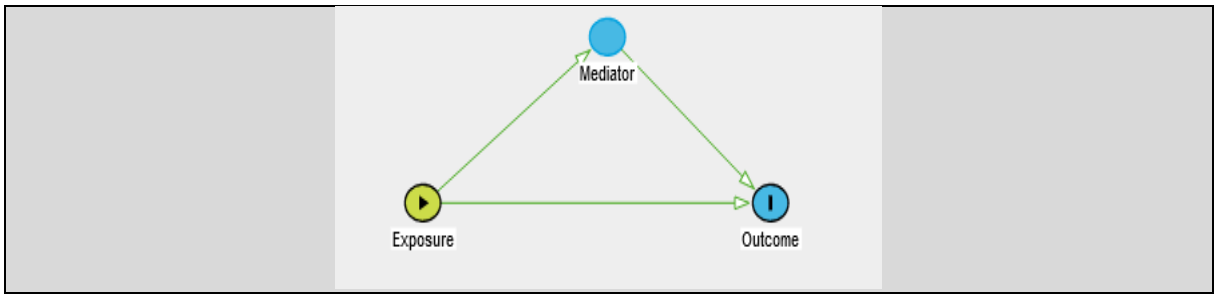


Figure 4 – Exposure-Mediator-Outcome relationship. Illustrating indirect effect. Created in Dagitty.

Thirdly, Cinelli et al. (2022) describes a collider association, here Z is being caused by X and Y. In this case, X and Y are independent of each other, but they are connected through Z. *Colliders* are important because by controlling for this factor using regression strategies, this will open an association between the X and Y. This kind of bias is referred to as collider bias.

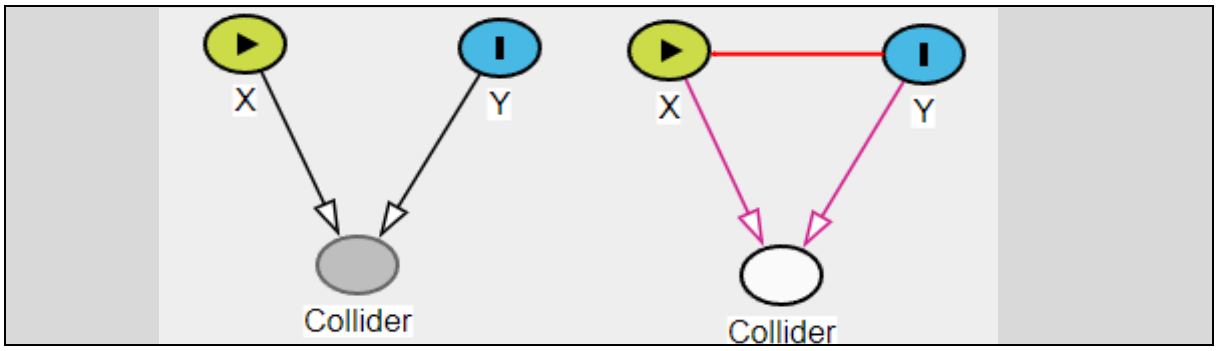


Figure 5 – Collider and Collider bias

1.6 Study Rational and Research question

The association between height and breast cancer has been studied extensively. Similar studies have been performed with the use of NOWAC data, through EPIC studies (Lahmann et al., 2004; Ritte et al., 2013). However, few studies have been performed on dedicated on Norwegian women in relation to height and breast cancer risk within the last decade, and even fewer have used causal framework for justifying inclusion of third variables.

Hence, this study aims to investigate the association between adult attained height and breast cancer in pre- and post-menopausal Norwegian women using a causal framework in the model building process illustrated through Directed Acyclic Graph (DAG).

Using a causal framework, we seek to establish subject-matter context to the statistical analysis, while communicating the assumptions transparently. With this approach we seek to alleviate the risk of overadjustment seen in many studies. We will investigate the influence of third variables on the model, and discuss the potential implications related to the direct and indirect hypothesis; describing the mechanism that may link adult attained height and breast cancer.

Primary objective:

Investigate the association between adult attained height and breast cancer incidence in Norwegian Women (Age: 30-70).

2 Methods: Material, Design and Analysis

2.1 Material – NOWAC Cohort Profile

2.1.1 Population, Recruitment, and sample

To answer the research question, and cover the objective stated in the introduction, we used data from the Norwegian Women and Cancer cohort study (NOWAC). A national prospective study containing over 172 000 Norwegian women aged 30-70 years.

The NOWAC cohort was randomly sampled from the Norwegian Central Person Register using an eleven-digit national personal identification number, unique to all citizens of Norway. This allows for linkage to all other national registries. Recruitment started in 1991, lasting until 2007, passive follow-up ongoing with its latest follow-up in 2020.

Potential participants were sent a study invitation containing general information regarding the study, including questionnaire with an integrated informed consent form. All candidates received a four-page questionnaire regarding: use of oral contraceptives (OC), use of hormonal replacement therapy (HRT), reproductive history, age at menarche (AAM) and menopause; smoking, physical activity (PA), alcohol consumption, anthropometric variables, socioeconomic status (SES), screening for breast cancer, family history of breast cancer, sunbathing habits and pigmentation and self-reported disease. Mailings after 1996 contained in addition to the original questionnaire, another four-page questionnaire regarding dietary habits(Lund et al., 2007). In the periods 1991-1997 and 2003-2006, 102 540 (57 % of invited) and 63 232 (48.4% of invited) responded to the study(Lund et al., 2007).

Active follow-up was performed through three different mailings with a six to seven-year interval between each mailing with an updated questionnaire(figure 7), while also having a passive follow-up by linkage to the Norwegian Cancer Registry and the Norwegian Cause of Death Registry(Lund et al., 2007). Further details regarding the cohort can be found on NOWACS webpage: https://uit.no/research/nowac.

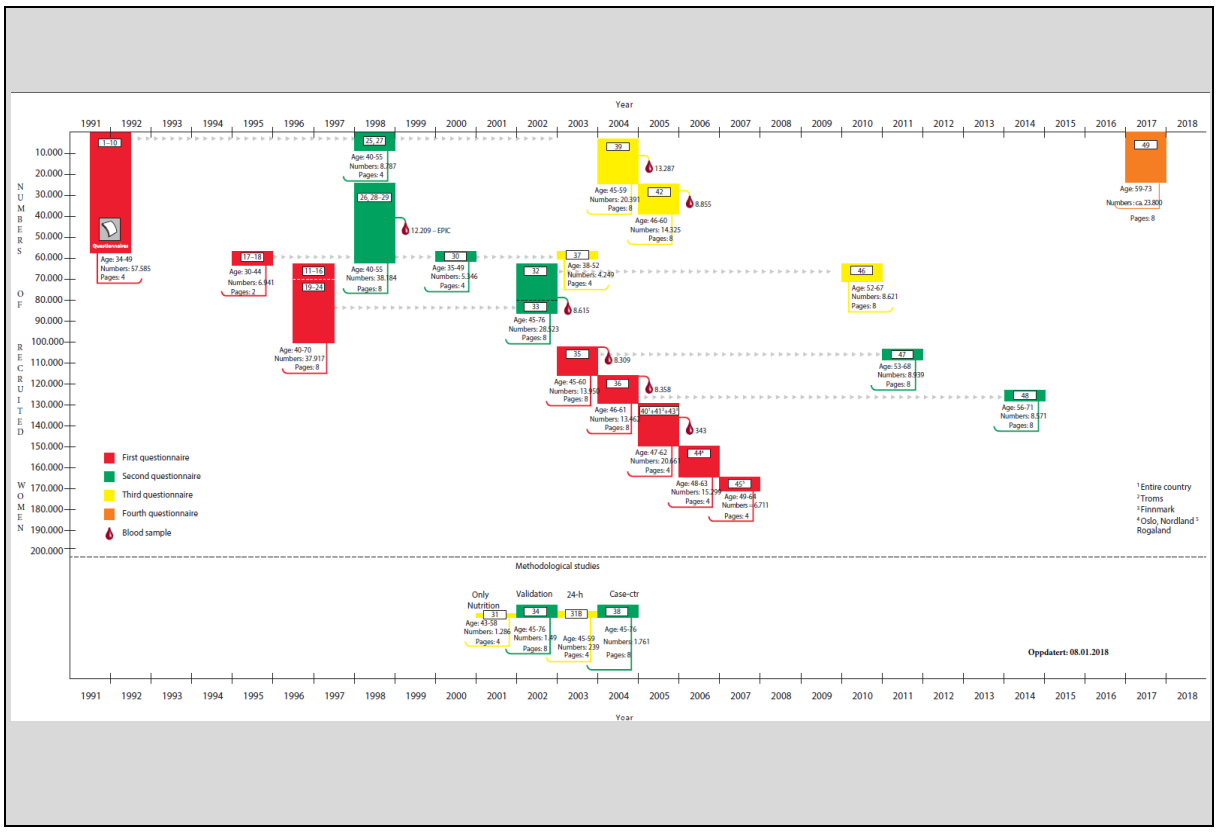


Figure 6 - Overview of NOWAC mailing series and participant distribution (2018).

2.1.2 Inclusion/exclusion criteria

Information from all first-time surveys were included into the study as they all contained information on height and linked to the Norwegian Cancer Registry. The NOWAC cohort had information of adult height for 170 428 women and had registered 11 582 breast cancer cases in total.

Participants that had prevalent cases (n= 6 570), BMI of 40 or below 14 (n=17), Alcohol intake exceeding 150g g/day was excluded (n= 626). Finally, all women not reporting their current height at baseline (n= 2 044), was excluded. The total analytic sample totaled 163 215 participants, and 8 872 cases.

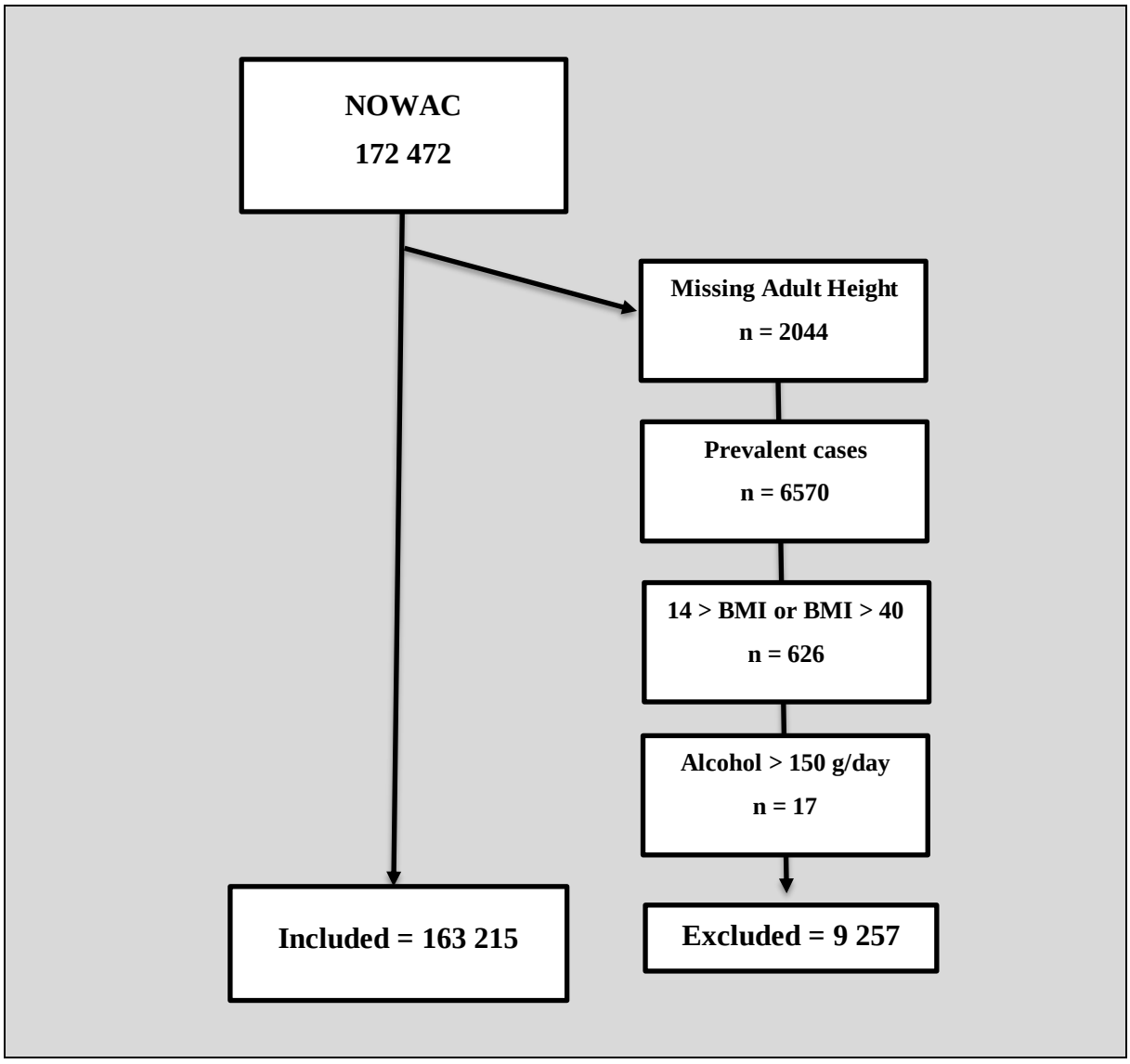


Figure 7 - Exclusion/Inclusion flow-chart.

2.2 Variables:

Variables included into the NOWAC dataset is outlined in Appendix 1.

2.2.1 Exposure

Adult attained height was measured through self-report questionnaire, the participants were asked "How tall are you (in cm)?", participants were then asked to fill this in as an integer. For the purposes of the statistical analysis, height at baseline was coded into a 10cm increment continuous variable.

2.2.2 Outcome

The outcome variable “Breast Cancer Diagnosis” was based on the cancer diagnosis classification provided by the Norwegian Cancer Registry. The diagnosis was sorted in four categories, C50, C500, C509 and NA, based on the ICD-10 classification system based on location of the tumor. C50 is specified to lesions in in the nipple and areola of the breast. C.509 is classified as an unspecified lesion in the breast which could be contained any quadrant or depth of the breast. The mentioned classification includes the connective tissue of the breast and mammary glands but excludes the skin.

The original IOC-150 classification from the Norwegian Cancer Registry was recoded into a binary variable, with “1” indicating diagnosis, and “0” indicating no diagnosis, treating all NAs in the original classification as a “0”.

The full data cleaning and coding procedure can be observed Appendix 10.

2.3 Modelbuilding and Statistical Methods

2.3.1 Modelbuilding: Direct Acyclic Graphs

Third variables for the model were decided by using Directed Acyclic Graphs (DAGs) informed by current literature as to argue for the inclusion/exclusion of third variables into the total effects model proper.

Literature searches were performed to create the initial map of a causal network of height and breast cancer development (appendix 2 & appendix 3). The literature used to justify the DAG is discussed in chapter 4.3. The DAG was simplified into figure 9, using Huntington-Kleins(Wang, 2023) approach to DAGs construction, which reduced the scope of the DAG to include the variables which we have available in the NOWAC dataset and variables had some plausible causal connection to both height and breast cancer. The exceptions to this are the variables for age, as it is assumed to be associated with all variables listed, and the unobserved variables parental SES and genetic factors, which is included on the basis of connecting variables in a cohesive manner and enable hypothetical causal pathways.

The age-adjusted model included the age at baseline. The total effect multivariate model included the covariables: age at menarche, childhood body shape, breast cancer in mother and years of education.

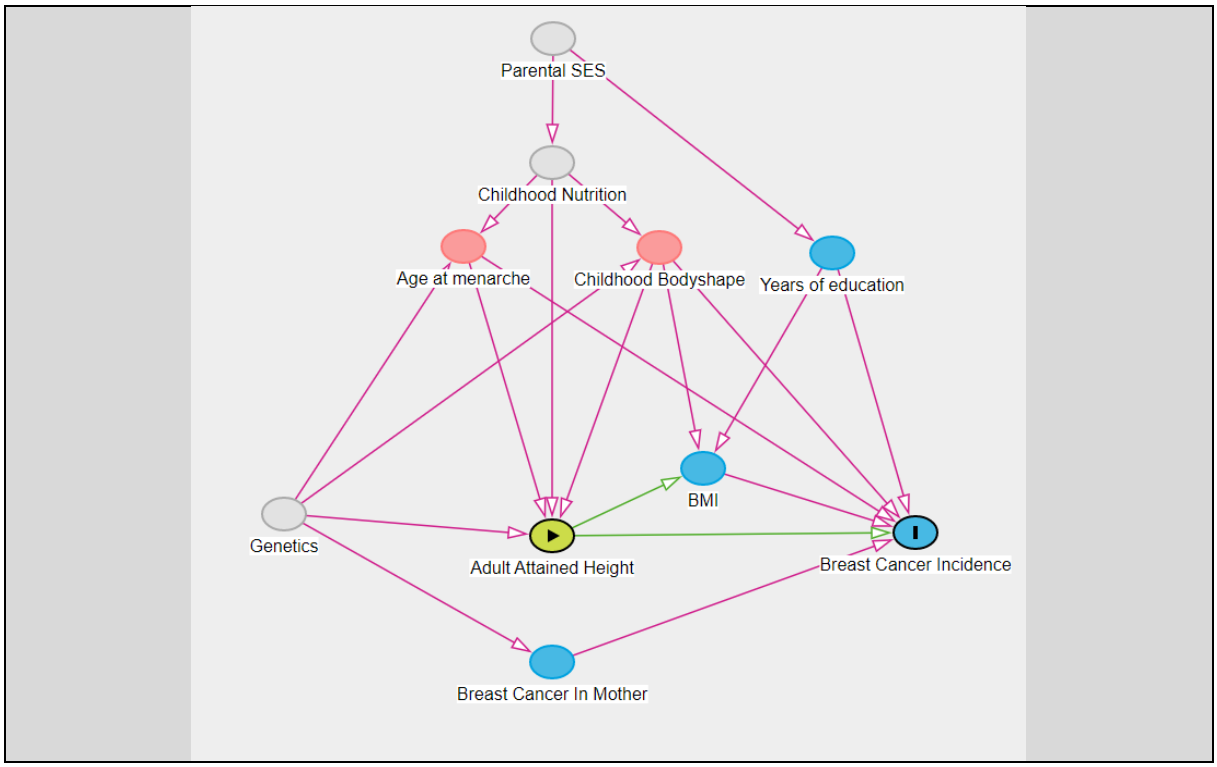


Figure 8 – DAG. Pink arrows indicate biasing pathways. Green arrows indicate unbiased pathways. Red nodes indicate a confounding variable. Blue nodes indicate ancestors of the outcome. Grey nodes indicate unobserved variables. Generated in Dagitty, R-code in Appendix 5.

2.3.2 Cox Proportional Hazards Regression

For descriptive analysis I will be presenting results from all variables shown in table 2. For estimating the association, we used multivariable cox proportional hazards regression model to calculate age-adjusted, and multivariable hazards ratios for developing breast cancer, stratified by menopausal status. The main exposure is height, expressed in 10 cm increments. We further expanded the main exposure by fitting restricted cubic splines to detect potential non-linear trends between height and breast cancer risk. For the splines, four knots were placed at the 5^{th}, 35^{th}, 65^{th}, and 95^{th} percentiles of the height variable. The proportional hazard assumption was checked using Martingale residual plots, and no violations were found.

2.3.3 Missing Data and imputation.

Under complete case-analysis on the all-breast cancer analysis there were 30 302 participants missing, for stratified analysis premenopausal and postmenopausal strata had 13 830 and 15 628 missing, respectively. To recover the loss of statistical power due to observations with missing values, multiple imputation was performed. Given the assumption that this data was missing at random (MAR) we performed multiple imputation to replace missing values in

included covariables. Fifteen datasets were imputed and results pooled using Rubin's rules (Harrel, 2015) for the Cox-Regression analysis to calculate hazard ratios and 95% confidence intervals.

For the imputation model, age at baseline, age at end of follow up, age at menarche, childhood body shape, height, breast cancer status mother, alcohol intake, smoke status, education years, use of oral contraceptives, use of hormonal replacement therapy, current physical activity level, total number of children, age at first child and binary menopausal status were included regardless of its inclusion in the model proper.

Based on recommendation described in (Heymans & Eekhout, 2019) All variables included in the imputation model was part of the analysis model or that were believed to be related to missingness. Variables was excluded if they were believed to be highly correlated or had high levels (>50%) of NA/unknown. All inclusion and exclusion criteria described in chapter 2.1.2., were implemented before the imputation.

For the data management, analysis, and multiple imputations the statistical programing language R through the RStudio software was utilized. Details of the imputation are available in the appendix 10.

2.3.4 Ethical considerations

All women participating in the study have given their informed consent for their submitted information to be used in research and the linkage to the cancer registry and the cause of death registry. Information regarding data handling and intent of utilization of the data was described in the consent form participants received upon invitation(Lund et al., 2007).

The NOWAC study has been approved by the Regional Committee for Medical Research Ethics (Lund et al., 2007). All data gathered by NOWAC is stored without the participants name or personal identification number. The linkage from the gathered data is handled by the Norwegian Statistical Bureau through a unique code given to each participant (NOWAC, 2022).

Approval for the study has been granted by the Regional Committee for Medical Research Ethics. Approval for access to data has been granted by NOWAC.

3 Results

3.1 Descriptive statistics

After exclusion 154 343 participants were included in the analysis, throughout the follow-up, 8 872 participants experienced the event of interest.

The mean height of the sample is 166.27, where a slight difference is observed between cases (m=166.23cm) and non-cases (166.84 cm). The mean age at baseline where 49.29, with a minor difference between cases and non-cases. BMI mean for the samples is estimated to be 24.20, with no statistically significant difference between the groups, and most individuals, 60.4%, reported their childhood body shape to have been "normal".

The participants in our sample had a slightly greater proportion of postmenopausal women (49.4%). 90% reported to have at least one child, where most participants (54.9%) reported to have given birth of their first child before the age of 25. When compared against participants experiencing an event versus non-event, the larger proportion of participants with a cancer diagnosis tended towards having births later than earlier and having fewer children.

23.9% of participants reported to ever have used hormonal replacement therapy. Diagnosed participants tended towards having a slightly higher proportion of ever and current users of HRT. Further, 55.1% claimed ever use of oral contraceptives, no difference of note between the groups.

On lifestyle variables, 34.2% of participants claim to have never been smokers, while former smokers and current smokers are reported to be 33.8% and 30.3%, respectively. A slight difference between the groups was observed, where cases had a larger proportion of former smokers.

Table 2 – Descriptive table of the NOWAC sample (n = 163 215) by cancer diagnosis status, complete case analysis.

Stratified by cancer diagnosis		**All**	**No (n = 154 343)**	**Yes (n = 8 872)**	
Mean (SD)					**P-value**
Standing height in cm		166.27 (5.65)	166.23 (5.65)	166.84 (5.60)	<0.001
Age at baseline		49.29 (8.45)	49.33 (8.47)	48.66 (8.16)	<0.001
BMI		24.20 (3.9)	24.20 (3.74)	24.14 (3.67)	0.135
Number (%)					
Menopausal status	Premenopausal	70 157 (45.6)	65 853 (42.7)	4 304 (48.5)	<0.001

	Postmenopausal	89 449	(52.3)	85 065	(55.1)	4 383 (49.4)	
	NA	3 610	(2.2)	3550	(2.2)	185 (2.1)	
Age at first live birth	<20	20 041	(12.3)	19 964	(12.4)	994 (11.2)	<0.001
	20-24	69 506	(42.6)	65 949	(42.7)	3 557 (40.1)	
	25-29	41 093	(25.1)	38 817	(25.0)	2 276 (25.8)	
	>30	16 470	(10.1)	15 421	(9.8)	1 049 (11.8)	
	NA	16 105	(9.9)	15 109	(9.8)	996 (11.2)	
Total number of children	0	16096	(9.9)	15 101	(9.8)	995 (11.2)	<0.001
	1	19 559	(12.0)	18 356	(11.9)	1203 (13.6)	
	2	67 965	(41.6)	64 176	(41.4)	3 789 (42.7)	
	3	41 950	(25.6)	39 853	(25.8)	2 097 (23.6)	
	4+	17 645	(10.8)	16 857	(10.9)	788 (8.9)	
Ever use of (HRT)	Yes	38 945	(23.9)	36 675	(23.8)	2270 (25.6)	<0.001
	No	134 009	(70.5)	109 027	(70.4)	6 088 (68.6)	
	NA	9 155	(5. 6)	8 641	(5.5)	514 (5.8)	
Ever use of Oral Contraceptives	Yes	89 993	(55.1)	84 856	(55.0)	5077 (57.2)	<0.001
	No	67 638	(41.8)	64 123	(41.5)	3512 (39.6)	
	NA	5644	(3.7)	5 892	(3.7)	414 (3.6)	
Breast Cancer Mother	Yes	10 027	(6.1)	9 271	(6.0)	756 (8.5)	<0.001
	No	143 259	(87.8)	135 729	(87.9)	7 530 (84.9)	
	NA	12607	(6.1)	9 343	(7.2)	586 (6.6)	
Body shape in childhood	Very Thin	7 587	(4.6)	7 084	(7.2)	661 (7.5)	<0.001
	Thin	28 272	(17.3)	26 627	(17.3)	1645 (18.5)	
	Normal	98 504	(60.4)	93 285	(60.4)	5 219 (58.8)	
	Fat	16 534	(10.1)	15 781	(10.2)	816 (9.2)	
	Very fat	554	(0.3)	526	(0.3)	28 (0.3)	
	NA	11 764	(7.2)	11892	(7.4)	802 (6.9)	
Self-reported physical activity	Low	18 472	(11.3)	17 393	(11.3)	1 079 (12.2)	<0.001
	Middle	80 157	(49.1)	75 590	(49.0)	4 567 (51.5)	
	High	46 940	(28.8)	44 582	(28.9)	2 358 (26.6)	
	NA	17 646	(10.8)	16 778	(10.9)	868 (9.8)	
Number of years in Education	<=9	35 152	(21.8)	33 413	(21.6)	1 739 (19.6)	<0.001
	10-12	53 040	(32.5)	50 130	(32.5)	2 910 (32.8)	
	13-16	42 877	(26.3)	40 428	(26.2)	2 449 (27.6)	
	>=17	23 032	(14.1)	21 745	(14.1)	1 287 (14.5)	
	NA	914	(5.6)	8 627	(5.6)	487 (5.5)	
Smoking status	Never	55 861	(34.2)	52 944	(34.3)	2 917 (32.9)	<0.001
	Former	55 246	(33.8)	52 189	(33.8)	3 057 (36.1)	
	Current	49 995	(30.3)	47 202	(30.6)	2 793 (31.5)	
	NA	2488	(1.4)	2 008	(1.5)	105 (1.2)	

3.2 Cox Proportional Hazards regression

Overall breast cancer risk was modeled using Cox proportional hazard regression, using attained-age as the time variable. Statistical significance was defined as $p < 0.05$. For the purposes of sensitivity analysis, models using both complete-case data and multiply imputed data were run, although no difference in hazard ratios exceeding 0.02 was observed between the two sets of data. Overall breast cancer risk was observed in the age-adjusted models to have a 23% increased risk of overall breast cancer per 10 cm increment, versus 21 % in premenopausal and 26 % in postmenopausal sub-cohorts.

Table 3 *– Cox Proportional Hazards Regression Results: hazards ratios (95% CI) of overall breast cancer, premenopausal breast cancer and postmenopausal breast cancer in relation to height in ten cm increments among NOWAC participants (n = 163 215).*

Complete case	**HR - Age-adjusted (95% CI)**	**HR - Multivariable (95% CI)[a]**
All breast cancer, n = 7 161	1.23 (1.19-1.28, p< 0.001)	1.21 (1.16-1.26, p< 0.001)
Premenopausal, n = 3 543	1.21 (1.15-1.28, p< 0.001)	1.21 (1.14-1.28, p< 0.001)
Postmenopausal, n = 3 470	1.26 (1.19-1.32, p< 0.001)	1.22 (1.15-1.30, p< 0.001)
Multiple imputation	**HR - Age-adjusted (95% CI)**	**HR - Multivariable (95% CI)[a]**
All breast cancer, n = 8 860	1.23 (1.19-1.27, p< 0.001)	1.21 (1.17-1.25, p< 0.001)
Premenopausal, n = 4 426	1.21 (1.16-1.26, p< 0.001)	1.20 (1.14-1.26, p< 0.001)
Postmenopausal, n = 4 434	1.26 (1.21-1.31, p< 0.001)	1.22 (1.16-1.28, p< 0.001)

[a] HR adjusted for: age at menarche(cont.), childhood body shape(cat.), years of education(cat.) & breast cancer in mother(cat.).

CI = confidence interval, cat. = categorical variable, cont. = continuous variable

The results from the multivariable models showed an increased risk of 21% for overall breast cancer, versus 20% and 22% for the premenopausal and postmenopausal sub-cohorts, respectively. As can be observed from table 3, the multivariable models exhibit notably smaller differences in hazard ratio between the pre- and postmenopausal sub-cohorts.

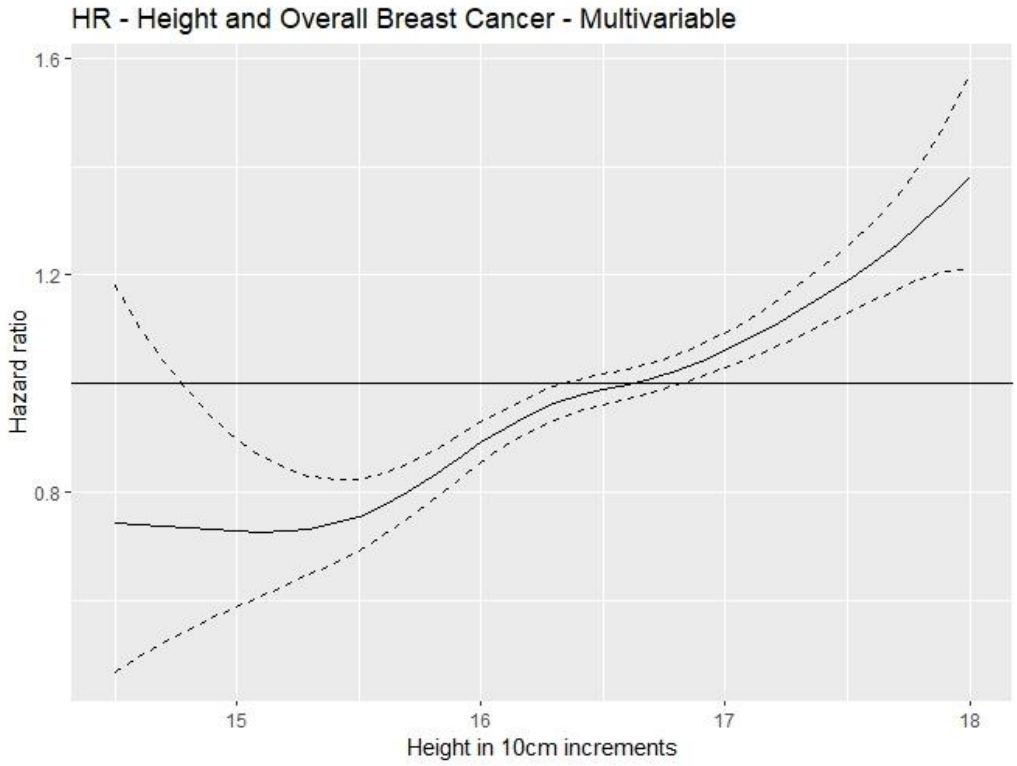


Figure 9 – Spline multivariable regression model for height and overall breast cancer risk. Solid line: HR, dashed lines: 95% CI.

Restricted cubic spline plot did not indicate any non-nonlinear behavior between height and breast cancer risk overall. No deviation from linearity was observed in premenopausal and postmenopausal stratifications. As a result, splines were not used to model height in the final analysis.

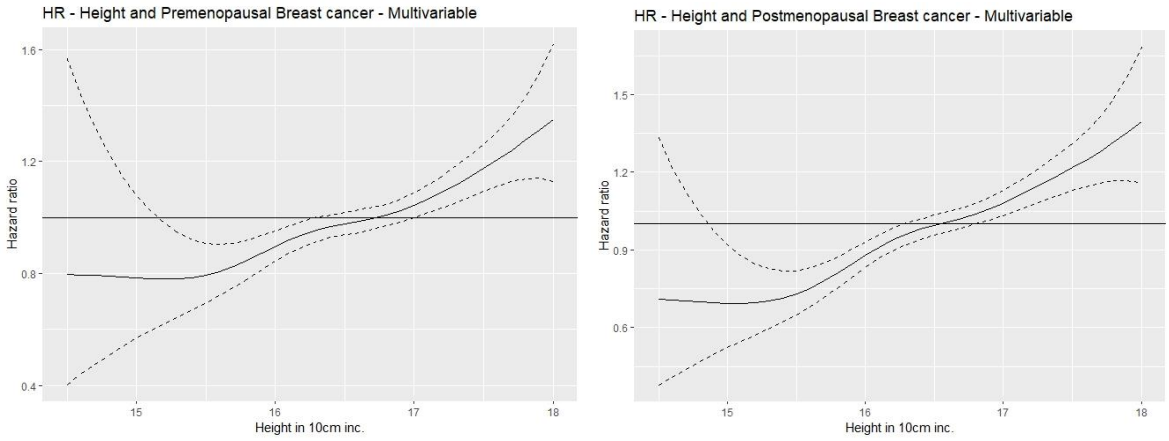


Figure 10 – Spline multivariable regression models for height and premenopausal(left) and postmenopausal(right) breast cancer risk. Solid line: HR, dashed lines: 95% CI.

4 Discussion:

The multivariable imputed regression analysis indicated association between every 10-centimeter increase in height and the risk of breast cancer, both in premenopausal and postmenopausal women and overall breast cancer risk. The sensitivity analysis provided evidence for breast cancer in mother, age at menarche, years of education did not change the height-breast cancer association to the predetermined threshold (10%). However, years of education did change the HR by13 % for postmenopausal breast cancer.

This is indicating a positive association between height and breast cancer incidence. However, this relationship seems to be expressed differently across model types. In the age-adjusted model we observe a difference in HR when comparing premenopausal (HR = 1.21(1.15-1.28)) and postmenopausal (HR = 1.26(1.19-1.32)) participants. This difference between stratifications flattens when observing the multivariable model.

4.1 Comparison to other results:

Choi et al. (2019) conducted a study with over 22 million participants linking several cancer types to height, including breast cancer, marking it to be the largest sample size to date, used to investigate the height and cancer association (Choi et al., 2019). The study found an association between height and breast cancer, like our findings (HR: 1.16, 95% CI: 1.15-1.17). On the other hand, the follow-up was relatively short. It had only a five-year follow up time reducing the ability to establish temporality between exposures and a slow developing outcome (prevalent cases was excluded), whereas our study did provide a longer follow-up, passing beyond 20 years.

Van den Brandt et al. (2021) investigated through a pooled analysis of 20 cohort studies the association between anthropometric variables, among them height, and the risk of various classifications of breast cancer. Van den Brandt et al. (2021), observed risk ratios for premenopausal (1.07, 95% CI: 1.04-1.10) and postmenopausal women (1.07, 95% CI: 1.04-1.10), that reflects our findings of a positive relationship between height and overall breast cancer. The authors performed a restricted cubic spline regression analysis, observing a linear trend for the association (*p* value for test of non-linearity > 0.14) with no statistically significant heterogeneity across cohorts (Van den Brandt et al., 2021).

(Ritte et al., 2013) There have not been any dedicated studies on height and breast cancer risk in NOWAC, however subsamples from the NOWAC cohort (n = 35 311) have been used as

part of the EPIC cohort(n = 306 600) (Ritte et al., 2013) investigated the association between height, age at menarche and risk of hormonal-receptor-stratified breast cancer. An association between height and breast cancer (HR 1.13, 95% CI: 1.10-1.17) was found.

Horn et al. (2014) performed a cohort study among Norwegian postmenopausal women (n = 18 562, breast cancer cases = 969), finding an increased risk of 8% for breast cancer per 5 cm increment (HR 1.08, 95% CI: 1.02-1.14). An earlier Norwegian study comparing the shortest (< 162cm) and tallest women(>167cm) found an increased risk (RR: 2.5, 95% CI 1.2-5.5) for taller women's development of breast cancer. (Nilsen & Vatten, 2001)

In contrast to our analysis, the studies mentioned above did not adjust for confounders between height and breast cancer, other than age. Our study is notable in that it has more participants and a longer study duration in comparison to the Norwegian cohort studies mentioned above, although several international studies have larger cohorts. Among these studies, hypotheses regarding the link between height and breast cancer have been divided into two groupings: direct effect and indirect effect (Giovannucci, 2019).

4.2 Direct Effect Hypothesis

The direct effect hypothesis (figure 11) suggests that adult attained height is related to organ size (Giovannucci, 2019; Nunney, 2018), as the epithelial tissue lining the organs has a rapid cellular turnover rate and thus there are more stem cells that can provide selfish-lineage mutations (Nunney, 1999, 2018). The hypothesis presumes a causal link between all variables and assumes the absence of confounding. In summary, the hypothesis proposes a direct unbroken causal path from height to breast cancer risk (figure 11).
The research of Ward et al. (2022), which found an association between height and increased mammographic total dense area (Ward et al., 2022), supports the direct effect hypothesis, as mammary gland size increased with height.

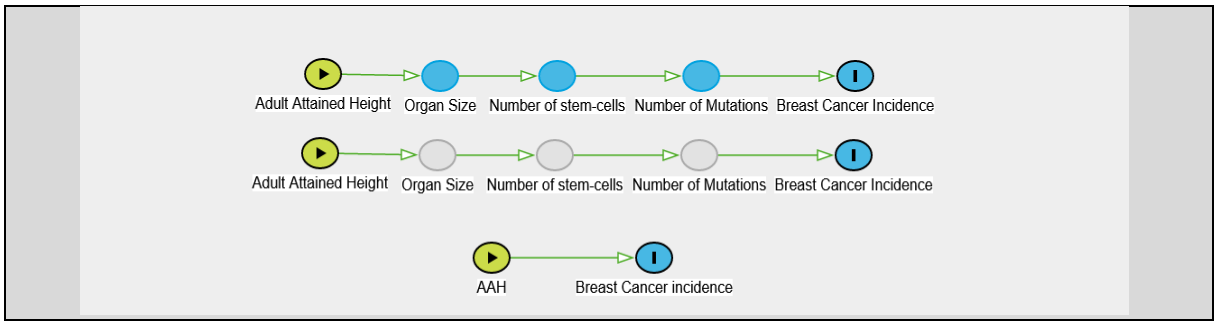


Figure 11 – Causal Path between Adult Attained Height and Breast Cancer

4.3 Indirect Effect Hypothesis

In contrast, the indirect effect hypothesis suggests that there are other causal factors either causing(confounding) or mediating the height-breast cancer relationship (Giovannucci, 2019; Michels & Willett, 2004). Most of the common risk factors are related to early-life growth periods.

Serum IGF-1 concentration is one of the variables that is often proposed as having a core physiologic role in both height ((Rosenfeld, 2003)), and breast cancer development ((Biro et al., 2021; Rosenfeld, 2003)). Serum levels of several growth hormones, including IGF-1, increase substantially during periods of growth. One of the primary actions of IGF-1 during these periods, is to promote cell proliferation. As such, this opens "windows of susceptibility" in which an increase in mitosis together with heightened suppression of apoptosis increases the risk of genetic instability and cancer-promoting mutations (Biro et al., 2021). However, a 2020 study by Parra-soto et al. of 414,923 subjects in the UK biobank did not find any mediating effect of adult serum levels of IGF-1 on the association between height and cancer incidence, suggesting that any mediating effect of IGF-1 may be limited to early growth periods. A mendelian randomization study linked common genetic factors and biological pathways between height and breast cancer independently of IGF-1, indicating that IGF-1 is not the sole explanatory factor for the height-breast cancer association (Parra-Soto et al., 2020; Zhang et al., 2015).

In our analysis, we attempted to account for indirect effect variables described by the literature as potential confounders of height and breast cancer (Figure 9). We performed sensitivity analyses to quantify the influence each variable had on the overall model. While we found that years of education decreased the association for postmenopausal estimates by 13% (Appendix 6), we observed limited (<10%) change for age at menarche, childhood body shape and breast cancer in mother.

Years of education serves as a proxy for the socioeconomic status, assuming that parental SES is predictive of educational level(Sirin, 2005). Socioeconomic status might be reflective of better childhood nutrition and living conditions potentially contributing to increased stature. A recent Norwegian descriptive study (Arntsen et al., 2023) investigated the growth trends for height and educational levels across 10-year cohorts (born 1930-1977) and found that height within cohort has been correlated to educational status. However, differences in height between educational levels decreased with more recent cohorts, suggesting that

increased standards of living could be central contributor to this trend (Arntsen et al., 2023). Consequently, the effect of years of education (as a proxy for socioeconomic status) on height and thus breast cancer risk may be minimized in developed countries due to an overall higher standard of living.

In our study, body shape in childhood can reflect childhood adiposity. Increased childhood adiposity has been linked to reduced incidence of pre- and postmenopausal breast cancer (Byun et al., 2022), younger age at menarche (Cousminer et al., 2013; Li et al., 2017), and increased adult stature (Cousminer et al., 2013). Byun et al. (2022) points out that this inverse relationship between early life adiposity and breast cancer is unique compared to other female oriented cancers (endometrial and ovarian) and that this is possibly independent of adult adiposity, suggesting a unique mechanism for childhood adiposity (Byun et al., 2022).

Age at menarche has been consistently observed in the literature as having an inverse association with overall breast cancer risk and positively with height (Osuch et al., 2010). An EPIC study (Ritte et al., 2013) on breast cancer risk found an interaction between age at menarche and height, and observed that tall women with early menarche, had a higher risk of breast cancer then tall women with a late menarche. Schoemaker et al. (2017) suggest with their observational study that age at menarche might be positively related to adult mammographic density and density area, further supported by Ward et al. (2022), describing a positive relationship between mammographic density, age at menarche, and adult attained height. A possible explanation for the association might be that breast development could be reliant on the length of the interval, rather than timing of onset between thelarche and menarche; where both are positively associated with increased adult mammographic density (Ghadge et al., 2021). Though this association, to the extent of the authors knowledge, has not yet been investigated in relation to adult attained height.

Genetic risk factors have been extensively related breast cancers through, mainly BRCA1 and BRCA2. This makes breast cancer in mother a strong predictor for breast cancer risk in their daughters. However, mendelian randomization combined with a meta-analysis identified several of the genetic predictors of height, and genetic predictors for breast cancer was shared (Zhang et al., 2015), suggesting that height might mediate some of the risk observed in between hereditary breast cancer.

4.4 Strengths and limitations

The current study offered a large and representative sample of Norwegian women, paired with a long follow-up and outcomes were delivered by national high-quality registries. For covariate selection DAGs were used to identify potential confounding associations. Main limiting factors is misclassification derived from questionnaire data, healthy participant bias and the indirect nature of the covariates.

The sample is large and with the exclusion criteria, over 162 000 participants and were included in the analysis. This provides substantial statistical power, instilling greater confidence in our regression estimates. The NOWAC cohort sampled randomly throughout the Norwegian population using women's individual 11-digit social identification numbers registered in the Norwegian population registry. Lund et al. (2007) found that participants have a tendency to be highly educated, from the north of Norway and tend to be younger (Lund et al., 2007). However, the sample overall is reported to be representative of the Norwegian populations (Lund et al., 2007; Lund et al., 2003), and the distribution of exposures was not significantly affected by response rates.(Lund et al., 2007). First-time mailings response rates were for the periods 1991-1997 and 2003- 2006, 57% and 48.4% respectively (Lund et al., 2007). Lund et al. (2003) investigated the responders of 1991-1997 mailings and found that the reported education and parity fitted with registry data and lifestyle-factors did not differ significantly from non-responders(Lund et al., 2003). Our outcome, breast cancer incidence, was derived from high-quality registry data from the Norwegian Cancer Registry. The study design allows for establishing of temporality, as we follow subjects over a period using passive follow-up. The current study provides a long follow-up period (26 years at the most) which is crucial for the detection of cancer development. Temporality aspect allows us to discuss the direction of the association, and the length of the follow-up provides the analysis with more statistical power through the accumulation of events.

A further strength to the study is the use of DAGs in its model building stage. DAGs have the potential for reducing bias as they provide explicit causal context for statistical testing. Sensitivity analysis is a powerful tool in establishing an influence of a covariate on the relationship between exposure and outcome. Finally, explicit presentation of DAGs in scientific work allows for easier reproducibility of the stated hypothesis, and clarity regarding the nature of the model assumptions.

This study has limitations increasing the risk of bias stemming both from the design and within the analysis. All information on exposure and covariables were collected via a self-administered questionnaire, making them vulnerable to misclassification as participants might under or overestimate the true value. Further, variables such as body shape at first grade is vulnerable to recall bias.

Our study is based on self-reported prospective data, being a voluntary effort there are dangers of selection bias through healthy participant bias. We can assume that a healthy individual is more willing to contribute to a study than a severely ill individual. Exposure and outcome variables were gathered independently, but a strong limitation to the current study is that all covariables are only measured at baseline, meaning the stratifying variable menopause will be potentially biased as women that originally answered premenopausal, might reach menopause during follow-up. Because of unrepresentative distribution of hormonal receptor status due to potential error in the dataset, we were unable to investigate the height and breast cancer by hormonal receptor status. When it comes to our overall model, lacking information on childhood information like nutrition and socioeconomic conditions opens us up for potential confounders that we were unable to control for. The intention to model de novo cancer incidence limited statistical power due to the exclusion of subjects with prevalent cancer.

4.5 Future Implications:

Our study adds on to the overall literature, where we observe an increased risk of breast cancer associated with increased height, and a linear trend between height and breast cancer risk. While we also adjust for relevant confounding variables. However, our study is limited to investigating total effects models, based on available variables from the NOWAC study.

Future studies should emphasize further looking into direct effect models by investigating potential mediation of height and mammographic density to further our understanding of direct effect from height and breast cancer risk. R

Regarding the indirect effect hypothesis to further our understanding of the indirect effect hypothesis; breast cancer risk and early life factors need increased focus, as information on growths periods and childhood nutrition are important factors suggested by the unique roles age at menarche and childhood adiposity has on breast cancer risk and their association with height. Our understanding of IGF-1 and growth hormones needs to be further understood by

long term prospective studies that allows observation through growth periods, leading into post-growth periods and adult life.

Finally, all subject matters should consider using causal framework for their development of statistical models. The further investigation into the height and breast cancer association could be important for our understanding of tumorigenesis and improve screening methods by understanding the causal link from anthropometric variables to breast cancer risk.

5 Conclusion:

Our study looked at the association between height and breast cancer risk. Our analysis found that adult attained height is linearly associated with overall breast cancer incidence (HR: 1.22 (CI 95%:1.17-1.25), an association that is maintained, though modified, after stratification by menopausal status. Evaluation of potential confounders showed that adjustment years of education had the largest effect on hazard estimates in postmenopausal strata (13%), while age at menarche, childhood body shape, and breast cancer in mother were limited in their impact.

Height and breast cancer association is a well-documented topic, one for which the etiology remains disputed. Our findings contribute to the cumulative results of previous observational studies, and our study findings are consistent with these earlier findings. The most noteworthy feature of this study, and its most significant contribution, is the causal framework being used to critically appraise included variables. This framework allows for the combination of subject matter knowledge and statistical methods, and transparently presents causal thinking and its fundamental assumptions. As with many epidemiological studies, the indirect nature of the variables involved limits the case for causal inference to the theoretical.

References:

Arntsen, S. H., Borch, K. B., Wilsgaard, T., Njølstad, I., & Hansen, A. H. (2023). Time trends in body height according to educational level. A descriptive study from the Tromsø Study 1979–2016. *PLOS ONE*, *18*(1), e0279965. https://doi.org/10.1371/journal.pone.0279965

Aune, D., Navarro Rosenblatt, D. A., Chan, D. S. M., Abar, L., Vingeliene, S., Vieira, A. R., Greenwood, D. C., & Norat, T. (2015). Anthropometric factors and ovarian cancer risk: A systematic review and nonlinear dose-response meta-analysis of prospective studies. *International journal of cancer*, *136*(8), 1888-1898. https://doi.org/https://doi.org/10.1002/ijc.29207

Biro, F. M., Huang, B., Wasserman, H., Gordon, C. M., & Pinney, S. M. (2021). Pubertal growth, IGF-1, and windows of susceptibility: puberty and future breast cancer risk. *Journal of Adolescent Health*, *68*(3), 517-522.

Bodewes, F. T. H., van Asselt, A. A., Dorrius, M. D., Greuter, M. J. W., & de Bock, G. H. (2022). Mammographic breast density and the risk of breast cancer: A systematic review and meta-analysis. *The Breast*, *66*, 62-68. https://doi.org/https://doi.org/10.1016/j.breast.2022.09.007

Byun, D., Hong, S., Ryu, S., Nam, Y., Jang, H., Cho, Y., Keum, N., & Oh, H. (2022). Early-life body mass index and risks of breast, endometrial, and ovarian cancers: a dose–response meta-analysis of prospective studies. *British Journal of Cancer*, *126*(4), 664-672. https://doi.org/10.1038/s41416-021-01625-1

Chen, X., Wang, Q., Zhang, Y., Xie, Q., & Tan, X. (2019). Physical Activity and Risk of Breast Cancer: A Meta-Analysis of 38 Cohort Studies in 45 Study Reports. *Value in Health*, *22*(1), 104-128. https://doi.org/https://doi.org/10.1016/j.jval.2018.06.020

Choi, Y. J., Lee, D. H., Han, K.-D., Yoon, H., Shin, C. M., Park, Y. S., & Kim, N. (2019). Adult height in relation to risk of cancer in a cohort of 22,809,722 Korean adults. *British Journal of Cancer*, *120*(6), 668-674. https://doi.org/10.1038/s41416-018-0371-8

Cinelli, C., Forney, A., & Pearl, J. (2022). A Crash Course in Good and Bad Controls. *Sociological Methods & Research*, *0*(0), 00491241221099552. https://doi.org/10.1177/00491241221099552

Cole, T. J. (2003). The secular trend in human physical growth: a biological view. *Economics & Human Biology*, *1*(2), 161-168.

Cousminer, D. L., Berry, D. J., Timpson, N. J., Ang, W., Thiering, E., Byrne, E. M., Taal, H. R., Huikari, V., Bradfield, J. P., Kerkhof, M., Groen-Blokhuis, M. M., Kreiner-Møller, E., Marinelli, M., Holst, C., Leinonen, J. T., Perry, J. R. B., Surakka, I., Pietiläinen, O., Kettunen, J., . . . Consortium, f. t. E. G. G. (2013). Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. *Human Molecular Genetics*, *22*(13), 2735-2747. https://doi.org/10.1093/hmg/ddt104

CRN. (2021). *Cancer in Norway*. Cancer Registry of Norway. https://www.kreftregisteret.no/globalassets/cancer-in-norway/2021/cin_report.pdf

CRN. (2023). *BreastScreen Norway*. Cancer Registry of Norway https://www.kreftregisteret.no/en/screening/BreastScreen_Norway/breastscreen-norway/

Ghadge, A. G., Dasari, P., Stone, J., Thompson, E. W., Robker, R. L., & Ingman, W. V. (2021). Pubertal mammary gland development is a key determinant of adult mammographic density. *Seminars in Cell & Developmental Biology*, *114*, 143-158. https://doi.org/https://doi.org/10.1016/j.semcdb.2020.11.011

Giovannucci, E. (2019). A growing link—what is the role of height in cancer risk? *British Journal of Cancer*, *120*(6), 575-576. https://doi.org/10.1038/s41416-018-0370-9
GLOBOCAN. (2020). *Breast Cancer Fact sheet.* Global Cancer Observatory WHO. https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf
Green, J., Cairns, B. J., Casabonne, D., Wright, F. L., Reeves, G., & Beral, V. (2011). Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *The Lancet Oncology*, *12*(8), 785-794. https://doi.org/https://doi.org/10.1016/S1470-2045(11)70154-1
Groenwold, R. H. H., Palmer, T. M., & Tilling, K. (2021). To Adjust or Not to Adjust? When a "Confounder" Is Only Measured After Exposure. *Epidemiology*, *32*(2), 194-201. https://doi.org/10.1097/ede.0000000000001312
Hanahan, D., & Weinberg, Robert A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, *144*(5), 646-674. https://doi.org/https://doi.org/10.1016/j.cell.2011.02.013
Harrel, F. (2015). *Regression Modeling Strategies* (2nd ed.). Springer publishing.
Heymans, M. W., & Eekhout, I. (2019). *Applied Missing Data Analysis with SPSS and (R) studio* https://bookdown.org/mwheymans/bookmi/
Horn, J., Alsaker, M. D. K., Opdahl, S., Engstrøm, M. J., Tretli, S., Haugen, O. A., Bofin, A. M., Vatten, L. J., & Åsvold, B. O. (2014). Anthropometric factors and risk of molecular breast cancer subtypes among postmenopausal Norwegian women. *International journal of cancer*, *135*(11), 2678-2686. https://doi.org/https://doi.org/10.1002/ijc.28912
Huang, J., Chan, P. S., Lok, V., Chen, X., Ding, H., Jin, Y., Yuan, J., Lao, X.-q., Zheng, Z.-J., & Wong, M. C. (2021). Global incidence and mortality of breast cancer: a trend analysis. *Aging (Albany NY)*, *13*(4), 5748.
Jing, Z., Hou, X., Liu, Y., Yan, S., Wang, R., Zhao, S., & Wang, Y. (2015). Association between height and thyroid cancer risk: A meta-analysis of prospective cohort studies. *International journal of cancer*, *137*(6), 1484-1490. https://doi.org/https://doi.org/10.1002/ijc.29487
Krieg, S., Roderburg, C., Krieg, A., Luedde, T., Loosen, S. H., & Kostev, K. (2022). The association between body height and cancer: a retrospective analysis of 784,192 outpatients in Germany. *Journal of Cancer Research and Clinical Oncology*. https://doi.org/10.1007/s00432-022-04335-0
Lahmann, P. H., Hoffmann, K., Allen, N., Van Gils, C. H., Khaw, K. T., Tehard, B., Berrino, F., Tjønneland, A., Bigaard, J., & Olsen, A. (2004). Body size and breast cancer risk: findings from the European Prospective Investigation into Cancer And Nutrition (EPIC). *International journal of cancer*, *111*(5), 762-771.
Lerro, C. C., McGlynn, K. A., & Cook, M. B. (2010). A systematic review and meta-analysis of the relationship between body size and testicular cancer. *British Journal of Cancer*, *103*(9), 1467-1474. https://doi.org/10.1038/sj.bjc.6605934
Li, W., Liu, Q., Deng, X., Chen, Y., Liu, S., & Story, M. (2017). Association between obesity and puberty timing: a systematic review and meta-analysis. *International Journal of Environmental Research and Public Health*, *14*(10), 1266.
Little, M. A. (2020). Evolutionary strategies for body size. *Frontiers in Endocrinology*, *11*, 107.
Loda, M., Mucci, L., Mittelstadt, M., Hemelrijck, M., & Cotter, M. (2016). *Pathology and Epidemiology of Cancer*. Springer Publishing. https://doi.org/DOI 10.1007/978-3-319-35153-7
Lund, E., Dumeaux, V., Braaten, T., Hjartåker, A., Engeset, D., Skeie, G., & Kumle, M. (2007). Cohort Profile: The Norwegian Women and Cancer Study—NOWAC—

Kvinner og kreft. *International journal of epidemiology*, *37*(1), 36-41. https://doi.org/10.1093/ije/dym137
Lund, E., Kumle, M., Braaten, T., Hjartåker, A., Bakken, K., Eggen, E., & Gram Inger, T. (2003). External validity in a population-based national prospective study – the Norwegian Women and Cancer Study (NOWAC). *Cancer Causes & Control*, *14*(10), 1001-1008. https://doi.org/10.1023/B:CACO.0000007982.18311.2e
Michels, K. B., & Willett, W. C. (2004). Breast cancer--early life matters. *N Engl J Med*, *351*(16), 1679-1681. https://doi.org/10.1056/NEJMe048229
NCD. (2016). A century of trends in adult human height. *Elife*, *5*, e13410. https://doi.org/10.7554/eLife.13410
Nilsen, T. I. L., & Vatten, L. J. (2001). Adult height and risk of breast cancer: a possible effect of early nutrition. *British Journal of Cancer*, *85*(7), 959-961. https://doi.org/10.1054/bjoc.2001.1946
NOWAC. (2022). *Personvern*. Norwegian Women and Cancer study. Retrieved 20.11.22 from https://uit.no/research/nowac#647323_buttoncont1
Nunney, L. (1999). Lineage selection and the evolution of multistage carcinogenesis. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*, 493 - 498.
Nunney, L. (2018). Size matters: height, cell number and a person's risk of cancer. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1889), 20181743. https://doi.org/doi:10.1098/rspb.2018.1743
Osuch, J. R., Karmaus, W., Hoekman, P., Mudd, L., Zhang, J., Haan, P., & Mikucki, D. (2010). Association of age at menarche with adult leg length and trunk height: Speculations in relation to breast cancer risk. *Annals of Human Biology*, *37*(1), 76-85. https://doi.org/10.3109/03014460903213845
Parra-Soto, S., Ho, F. K., Pell, J. P., & Celis-Morales, C. (2020). Does insulin-like growth factor moderate the association between height and risk of cancer at 24 sites? *British Journal of Cancer*, *123*(11), 1697-1704.
Richiardi, L., Bellocco, R., & Zugna, D. (2013). Mediation analysis in epidemiology: methods, interpretation and bias. *International journal of epidemiology*, *42*(5), 1511-1519. https://doi.org/10.1093/ije/dyt127
Ritte, R., Lukanova, A., Tjønneland, A., Olsen, A., Overvad, K., Mesrine, S., Fagherazzi, G., Dossus, L., Teucher, B., Steindorf, K., Boeing, H., Aleksandrova, K., Trichopoulou, A., Lagiou, P., Trichopoulos, D., Palli, D., Grioni, S., Mattiello, A., Tumino, R., . . . Kaaks, R. (2013). Height, age at menarche and risk of hormone receptor-positive and -negative breast cancer: A cohort study. *International journal of cancer*, *132*(11), 2619-2629. https://doi.org/https://doi.org/10.1002/ijc.27913
Rosenfeld, R. G. (2003). Insulin-like growth factors and the basis of growth. *New England Journal of Medicine*, *349*(23), 2184-2186.
Rudolph, K. E., & Lau, B. (2021). Epidemiologic Methods: Seeing the Forest and the Trees. *Epidemiologic reviews*, *43*(1), 1-3. https://doi.org/10.1093/epirev/mxab010
Schoemaker, M. J., Jones, M. E., Allen, S., Hoare, J., Ashworth, A., Dowsett, M., & Swerdlow, A. J. (2017). Childhood body size and pubertal timing in relation to adult mammographic density phenotype. *Breast Cancer Research*, *19*(1), 13. https://doi.org/10.1186/s13058-017-0804-y
Sherwood, L. (2016). *Human Physiology: From Cells to systems* (9th ed.). Cengage learning.
Shrestha, R. M., Mizoue, T., Sawada, N., Matsuo, K., Wada, K., Tanaka, K., Lin, Y., Sugawara, Y., Takimoto, H., Kimura, T., Ito, H., Kitamura, T., Sakata, R., Tanaka, S., Inoue, M., Development, f. t. R. G. f. t., & Japan, E. o. C. P. S. i. (2021). Adult height in relation to the risk of colorectal cancer among the Japanese population: an

evaluation based on systematic review and meta-analysis. *Japanese Journal of Clinical Oncology*, *52*(4), 322-330. https://doi.org/10.1093/jjco/hyab203

Silventoinen, K. (2003). DETERMINANTS OF VARIATION IN ADULT BODY HEIGHT. *Journal of Biosocial Science*, *35*(2), 263-285. https://doi.org/10.1017/S0021932003002633

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research*, *75*(3), 417-453.

Soerjomataram, I., & Bray, F. (2021). Planning for tomorrow: global cancer incidence and the role of prevention 2020–2070. *Nature Reviews Clinical Oncology*, *18*(10), 663-672. https://doi.org/10.1038/s41571-021-00514-z

Van den Brandt, P. A., Ziegler, R. G., Wang, M., Hou, T., Li, R., Adami, H.-O., Agnoli, C., Bernstein, L., Buring, J. E., Chen, Y., Connor, A. E., Eliassen, A. H., Genkinger, J. M., Gierach, G., Giles, G. G., Goodman, G. G., Håkansson, N., Krogh, V., Le Marchand, L., . . . Smith-Warner, S. A. (2021). Body size and weight change over adulthood and risk of breast cancer by menopausal and hormone receptor status: a pooled analysis of 20 prospective cohort studies. *European journal of epidemiology*, *36*(1), 37-55. https://doi.org/10.1007/s10654-020-00688-3

Wang, F., Xu, X., Yang, J., Min, L., Liang, S., & Chen, Y. (2017). Height and lung cancer risk: A meta-analysis of observational studies. *PLOS ONE*, *12*(9), e0185316. https://doi.org/10.1371/journal.pone.0185316

Wang, Y. S. (2023). The Effect: An Introduction to Research Design and Causality: Huntington-Klein N. 2022. Boca Raton, FL: Routledge. ISBN: 1032125780. *Data Science in Science*, *2*(1). https://doi.org/10.1080/26941899.2023.2167433

Ward, S. V., Burton, A., Tamimi, R. M., Pereira, A., Garmendia, M. L., Pollan, M., Boyd, N., dos-Santos-Silva, I., Maskarinec, G., Perez-Gomez, B., Vachon, C., Miao, H., Lajous, M., López-Ridaura, R., Bertrand, K., Kwong, A., Ursin, G., Lee, E., Ma, H., . . . Stone, J. (2022). The association of age at menarche and adult height with mammographic density in the International Consortium of Mammographic Density. *Breast Cancer Research*, *24*(1), 49. https://doi.org/10.1186/s13058-022-01545-9

WCRF. (2017). *Diet, nutrition, physical activity and breast cancer*. https://www.wcrf.org/wp-content/uploads/2021/02/Breast-cancer-report.pdf

WCRF. (2018). *Height and birthweight and the risk of cancer*. World Cancer Research Fund.

Wysocki, A. C., Lawson, K. M., & Rhemtulla, M. (2022). Statistical Control Requires Causal Justification. *Advances in Methods and Practices in Psychological Science*, *5*(2), 25152459221095823. https://doi.org/10.1177/25152459221095823

Zhang, B., Shu, X.-O., Delahanty, R. J., Zeng, C., Michailidou, K., Bolla, M. K., Wang, Q., Dennis, J., Wen, W., Long, J., Li, C., Dunning, A. M., Chang-Claude, J., Shah, M., Perkins, B. J., Czene, K., Darabi, H., Eriksson, M., Bojesen, S. E., . . . Zheng, W. (2015). Height and Breast Cancer Risk: Evidence From Prospective Studies and Mendelian Randomization. *JNCI: Journal of the National Cancer Institute*, *107*(11). https://doi.org/10.1093/jnci/djv219

Zhou, E., Wang, L., Santiago, C. N., Nanavati, J., Rifkin, S., Spence, E., Hylind, L. M., Gills, J. J., La Luna, L., Kafonek, D. R., Cromwell, D. M., Drewes, J. L., Sears, C. L., Giardiello, F. M., Mullin, G. E., & Consortium, t. B. S. (2022). Adult-Attained Height and Colorectal Cancer Risk: A Cohort Study, Systematic Review, and Meta-Analysis. *Cancer Epidemiology, Biomarkers & Prevention*, *31*(4), 783-792. https://doi.org/10.1158/1055-9965.Epi-21-0398

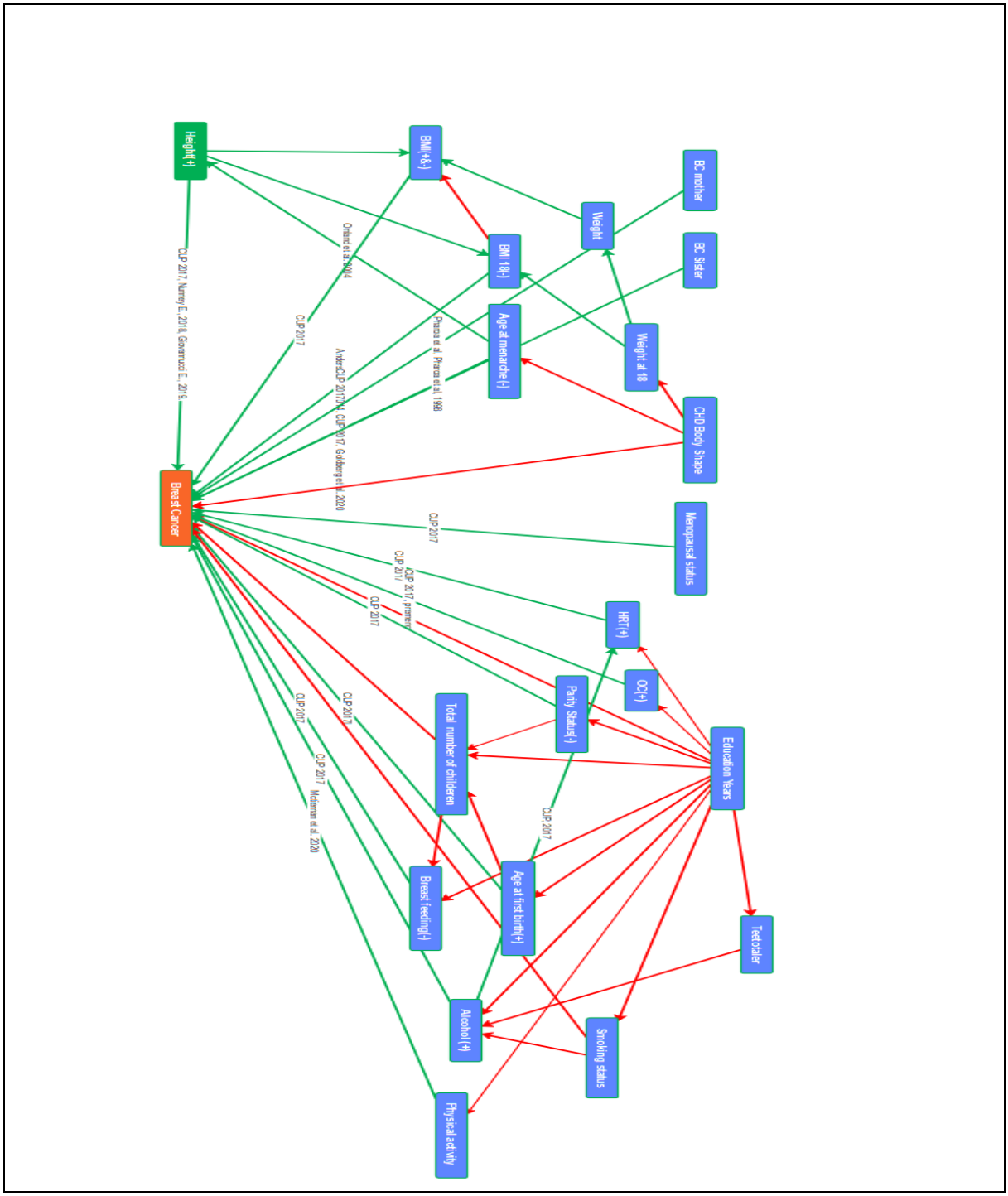
Appendix 1

NOWAC variable list

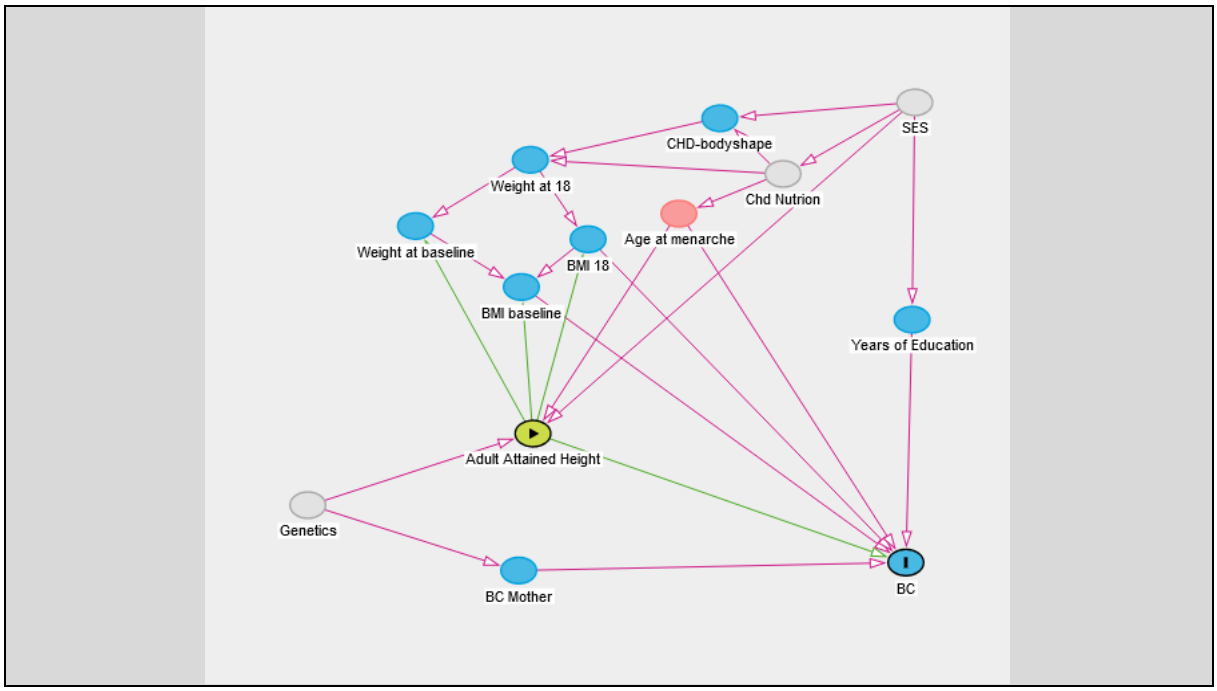
Table 1 - Variable list		
Variable group	**Variable**	**ID**
Cancer	Diagnosis code (ICD-10)	LOK_ICD10_KKCA
	Vital status	VITAL_STATUS_KKCA
	Date of emigration from the National Population Register	EMIG_DATE_KKCA
	Estrogen receptor status	ER_STATUS_KKCA
	Progesterone receptor status	PR_STATUS_KKCA
	Human epidermal growth factor receptor 2 status	HER2_STATUS_KKCA
	Date of diagnosis	DIAG_DATE_KKCA
Death	Date of death from the Norwegian Cause of Death Registry	DEATH_DATE_DC_KKDT
Anthropometry	Current self-reported body height	HEIGHT_KK1ME
	Current self-reported body weight	WEIGHT_KK1ME
	Self-reported body weight at age 18	WEIGHT_18_KK1ME
	Body shape at first grade.	BODY_SHAPE_KK1ME
Reproductive history	Age at menarche	MENARCHE_AGE_KK1ME
	Age at menopause	MENOPAUSE_AGE_KK1ME
	Total number of children	CHD_TOTAL_NUMBER_KK1DE
	Age at first birth	CHD_FIRST_AGE_MOTHER_KK1DE
	Menopausal status	MENOPAUSE_STATUS_KK1DE
Use of contraceptives	Ever oral contraceptive use	OC_EVER_KK1DE
Use of postmenopausal hormones	Ever menopausal hormone use	MHT_EVER_KK1DE
	Current menopausal hormone use	MHT_CURR_KK1DE
Family history of cancer	Has your mother been diagnosed with breast cancer?	BR_CANCER_MOTHER_2_KK1ME
	Has your mother been diagnosed with breast cancer?	BR_CANCER_MOTHER_1_KK1ME
	Have any of your sisters been diagnosed with breast cancer?	BR_CANCER_SISTER_KK1ME
Beverages	Daily intake of alcohol in grams	ALCO_INTAKE_GRAMS_KK1DE
	Are you a teetotaler?	TEETOTALER_KK1ME
Smoking	Smoking status	SMOKE_STATUS_KK1DE
Physical activity	Physical activity score at present	PHYSICAL_ACTIVITY_NOW_KK1ME
Socioeconomic conditions	Years of education in categories	EDUCATION_YEARS_CAT_KK1DE
General information	Age at enrolment	AGE_AT_ENROLMENT_KK1DE

Appendix 2

Initial DAG. All arrows were suggesting causal paths. Green arrows indicate causal path with meta-analysis to support connection. Red arrows is not supported by literature. Causal path related to both height and breast cancer was prioritized for investigation.



Appendix 3


SES
CHD-bodyshape
Weight at 18
Chd Nutrion
Weight at baseline
Age at menarche
BMI 18
BMI baseline
Years of Education
Adult Attained Height
Genetics
BC Mother
BC

Appendix 4

Covariable Details

Menopausal status:

Menopausal status is a derived variable based on several variables was measured in the questionnaire based on the following questions: “Are you still having regular menstruation?”, with the options to answer, “Yes”, “I have irregular menstruation”, “I don’t know”, “I am using hormonal therapy” or “No”.

In addition, if the participant answered “No”, follow up questions was presented as: “Did [menstruation] end by itself?”, “Have you had your ovaries surgically removed?”, “Have you had your uterus removed?”, “other”.

1. Premenopausal
2. Perimenopausal
3. Postmenopausal
4. Missing Information
5. Hysterectomy Before age 53
6. Use of MHT before age 53

The variable for menopausal status was recoded into a binary variable where premenopausal was recoded into “1”, Missing Information was recoded into “NA” and all remaining variables where coded “0”. Then “1” and “0” was labeled “Premenopausal” and “Postmenopausal”, respectively.

Years of Education

Years of education were measured through self-reported data. Participants were asked “How many years of education do you have in total?”. The participants replied in integers. NOWAC then categorized the results into four categories. The variable was coded 1-4 by NOWAC, for the analysis the variables were recoded to categorical variables.

1. <=9
2. 10-12
3. 13-16
4. >=17

The variable was coded 1-4 by NOWAC, for the analysis the variables were recoded to categorical variables.

Age at Menarche

All participant where asked: “At what age did you have your first menstruation?”.

As it already were a continuous variable, nothing was done to this variable.

Body shape at Childhood

All participant where asked: “What was your body type in 1st grade”. In the Norwegian school system, student normatively attend 1st grade the year they turn six years old.

0. Very thin
1. Thin
2. Normal
3. Fat
4. Very Fat

The variables where recoded into categorical variables corresponding to the categories described.

Years of Education

Years of education were measured through self-reported data. Participants were asked “How many years of education do you have in total?”. The participants replied in integers. NOWAC then categorized the results into four categories. The variable was coded 1-4 by NOWAC, for the analysis the variables were recoded to categorical variables.

Age at Menarche

All participants were asked: “At what age did you have your first menstruation?”.

As it already were a continuous variable, nothing was done to this variable.

Appendix 5

```
testImplications <- function( covariance.matrix, sample.size ){

        library(ggm)

         tst <- function(i){ pcor.test( pcor(i,covariance.matrix), length(i)-2, sample.size )$pvalue }

tos <- function(i){ paste(i,collapse=" ") }

implications <- list(c("Age at menarche",

                          "BMI",

                           "Adult Attained Height",

                           "Childhood Bodyshape",

                           "Years of education"),

                        c("Years of education",

                            "Breast Cancer In Mother"),

                        c("BMI",

                           "Breast Cancer In Mother",

                            "Years of education",

                            "Childhood Bodyshape",

                             "Adult Attained Height"))

        data.frame( implication=unlist(lapply(implications,tos)),

                  pvalue=unlist( lapply( implications, tst ) ) )

        }
```

Appendix 6

Sensitivity analysis of percentage change of effect estimates (HR) of covariable inclusion to the multivariable model.

Covariate	Overall Breast cancer	Premenopausal	Postmenopausal
Years of education	9,50 %	4,70 %	13 %
Age at menarche	5 %	5 %	8,70 %
Childhood Body shape	4,70 %	0 %	0 %
Breast cancer in mother	4,70 %	4,80 %	4,30 %