UiT The Arctic University of Norway

Faculty of Health Sciences, Department of Community Medicine

The mediating impact of lifestyle factors in the relationship between socioeconomic status and self-reported health in a Norwegian cohort of women

Anup Neupane

Master's thesis in Public Health HEL-3950 December 2020

Supervisor: Kristin Benjaminsen Borch Co-supervisor: Tonje Braaten

Table of Contents

List of Tables

List of Figures

Acknowledgments

I want to thank my supervisors, Kristin Benjaminsen Borch and Tonje Braaten, who guided me throughout my thesis journey. I am grateful for their comments and suggestions on the way to develop my thesis.

I am grateful to the University of Tromso for these two years of learning that have developed more academic and professional public health knowledge.

Manjila, thank you for your support and patience during the period of learning. More grateful for the emotional and mental support during the thesis writing process.

I want to thank everyone on my thesis writing journey and dedicate this work to my father.

Abstract

Introduction: Information on self-reported health (SRH) with a simple question on “How do you rate your health?” provides sufficient information to be a reasonable proxy of health. Socioeconomic inequalities in health are often monitored via SRH. Individuals with better socioeconomic status (SES) more likely to rate their SRH as good. The part of the effect of SES (income level and education) operates through lifestyle factors along with the physical environment, social environment, social support, and psychological (cognitive, emotional, and social capabilities) development. The extent of the mediation on the relation between SES and SRH with the individual lifestyle factor’s impact will provide more insights into the relationship between SES and SRH through lifestyle factors.

Material and methods: Data were extracted from the Norwegian Women and Cancer Study (NOWAC). After the inclusion/exclusion criteria, 53,941 participants were included in the analytical study sample. The descriptive statistics were presented for different education and income levels, using percentages, mean and standard deviation, and chi-square tests to test the difference. Logistic regression was used to establish the association between SES and SRH and reported with Odds Ratios (ORs) with 95% confidence interval (CI) using the statistical software SPSS, version 26. Mediation analysis was performed using the Medflex package in R applying a counterfactual approach, which was reported with ORs with 95% bootstrap CI along with the proportion of mediation.

Results: The odds of reporting poor SRH were nearly half with a higher household income level compared to the lowest household income level. Furthermore, in the highest income level, the odds of poor SRH were lowered by 80% compared to the lowest household income level. The odds of reporting poor SRH with a high level of education compared to the lowest education level were about 35% less. With the highest education level, the odds of reporting poor SRH were 53% less compared to the lowest education group.

About 20% of the association between income level and SRH was identified to go through an indirect path via lifestyle factors (smoking, body mass index (BMI), physical activity (PA)), whereas this was 20%-30% when investigating the model with education. Smoking and BMI contributed more to the indirect effect in the model with education, where PA contributed relatively less. Whereas in the model with income, PA had the greatest contribution, followed by BMI and smoking. The proportion of mediation by all the lifestyle factors (smoking, BMI, and PA) on the association between income level and SRH was statistically significant and quite similar across different income levels. There was an increasing gradient in the proportion of mediation across increasing education levels in the model with education and lifestyle factors (smoking, BMI, and PA).

Conclusion: SES (education and income level) had a prominent and statistically significant effect on SRH. Better SRH was observed with both higher income levels and years of education. Household income level shows a more prominent effect than years of education on differentiating across groups of SES. Each lifestyle factors (smoking, BMI, and PA) individually and jointly mediated the association between SES and SRH. Any interventions focusing on improving the healthy lifestyle of the population should comprehend the impact of SES.

Keywords: socioeconomic status, self-reported health, lifestyle factors, mediation effect, and counterfactual framework.

Abbreviations

Self-rated health (SRH)

Socioeconomic status (SES)

Body mass index (BMI)

Physical activity (PA)

Cardiovascular disease (CVD)

Norwegian women and cancer (NOWAC)

Randomized controlled trial (RCT)

Natural direct effect (NDE)

Natural indirect effect (NIE)

Total effect (TE)

1 Introduction

Self-rated health (SRH) is a global health quality of life measure that has been extensively used in medical and public health research, exploring different aspects of health (1-3). Being a part of the population survey, SRH needs to have reliability and validity to be used as a reasonable estimate of the population's health status. Due to its ease of collection and independent predictor of mortality and morbidities (1, 4), SRH serves as a proxy for health status (5, 6). Self-rated health correlates with other health scales, such as the Sickness Impact Profile and the Perceived Well-Being Scale, which indicates a high degree of construct validity (7). It is challenging to identify the reference that patients and individuals have in their minds while answering SRH. This may include illness, physical functioning, or behaviors and bring us to a broader construct and gaze of the SRH. People rate their health on a very poor scale to very good, which may vary across different populations, so countries may not be comparable, especially in developing and developed countries due to cross-cultural differences (8). A study supports the construct validity of SRH in developing countries like India to approve SRH as a measure to monitor population health (9). Likewise, SRH is a well-established measure in developed nations to monitor population health. However, despite the use in different country settings, we must be aware of the direct comparability as they may capture health differently and have cross-cultural differences. SRH can serve as a reasonable proxy for health in community-based health and population-level health. It can prioritize the public decision policy and disparities in cases where we often lack individualized health data (10). The need to understand the population perception of health by medical society is highlighted by Fylkesnes & Førde as a way of discussing predictors of SRH in the Tromso study (11). The use of SRH as a health measure while discussing the socioeconomic inequalities in health in Norway's public health report demonstrates that SRH is an accepted measure of proxy of objective health status in Norway (12).

1.1 Aspects of SRH

SRH is more of an individual perception of health, which can be viewed as physical health and mental health. Both physical and mental health conditions have their individual and collective impact while individuals assess their SRH (13).

1.1.1 Physical health

An individual would be likely to rate poor SRH if they have certain disease conditions. Most studies show that having multiple diseases and functional limitations are associated with poorer SRH (8, 14). Diabetes mellitus, cancer, heart disease, emphysema, hypertension, stroke, and arthritis; in a study to decompose the association of these chronic conditions to SRH by age, these chronic conditions were significant in all age group categories (15). Also, activity limitation of any kind is associated with poorer SRH (15).

1.1.2 Mental health

Following the construct of health itself, the mental health aspect is of equal importance as physical health in shaping health in an individual (13, 16). A linear improvement was observed in overall composite mental health attributes in a community-based study (17). Happiness Subscale and Reverse-coded Anxiety Scale and Depression Scale showed a linear relationship but inverted U-shaped pattern for the 36-Item Short Form Survey (SF-36) mental composite (17). Similarly, U-shaped relationships increasing with age and then a subtle decline after the 60s are seen in a prospective study looking at social inequality in health (18). With old age around 74 and over, mental health's impact on perceiving SRH is slightly more than functional limitations and chronic conditions (16).

The impact of physical and mental health plays a different role during the life process, where mental health becomes more prominent with age (17). Physical conditions are strongly related to perceived health, along with lifestyles and socioeconomic factors (7).

1.2 Predictors of self-rated health

In a study to assess self-rated global health in the general Norwegian population, better global health was associated with higher education, being employed, and living with a spouse or partner (2). SRH has been used extensively to study health and inequalities between men and women (19, 20). In change over time in the status of SRH, compositional variables such as age, sex, and education turned out to be significantly related to SRH (21).

1.2.1 Socioeconomic factors

Socioeconomic status(SES) is a measure of one's access to collectively desired resources and is a fundamental construct in the social and health sciences (22). SES is abundantly used in quantifying social inequalities. SES can be measured as univariate measurements such as education, income, occupation, wealth, poverty, and area-level measure also as composite measures such as Duncan Socioeconomic Index & Nam-Powers Occupational Status Score, Household prestige scale, CAPSES, Cambridge scale (22). Education, income level, and occupation status are the most common SES proxies.

Different indicators of SES such as education, income, and occupation may be embedded and operating within key domains such as (a) physical environment in which one lives and works and associated exposure to pathogens, carcinogens, and other environmental hazards; (b) the social environment and associated vulnerability to interpersonal aggression and violence as well as the degree of access to social resources and supports;(c) socialization and experiences that influence psychological development and ongoing mood, affect, and cognition; and (d) health behaviors(23). Similarly by Adler and Ostrove on SES health model of pathways by which SES influence health focused on (a) environmental resource and constraints with the external environment, social environment, and resources (b) psychological constraint with effect and cognition developing into exposure to carcinogen and pathogen, (c) preference of health-relevant behaviors along with (d) the central nervous system and endocrine to shape health and illness (24).

Many studies investigating the constructs and predictors of SRH have found that the socio-demographic factors are crucial in shaping SRH (25, 26). Age is one of the significant predictors, where people tend to report poor SRH with the increase in their age (25, 27-29). At the same time, education is a crucial aspect of primary health-related outcomes and has a significant association with SRH. Studies show, with a higher level of education, individuals report good SRH (25, 29, 30). Furthermore, higher-income is also associated with good SRH (20). The Whitehall study of British civil servants showed physical health deteriorated more rapidly for men and women from lower occupational grades. However, mental health improved with age; the improvement rate was slower for men and women in lower gradient (18). In Norway, educational differences in general and mental health were observed, with a larger gradient for general than mental health(31). The socioeconomic factors and their impact tend to accumulate over time, so a higher impact of socioeconomic factors can be seen in the elderly population (15). Apart from that, social ties and family life also plays a role in shaping individual SRH.

1.2.2 Lifestyle factors

Lifestyle factors are modifiable habits and ways of life that can influence overall health and well-being. Alameda county study on health practices showed never smoking, drinking less than five drinks at one sitting, sleeping 7-8 hours a night, exercising, maintaining desirable weight for height, eating breakfast regularly, and avoiding snacks, were predictors of good health (32). Different lifestyle factors have different impacts in our health. A systematic review with meta-analysis with smoking, obesity, exercise, diet, and drinking alcohol as lifestyle factors showed the number of healthy lifestyle behaviors people adopt is inversely related to the risk of all-cause mortality. Adherence to a healthy lifestyle leads to better health and the other way around with unhealthy lifestyle choices. In general, smoking is a significant risk factor in cardiovascular diseases, cancer, and respiratory diseases (33, 34). In a Swedish study, no smoking and vegetables in the diet were significantly associated with good SRH; obesity and underweight in young adults associated with poor SRH when adjusted for health problems (35). Maintaining moderate/high physical activity (PA) levels decreases the risk of cardiovascular disease (CVD) and several cancers (36), as shown by a systematic review (37), and physical activity holds an association with SRH (7).

Poor lifestyle habits and behavior will have certain and defined effects on an individual's health. Still, it may not be reflected with SRH as the construct and reporting SRH allows a wider gaze than medical problems and health behaviors. Individuals may have comparative health assessments among peers; for instance, despite having some issues, one may rate their health as good SRH.

1.3 Socioeconomic status and lifestyle choices

We are aware that people tend to make their lifestyle choices depending on their education and income. Studies show that higher SES is associated with healthy lifestyle choices such as reduced smoking, higher physical activity levels, and consciousness about their health (38). In a literature review, low SES in adolescence was associated with poorer diets, less physical activity, and higher cigarette smoking (39). A study on health-related behaviors in older adults showed multiple health risk behaviors were less common among individuals with higher SES (40). Poor adult health behaviors and psychological characteristics were more prevalent among men whose parents were poor, which shows childhood conditions reflect life course development of SES impact (41).

A study from China assessing social capital and health found that lifestyle factors fairly mediated the relationship between social trust and social relationships and health (42). Also, these lifestyle choices play roles in determining the overall health of an individual. People report their health as poor if they have unhealthy choices such as smoking, drinking, less PA (7, 35, 37). Cockerham's seminal work on assessing health and lifestyle's pathways highlights the important role of class circumstances, in general, SES, on shaping lifestyle behaviors (43). This then brings to the need for accounting for the association between SES and lifestyle while examining the relationship between SES and SRH. With higher SES, better lifestyle behaviors are observed. This provides ample evidence to support the role of lifestyle factors in mediating the relationship between SES and SRH. Despite the widespread use of SRH in health research, most of the literature is based on a cross-sectional study (44-46), which imposes reverse causality. Is the poor SRH due to an unhealthy lifestyle or poor SRH resulting in these unhealthy lifestyles? A significant issue of temporality is persisting in studies related to SRH.

1.4 Research Question

- What is the relationship between SES and SRH in a Norwegian cohort of women?
- To what extent do lifestyle factors mediate this relationship?

1.5 Aims

- To examine the association between SES (education and household income) and SRH.
- To examine the mediation effect of all lifestyle factors (smoking, PA, and BMI) independently and jointly on the association between SRH and household income/education as a proxy of SES.

1.6 Mediation

A mediator is a variable that accounts for a certain extent of the relationship between predictors and outcome. Traditional approaches to mediation analysis stretch back to the 1980s by Baron and Kenny(47).

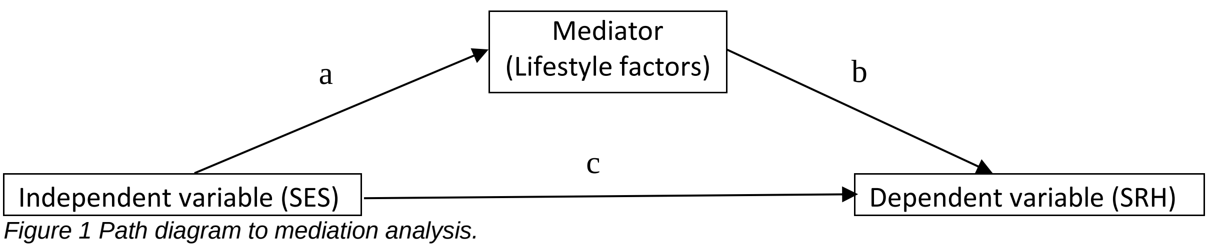


Figure 1 Path diagram to mediation analysis.

The causal chain involved in the mediation of our study is diagrammed in Figure 1. This model has two causal paths that are feeding into the outcome variable (SRH): the direct impact of the independent variable (SES) (Path c) and the impact of the mediators (Lifestyle factors) (Path b). There is also a path from the independent variable (SES) to the mediator

(Lifestyle factors) (Path a). Direct effect estimate results from direct impact via (Path c), indirect effect estimate is the product of the coefficients of (Path a) and (Path b).

There are certain conditions to be fulfilled to prove mediation, (a) Variations in levels of the independent variable significantly account for variations in the presumed mediator (Path a), (b) Variations in the mediator significantly account for variations in the dependent variable (Path b), (c) When both the paths mentioned above are controlled, a previously significant relation between the independent and dependent variables is no longer significant (Path c), with the strongest demonstration of mediation occurring when the regression coefficient is zero if not zero then this indicates the operation of multiple mediating factors(47). Baron and Kenny suggested providing an approximate significance test (Sobel method) for direct and indirect effects. The Sobel method uses the regression coefficient of the independent variable and the dependent variable and the coefficient describing the relationship between the mediating variable and the dependent variable after controlling for or taking into account the effect of the independent variable on the dependent variable and standard errors with those regression coefficients (48).

Baron and Kenny's procedure had limitations as they had a conservative approach that could not incorporate the wider dimension of variables such as binary variables that are used frequently. Their approach is only justified for continuous variables and linear settings. Different extensions, such as the popular Hayes approach to mediation, added the flexibility of having a binary outcome and exposure but demanded a continuous mediator (49). With the application of path analysis with Causal flow diagrams, the so-called difference and product of coefficient methods came into practice, which is more prevalent in literature. These methods are equivalent and justified when both the models for outcome and mediator are linear and with no interaction. They differ and raise validity concerns when one or both of these models are non-linear (50).

1.6.1 Counterfactual outcomes and effect decomposition

A potential outcome is an outcome that would be observed if the individual received a specific value of the treatment. One can generally observe only one, but not both, of the two potential outcomes for each particular individual. The unobserved outcome is called the counterfactual outcome (51). The counterfactual framework provides the opportunity to decompose the total causal effect into the natural direct and natural indirect effect irrespective of the effect's data distribution or scale. Randomization provides greater strength to

Randomized Controlled Trial (RCT); however, in observational and social science, lack of randomization brings up a solution via counterfactual framework (52). In this current study, the Medflex package is used as this package was developed to cover the limitation of existing packages that dealt with mediation analysis (53). With the imputation-based approach, Medflex can virtually deal with any variable as it does not require a mediator model specification and is based on an appropriate model for the outcome mean.

2 Materials and Methods

2.1 The Norwegian Women and Cancer Study (NOWAC)

The Norwegian Women and Cancer (NOWAC) study is a nationwide cohort, initially set up to examine breast cancer risk with oral contraceptive use (54). The cohort is described in detail in Lund et al. (2008) (ref51), but a brief overview is given here. The series of questionnaires are grouped into three mailings. In the first mailing, 179,387 women were sent a letter of invitation in the period from 1991 to 1997, of which 102,540 (57.2%) responded. Cohort expansion was done between 2003 to 2006, where additional 130,577 women were invited with a response rate of (48.4%) (63,282). A second mailing (first follow-up) was performed from 1998 to 2002 to update the exposure information for women enrolled from 1991 to 1997 and in 2011 to update exposure information for women enrolled during 2003. During 2003 – 2005, the third mailing (second follow up) was done for women enrolled in 1991-1995; flowchart on the mailings and follow up from cohort profile is in appendix 1. The NOWAC cohort has collected data on various lifestyle factors, diet, self-reported illness, social background, smoking, physical activity, and information regarding reproductive health through a questionnaire with informed consent. The NOWAC cohort is linked with the Cancer Registry of Norway for cancer information.

2.2 Study sample

This study's data includes the baseline of the first and second waves of NOWAC 1991-1997 and 2003-2005, with the first follow-up in 1997-2002 and 2011, which gave a today study sample of 101,316. Individuals should have answered the question regarding SRH at baseline and follow-up questionnaires; this led to the loss of 36,086 participants and resulted in a remaining of 65,230 participants, as shown in the flowchart below (figure 2). At baseline, good SRH was the inclusion criteria, which led to an additional loss of 4794 participants and resulted in 60,436 participants in the study. Women with missing information on any variable were excluded for complete case analysis, which brought to a final study sample of 53,941 participants (figure 2).

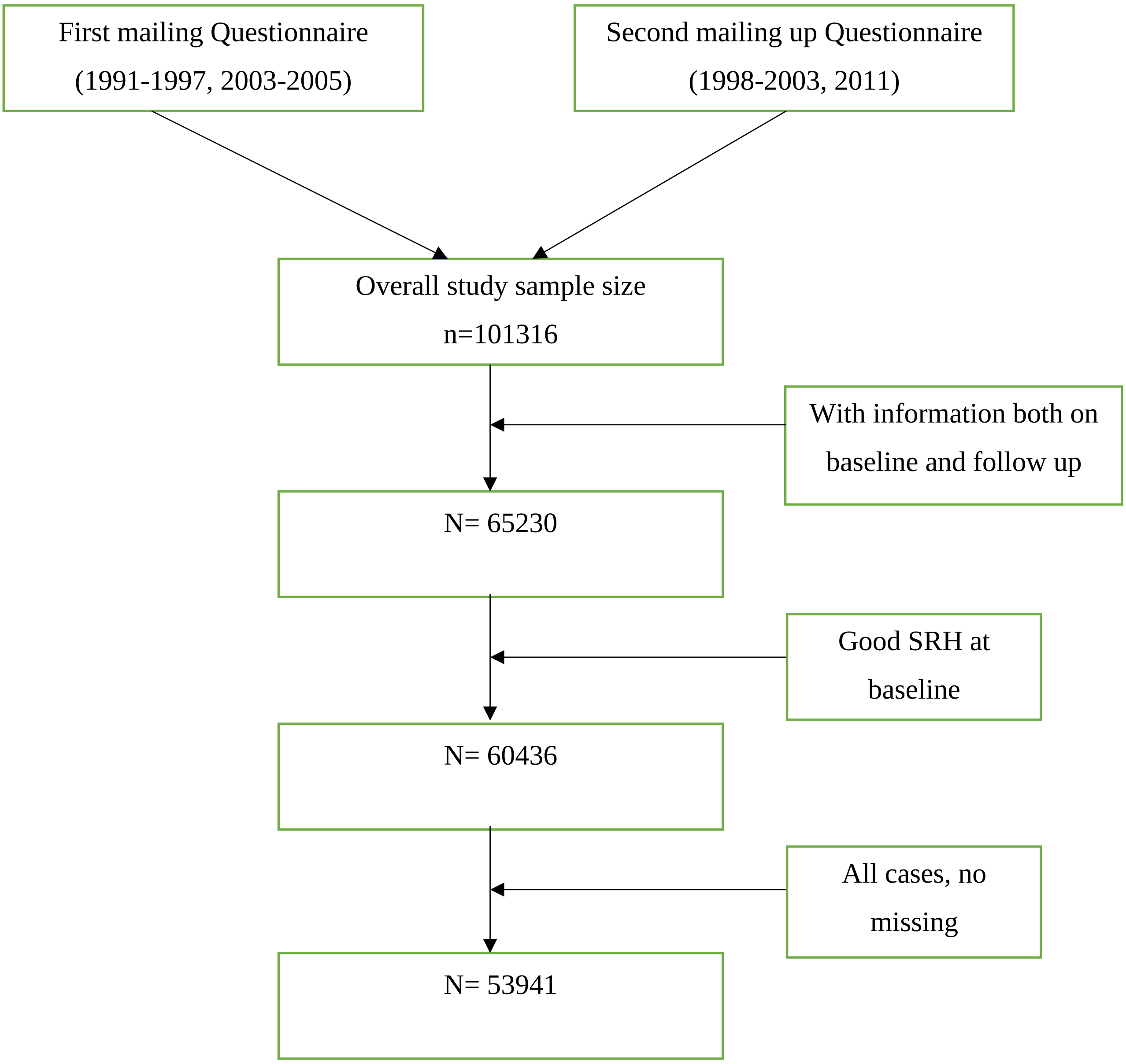


Figure 2 Flow diagram for study population.

2.3 Variables

2.3.1 Outcome

The dependent variable SRH is framed as “How would you rate your current state of health” with four levels as an option “very good,” “good,” "poor,” “very poor.”

2.3.2 Exposure

Household income

Data on household gross annual income is collected in NOWAC. It has categorical options as less than 150000kr, 151000-300000kr, 301000-450000kr, 451000-600000, more than 600000kr.

Education

Years of education information are collected from the baseline questionnaire. Information on years of education is collected.

2.3.3 Lifestyle factors

Smoking

Information on smoking status is collected via two questions; “Have you ever smoked?” and “Do you smoke currently?”

Physical activity

NOWAC inquiries about PA on a scale of 1 to 10, very low to very high, respectively.

BMI

NOWAC collects information on self-reported height (cm) and weight (kg). BMI was calculated using relation BMI = weight (kg)/height $(m)^2$.

2.3.4 Other factors

Age

The age of the individual was calculated from the birth year and date of answering the first questionnaire. The birth year is obtained from the Norwegian Central Person Register.

Marital status

Information regarding marital status was collected as a married, cohabitant, and living alone.

2.4 Statistical analysis

The outcome SRH is dichotomized into two groups, “good” by combining “very good” and “good” and “poor” by combining “poor” and “very poor.” In NOWAC, some questionnaires of SRH were constructed by self-rated physical health and self-rated mental health separately instead of overall SRH. A new variable for SRH was constructed by merging good and very good of both physical and mental health and merged into the SRH variable. For education, information on total years of education is divided into four categories, less than nine years, 9 to 12 years, 13 to 16 years, and more than or equals to 17 years. Household income levels were used as collected without recoding. Regarding smoking status, based on the questions "Have you ever smoked?" and "Do you smoke currently?" with binary response yes/no, we created categories for current, former, and never-smokers. PA is recoded by combing 1, 2, and 3 for low, 4, 5, 6, and 7 for moderate, 8, 9, and 10 for high physical activity levels. BMI is modeled as a categorical variable by creating levels as underweight (less than 19.9 kg/m^2, normal weight (20-24.9 kg/m^2), overweight (25-29.9 kg/m^2), and obese (≥30 kg/m^2). Marital status was dichotomized by combining married and cohabitant as living together and living alone. Age is the only continuous variable in our model.

Cross tabulations of baseline distribution of lifestyle factors (BMI, smoking status, and physical activity), age, and marital status by education level and household income were performed. The frequency distributions between groups were tested using the chi-square test and spearman correlation. Three logistic models, the first model with age, the second model with the socioeconomic status, and the third model with age, mediators (BMI, smoking status, and PA), and SES are analyzed in the model with SES as education. Whereas in the model with income level as SES, additional adjustment with marital status was performed will all the logistic models. The analysis was done using the statistical software SPSS 26.

Additionally, the counterfactual framework was used to analyze the natural direct effect (NDE) and natural indirect effect (NIE) for each SES proxy (i.e., education and household income) with lifestyle factors as mediators adjusting for age and marital status. The details on the calculation and estimation of NDE and NIE is in appendix 2. Estimation of natural direct and indirect effects with bootstrap 95% CI was performed using the Medflex package in the R 3.6.3 version. The mediated proportion was calculated using the relation:

Mediated proportion (%) = $\frac{NIE}{TE} * 100\%$

Where total effect (TE) = NIE +NDE

Sequential addition of lifestyle factors, smoking, BMI, and PA was done on the mediation model to quantify the individual impact of lifestyle factors.

2.5 Ethical perspectives

NOWAC has received approval from The Regional Committee for Medical Research Ethics (REK) to collect and store questionnaire information (32). For this study, data was provided by NOWAC in an anonymized form, and REK application was not required.

3 Result

3.1 Baseline characteristics of the study population

The cohort of 53941 women had a mean age of 51years (range 41-71) (table 1). Furthermore, table 1 shows that the proportion of never, former and current smokers were 38%, 34.9%, and 27.1%, respectively. The majority of women were in the normal BMI category (55%), whereas the underweight, overweight, and obese proportion was 6.3%, 29.4%, and 8.8%, respectively. Similarly, a greater proportion (73.4%) of women had a moderate PA level, with 11.8% having low PA levels and 14.8% with high PA levels. The proportion of married or cohabitant women was considerably high (82.4%) (Table 1).

Table 1 Baseline characteristics of the study population from the Norwegian Women and Cancer Study

Baseline characteristics (N = 53941)	Number	Percentage
Age		
Mean (sd)	51.08	6.01
Education[a]		
≤9	11508	21.3
10 to 12	18371	34.1
13 to 16	15931	29.5
≥17	8131	15.1
Household income[c]		
<150000	3643	6.8
151000-300000	14939	27.7
301000-450000	15357	28.5
451000-600000	11824	21.9
>600000	8178	15.2
Marital status		
Married or cohabitant	44442	82.4
Alone	9499	17.6
Smoking		
Never	20476	38
Former	18839	34.9
Current	14626	27.1
Body mass index		
Under weight (<20 kg/m^2)	3406	6.3
Normal (20 kg/m^2-24.9 kg/m^2)	29913	55.5
Overweight (25 kg/m^2-30 kg/m^2)	15879	29.4
Obese (>30 kg/m^2)	4743	8.8
Physical activity[b]		
Low	6351	11.8
Moderate	39614	73.4
High	7976	14.8

[a]*Years of education.*
[b]*Measured physical activity on a scale from 1-10, low (≤3), medium (4-7), high (≥8).*
[c]*Gross income per household in Norwegian kroner.*

3.2 Distribution of baseline characteristics across SES indicator

3.2.1 Education as SES

Table 2 shows the distribution of variables across the education categories. Across different education categories ≤9, 10 to 12, 13 to 16, ≥17, the proportion of women was 21.3%, 34.1%, 29.5%, 15.1%, respectively. With the increasing level of education, the proportion of never and former smokers increased, whereas, with current smokers, it was the opposite. Similarly, for BMI categories, the proportion of women in underweight and normal BMI increased, whereas overweight and obese individuals were decreasing with an increase in education levels. The highest proportion of low and high PA levels was observed in the lowest education group, and the proportion was decreasing with increasing education levels. There was no pattern of increase or decrease with different education levels for moderate PA levels (Table2).

Table 2 Distribution of study variable according to education groups from the Norwegian Women and Cancer Study

	Education levels[a] (N (%))				
	≤9	10 to 12	13 to 16	≥17	p value
Total (n)	11508	18371	15931	8131	
Age (years)					
Mean (sd)	53.57 (6.74)	50.81 (5.81)	50.15(5.34)	49.97(5.17)	
Smoking					<0.001[c]
Never	3629(31.5)	6163(33.5)	6801(42.7)	3883(47.8)	
Former	3656(31.8)	6434(35.0)	5700(35.8)	3049(37.5)	
Current	4223(36.7)	5774(31.4)	3430(21.8)	1199(14.7)	
Body mass index[d]					<0.001[e]
Under weight (<20)	591(5.1)	1032(5.6)	1694(6.9)	689(8.5)	
Normal (20-24.9)	5556(48.3)	10043(54.7)	9295(58.3)	5019(61.7)	
Overweight (25-30)	3967(34.5)	5601(30.5)	4386(27.5)	1925(23.7)	
Obese (>30)	1394(12.1)	1695(9.2)	1156(7.3)	498(6.1)	
Physical activity[b]					<0.001[e]
Low	1671(14.8)	2167(11.8)	1622(10.2)	891(11.0)	
Moderate	7970(69.3)	13689(74.5)	11980(75.2)	5975(73.5)	
High	1867(16.2)	2515(13.7)	2329(14.6)	1265(15.6)	
Marital status					<0.001[c]
Married or cohabitant	9526(82.8)	15413(83.9)	13123(82.4)	6380(78.5)	
Alone	1982(17.2)	2958(16.1)	2808(17.8)	1751(21.8)	

[a]*Years of education.*
[b]*Measured physical activity on a scale from 1-10, low (≤3), medium (4-7), high (≥8).*
[c]*The chi-square test statistics, significant at the 5% level.*
[d]*Body mass index unit kg/m^2.*
[e]*Spearman correlation significant at 5% level.*

3.2.2 Household income as SES

The distribution of lifestyle factors across different household income level is shown in table 3. The proportion of women across different income categories ie ≤1500000, 151000 to 300000, 301000 to 450000, 451000 to 600000 and >600000 was 6.8%, 27.7%, 28.5%, 21.9% and 15.2% respectively. The trend of smoking across the groups of income was similar to education levels, with a higher proportion of never and former smokers, while current smokers decreased with the increase in income levels. The proportion of obese and overweight women decreased with an increase in income levels, whereas for those in the normal BMI group, the proportion increased with increasing income levels, while no trend was observed for those in the underweight BMI group. The lower proportion of low PA levels was observed in higher income levels, whereas in the high PA category, a similar proportion was observed except for the highest income level. Distinct effect of being married or cohabitant on having a higher income is seen as more than 95% of women were cohabitant or married who were in the higher-income category whereas in the lower-income category (≤ 150000) the proportion was quite similar with married or cohabitant and living alone with 49.0% and 51.0% respectively and shifted in more proportion of married and cohabitant with increasing income level.

Table 3 Distribution of study variables according to household income levels from the Norwegian Women and Cancer Study

	Income levels[a] (Number, proportion)					
	≤150000	151000 to 300000	301000 to 450000	451000 to 600000	≥600000	p value
Total (n)	3643	14939	15357	11824	8178	
Age (years)						
Mean (sd)	55.45(7.85)	51.96(6.61)	50.46(5.51)	49.78(5.05)	50.56(4.84)	
Smoking						<0.001[c]
Never	1293(35.5)	5431(36.4)	5807(37.8)	4508(38.8)	3357(41.0)	
Former	1005(27.6)	4686(31.4)	5234(34.1)	4452(37.7)	3462(42.3)	
Current	1345(36.9)	4822(32.3)	4316(28.1)	2784(23.5)	1359(16.6)	
Body mass index[d]						<0.001[e]
Under weight (<20)	260(7.1)	957(6.4)	914(6.0)	720(6.1)	555(6.8)	
Normal (20-24.9)	1653(45.4)	7763(52.0)	8436(54.9)	6939(58.7)	5122(62.6)	
Overweight (25-30)	1254(34.4)	4625(31.0)	4619(30.1)	3333(28.2)	2048(25.0)	
Obese (>30)	476(13.1)	1594(10.7)	1388(9.0)	832(7.0)	453(5.5)	
Physical activity[b]						<0.001[e]
Low	752(20.6)	1981(13.3)	1674(10.9)	1212(10.3)	732(9.0)	
Moderate	2363(64.9)	10780(72.2)	11533(75.1)	8916(75.4)	6022(73.6)	
High	528(14.5)	2178(14.6)	2150(14.0)	1696(14.3)	1424(17.4)	
Marital status						<0.001[c]
Married or cohabitant	1784(49.0)	9374(62.7)	13729(89.4)	111488(97.2)	8067(98.6)	
Alone	1859(51.0)	5565(37.3)	1628(10.6)	336(2.8)	111(1.4)	

[a]*Gross income per household in NOK.*
[b]*Measured physical activity on a scale from 1-10, low (≤3), medium (4-7), high (≥8).*
[c]*The chi-square test statistics, significant at the 5% level.*
[d]*Body mass index unit kg/m^2.*
[e]*Spearman correlation significant at 5% level.*

3.3 Logistic regression analysis

3.3.1 Education as an indicator of socioeconomic status

All education levels effect estimates were significant in the baseline model with adjustment for age and marital status. The odds of having poor SRH was 0.65 (95% CI 0.60 to 0.70) for education level 10 to 12 years, 0.47 (95% CI 0.43 to 0.52) for education level 13 to 16 years, 0.46 (95% CI 0.41 to 0.51) for education level ≥17 years compared to education level less than nine years. With the addition of lifestyle factors, the odds increased for all education levels, 0.70 (95% CI 0.64 to 0.76), 0.55 (95% CI 0.50 to 0.61), 0.56 (95% CI 0.50 to 0.63) for education levels 10 to 12, 13 to 16, and ≥17, respectively (Table 4).

Table 4 Odds ratios (ORs) and 95% confidence interval for poor self-reported health across different education levels in the baseline and final model from the Norwegian Women and Cancer Study

	Baseline model OR (95% CI)	Final model OR (95% CI)
Education levels[a]		
≤9	1	1
10 to 12	0.65 (0.60 to 0.70)	0.70 (0.64 to 0.76)
13 to 16	0.47 (0.43 to 0.52)	0.55 (0.50 to 0.61)
≥17	0.46 (0.41 to 0.51)	0.56 (0.50 to 0.63)
Smoking status		
Never		1
Former		1.14 (1.05 to 1.23)
Current		1.57 (1.45 to 1.71)
Physical activity[b]		
Low		1
Moderate		0.34 (0.32 to 0.37)
high		0.23 (0.20 to 0.26)
Body mass index[c]		
Normal (20 – 24.9)		1
Underweight (<20)		1.34 (1.17 to 1.53)
Overweight (25 - 30)		1.35 (1.25 to 1.40)
Obese (>30)		2.31 (2.10 to 2.54)

[a]Years of education.
[b]Measured physical activity on a scale from 1-10, low (≤3), medium (4-7), high (≥8).
[c]Body mass index unit kg/m^2.
Both models adjusted for age.

3.3.2 Household income as an indicator of socioeconomic status

The odds for having poor SRH was 0.51 (95% CI 0.46 to 0.57), 0.31 (95% CI 0.32 to 0.40), 0.26 (95% CI 0.23 to 0.29), 0.20 (95% CI 0.17 to 0.23) for household income levels 151000 to 300000, 301000 to 450000, 451000 to 600000 and >600000 respectively compared to income less than 150000 adjusted for age and marital status. With introduction of lifestyle factors the odds increased to 0.57 (95% CI 0.52 to 0.64), 0.43 (95% CI 0.38 to 0.48), 0.32

(95% CI 0.28 to 0.37) and 0.27 (95% CI 0.23 to 0.32) for 151000 to 300000, 301000 to 450000, 451000 to 600000 and >600000 respectively (Table 5).

Table 5 Odds ratios (ORs) and 95% confidence interval for poor self-reported health across different household income levels in the baseline and final model from the Norwegian Women and Cancer Study

	Baseline model OR (95% CI)	final model OR (95% CI)
Household income levels[a]		
≤150000	1	1
151000-300000	0.51 (0.46 to 0.57)	0.57 (0.52 to 0.64)
301000-450000	0.36 (0.32 to 0.40)	0.43 (0.38 to 0.48)
451000-600000	0.26 (0.23 to 0.29)	0.32 (0.28 to 0.37)
600000+	0.20 (0.17 to 0.23)	0.27 (0.23 to 0.32)
Smoking status[d]		
Never		1
Former		1.19 (1.09 to 1.29)
Current		1.62 (1.49 to 1.75)
Physical activity[b]		
Low		1
Moderate		0.35 (0.33 to 0.38)
High		0.24 (0.21 to 0.27)
Body mass index[c]		
Normal (20 - 24.9)		1
Underweight (<20)		1.29 (1.13 to 1.48)
Overweight (25 - 30)		1.34 (1.24 to 1.44)
Obese (>30)		2.23 (2.05 to 2.49)

[a]Gross income per household in Norwegian Kroner.
[b]Measured physical activity on a scale from 1-10, low (≤3), medium (4-7), high (≥8).
[c]Body mass index unit kg/m^2.
Both models adjusted for age and marital status.

3.4 Mediation analysis

3.4.1 Education as an indicator of socioeconomic status

The proportion mediated by smoking was 7.80%, 11.47%, and 15.08% for education levels 10 to 12, 13 to 16, and ≥17, respectively, adjusted for age. With BMI and smoking in the model, the proportion mediated due to smoking and BMI was 18.67%, 22.49%, and 30.75%

across different education levels adjusted for age. The change in proportion accounted for BMI was 10.87%, 11.02%, and 15.07% across different education levels.

For all three mediators, there was an increasing gradient in the proportion of mediation with values 20.09%, 24.05%, 28.92% across different education levels Change in proportion accounted for PA was 2.27%, 2.45%, and -0.88% across different education levels. For education levels, 10 to 12, the OR for NDE, NIE, and TE was 0.71 (95% CI 0.65-0.77), 0.91 (95% CI 0.89-0.93), and 0.65 (95% CI 0.60-0.70), respectively, adjusted for age. The OR for NDE, NIE, and TE were 0.57(95% CI 0.52-0.63), 0.83 (95% CI 0.81-0.85), and 0.48 (95% CI 0.44-0.52) for education level 13 to 16 and 0.59 (95% CI 0.53-0.66), 0.79 (95% CI 0.77-0.82), and 0.48 (95% CI 0.42-0.52) for education ≥17.

Table 6 ORs for NDE, NIE and TE; proportion mediated and change in proportion mediated for different education levels on SRH mediated by lifestyle factors (smoking, BMI, and PA) with bootstrap 95% confidence interval from the Norwegian Women and Cancer Study.

Mediator	Education levels	Effect	OR (95% CI)	Proportion mediated (95% CI)	Change in proportion mediated
Smoking					
	Ref (<9)				
	9 to 12	NDE	0.67 (0.62 to 0.73)	7.80 (8.18 to 7.62)	
		NIE	0.96 (0.95 to 0.97)		
		TE	0.65 (0.60 to 0.71)		
	13 to 16	NDE	0.52 (0.48 to 0.57)	11.47 (12.10 to 10.74)	
		NIE	0.91 (0.90 to 0.93)		
		TE	0.48 (0.44 to 0.52)		
	17 +	NDE	0.53 (0.47 to 0.59)	15.68 (16.12 to 12.29)	
		NIE	0.88 (0.87 to 0.90)		
		TE	0.47 (0.42 to 0.53)		
Smoking and BMI					
	Ref (<9)				
	9 to 12	NDE	0.70 (0.65 to 0.76)	18.67 (18.24 to 19.44)	10.87
		NIE	0.92 (0.91 to 0.93)		
		TE	0.65 (0.60 to 0.70)		
	13 to 16	NDE	0.56 (0.51 to 0.62)	22.49 (22.26 to 22.75)	11.02
		NIE	0.84 (0.83 to 0.86)		
		TE	0.48 (0.43 to 0.52)		
	17 +	NDE	0.59 (0.53 to 0.66)	30.75 (29.89 to 32.02)	15.07
		NIE	0.79 (077 to 0.81)		
		TE	0.47 (0.42 to 0.52)		
Smoking, BMI and PA					
	Ref (<9				
	9 to 12	NDE	0.71 (0.65 to 0.77)	20.94 (21.18 to 20.41)	2.27
		NIE	0.91 (0.89 to 0.93)		
		TE	0.65 (0.60 to 0.70)		
	13 to 16	NDE	0.57 (0.52 to 0.63)	24.94 (25.10 to 24.92)	2.45
		NIE	0.83 (0.81 to 0.85)		
		TE	0.48 (0.44 to 0.52)		
	17 +	NDE	0.59 (0.53 to 0.66)	29.87 (29.56 to 30.44)	-0.88
		NIE	0.79 (0.77 to 0.82)		
		TE	0.47 (0.42 to 0.52)		

NDE= Natural direct effect, NIE= Natural indirect effect, TE= Total effect, ORs= Odd ratios, BMI= Body mass index, PA= Physical activity, SRH= self reported health.
Education levels= Education in years.

3.4.2 Household income as an indicator of socioeconomic status

The proportion mediated by smoking was 3.78%, 4.15%, 4.43%, and 5.20% for different household income level 151000 to 300000, 301000 to 450000, 451000 to 600000 and >600000 respectively, adjusted for age and marital status.

With BMI and smoking in the model, the proportion mediated was 9.97%, 11.38%, 12.96%, and 14.62% across different household income levels adjusted for age and marital status. The change in proportion accounted for BMI was 6.19%, 7.23%, 8.53%, and 9.42% across different household income levels.

With all three mediators smoking, BMI and physical activity, the proportion mediated was 21.75%, 21.07%, 19.95% and 21.68% across different household income levels. The change in proportion of PA was 11.78%, 9.69%, 7.33% and 7.06% across different household income level. The OR for NDE was 0.59 (0.53-0.65), 0.44(0.39 – 0.49), 0.34(0.30-0.38), and 0.29(0.25-0.33) across different household income levels adjusted for age and marital status. OR for NIE was 0.86(0.84-0.88), 0.80(0.78-0.82), 0.76(0.74-0.79) and 0.70(0.68-0.73) across different household income levels. OR for TE was 0.51(0.46-0.56), 0.35(0.32-0.40), 0.26(0.23-0.29) and 0.20(0.17-0.23) across different household income levels.

Table 7 ORs for NDE, NIE, TE; proportion mediated and change in proportion mediated for different household income levels on SRH mediated by lifestyle factors (smoking, BMI and PA) with bootstrap 95% confidence interval from the Norwegian Women and Cancer Study.

Mediator	Income Level	Effect	OR (95% CI)	Proportion Mediated (95% CI)	Change in proportion mediated
Smoking					
	Ref (≤ 150,000)				
	151000-300000	NDE	0.53 (0.47 to 0.58)	3.78 (4.45 to 3.03)	
		NIE	0.97 (0.96 to 0.98)		
		TE	0.51 (0.46 to 0.57)		
	301000-450000	NDE	0.37 (0.33 to 0.42)	4.15 (4.73 to 3.30)	
		NIE	0.95 (0.94 to 0.96)		
		TE	0.36 (0.32 to 0.40)		
	451000-600000	NDE	0.27 (0.24to 0.31)	4.43 (5.10 to 3.66)	
		NIE	0.94 (0.92 to 0.95)		
		TE	0.26 (0.23 to 0.29)		
	600000+	NDE	0.22 (0.19 to 0.25)	5.20 (5.96 to 4.40)	
		NIE	0.92 (0.90 to 0.93)		
		TE	0.20 (0.17 to 0.23)		

Smoking and BMI					
	Ref (≤ 150,000)				
	151000-300000	NDE	0.55 (0.49 to 0.61)	9.97 (9.92 to 11.88)	6.19
		NIE	0.93 (0.92 to 0.95)		
		TE	0.51 (0.46 to 0.57)		
	301000-450000	NDE	0.40 (0.36 to 0.45)	11.38 (10.64 to 11.98)	7.23
		NIE	0.89 (0.87 to 0.45)		
		TE	0.36 (0.32 to 0.40)		
	451000-600000	NDE	0.31 (0.27 to 0.35)	12.96 (12.09 to 13.17)	8.53
		NIE	0.84 (0.82 to 0.86)		
		TE	0.26 (0.23 to 0.29)		
	600000+	NIE	0.26 (0.22 to 0.30)	14.62 (14.11 to 15.03)	9.42
		NIE	0.79 (0.77 to 0.81)		
		TE	0.20 (0.17 to 0.23)		
Smoking, BMI and PA					
	Ref (≤ 150,000)				
.	151000-300000	NDE	0.59 (0.53 to 0.65)	21.75 (22.25 to 21.10)	11.78
		NIE	0.86 (0.84 to 0.88)		
		TE	0.51 (0.46 to 0.56)		
	301000-450000	NDE	0.44 (0.39 to 0.49)	21.07 (21.42 to 20.57)	9.69
		NIE	0.80 (0.78 to 0.82)		
		TE	0.35 (0.32 to 0.40)		
	451000-600000	NDE	0.34 (0.30 to 0.38)	19.95 (20.33 to 19.39)	7.33
		NIE	0.76 (0.74 to 0.79)		
		TE	0.26 (0.23 to 0.29)		
	600000+	NDE	0.29 (0.25 to 0.33)	21.68 (21.85 to 21.50)	7.06
		NIE	0.70 (0.68 to 0.73)		
		TE	0.20 (0.17 to 0.23)		

NDE= Natural direct effect, NIE= Natural indirect effect, TE= Total effect, ORs= Odd ratios, BMI= Body mass index, PA= Physical activity, SRH= self reported health.
Income level = Gross income per household in Norwegian Kroner

4 Discussion

4.1 Summary of main findings

In this study, the association between SES and SRH and the mediation effect of lifestyle factors on the association between SES and SRH were investigated. The main findings showed that SES (education and income level) had a prominent and statistically significant effect on SRH. Better SRH was observed with a higher income level and higher education. For mediation, each lifestyle factor individually and jointly mediated the relationship between SES (education and income levels) and SRH significantly with different degrees of the proportion of mediation. About 20% of the association between income level and SRH went through an indirect path via lifestyle factors, whereas 20%-30% on the model with education. Smoking and BMI contributed more to the indirect effect in the model with education, where PA contributed relatively less. Whereas in the model with income, PA had the greatest contribution, followed by BMI and smoking.

4.1.1 SES and SRH

Income level had a prominent effect on SRH in this study. The odds of reporting poor SRH were nearly half with a higher household income level compared to the lowest household income level. Furthermore, in the highest income level, the odds of poor SRH were lowered by 80% compared to the lowest household income level. Higher household income has been found to be significantly associated with better SRH in multiple studies (20, 28, 55-57). International research, including Norway, showed that higher household equivalent income was associated with better SRH in all the countries (57). In a study that assessed social inequalities in ill health, more reporting of ill health was observed in lower-income and unskilled workers (58). But having poor health is not necessarily equivalent to poor SRH as ill-health is the predictor of SRH, and reporting behavior may be different (59).

The odds of reporting poor SRH with a high level of education compared to the lowest education level were about 35% less. With the highest education level, the odds of reporting poor SRH were 53% less to the lowest education group. Similar, in a study that assessed self-rated health in the general Norwegian population, higher education was a significant predictor of better SRH (2). A globally representative population-based study by WHO on SRH assessment with education also confirms similar findings that higher years of schooling relates to better health (60). Different factors such as age, gender, race/ethnicity are often cofounding the relationship between SES and SRH. With age adjustment and only women included in the study, it is free of confounding by age and gender. There is no information on race/ethnicity in the NOWAC study questionnaire, thus no inference and adjustment could account for that specific parameter. For the immigrant population, the perception of health and use and access of healthcare may differ, resulting in a different relationship between SES and health.

Many proxies for SES, such as current income, wealth, education, occupation, and composite measures, are used in health research. While choosing the most appropriate variable for SES measurement, it is dependent on relevance to the population and outcomes under study (61). The collection of variable information on NOWAC allows the use of both education and income as proxies of SES in a single study. Although this study does not aim to seek the differences in education and income on SRH, household income shows a more prominent effect on differentiating across SES. Similar results were observed in a study exploring the relative strength of income on health compared to that of other socioeconomic measures; income appears to be a better discriminator of health status than education or occupation (55). Other studies show that higher education is a stronger predictor of good health than income or occupation (62). In the Norwegian context, education and income level are a justified proxy of SES, but education is more prevalent as described in the literature trying to quantify socioeconomic inequalities (31, 63). A 2019 report on addressing the social inequalities in Norway's population health pointed out that the wage and income differences have increased in Norway (64). Whereas looking at educational attainment, there is an increment in the number of individuals with higher education, undergraduate level, and graduate-level (65). Women's proportion in attaining graduate-level education increased to a greater extent from 1980 to 2018 (65). This study also reflects that income is better in reflecting the socioeconomic difference in health. Investigating the causal effect of SES on the quality of care under a universal health insurance system showed income was a determinant of the quality of care received (66). Despite public healthcare in Norway, access to care is often affected by SES, people with higher SES opting for private insurance to avoid waiting time to receive care.

4.1.2 Mediation

The total effect of SES on SRH was similar within different education levels and income levels, with BMI and PA, added sequentially into the model with smoking. The proportion of mediation by all the lifestyle factors (smoking, BMI, and PA) on the association between income level and SRH was statistically significant and quite similar across different income levels.

A sequential approach was performed to investigate the proportion of mediation through different lifestyle factors (smoking, BMI, and PA). Based on the relative importance/significance of predictors in studies, smoking, BMI, and PA order were chosen for the sequential approach. To the best of our knowledge, this is the first study to quantify the proportion of mediation through individual lifestyle factors on the association between SES and SRH.

All three models, first with smoking, second with smoking and BMI, and third the joint model with smoking, BMI, and PA have statistically significant estimates for direct, indirect, and total effect for both proxies of SES (education and income level). Inspection of direct and indirect effect estimates with increasing household income levels with smoking as a mediator showed an increasing gradient. An increasing proportion of meditation was observed with higher income levels in the model with smoking and BMI as mediators. This increase in proportion reflects more of the relationship going through an indirect path; the higher income has higher the impact of smoking and BMI on the association between SES and SRH. In contrast, with physical activity, the proportion of mediation was highest in the low-income group compared to the higher income level. This decrease in the proportion of mediation implies the greater extent of the association between SES and SRH acting through a direct path with an increase in household income level resulting in a lower proportion of mediation with PA. Similarly, in the model with education, a similar pattern of increase and decrease in the proportion of mediation was observed across different education levels for all three lifestyle factors.

Smoking and BMI greatly impacted the model with education, where PA contributed relatively less. Whereas in the model with income, PA had the greatest contribution, followed by BMI and smoking. The impact of smoking being a lifestyle factor meditating the association between SES (education) and SRH is justified as the Norwegian public health report in 2018 stated smoking was the strongest lifestyle factor with a social difference where socioeconomic inequality was analyzed through education(12).

In the final model with all lifestyle factors as mediators, The Medflex package provided adjustment between mediators allowing interaction between the lifestyle factors. Individual lifestyle factors had an impact on mediating the association between SES and SRH. With the final model, all lifestyle factors were taken into account; in the model with household income, the proportion of mediation across household income was around 20%. In the model with education, an increasing gradient of the proportion of mediation was observed 20%-30% with increasing education level. There was an increasing gradient in the proportion of mediation across increasing education levels in the model with education and lifestyle factors (smoking, BMI, and physical activity). Our results were confirmed by a study assessing the mediation effect of lifestyle factors on the association between SES and SRH in a Korean population. It showed that regular exercise and underweight significantly mediated the relationship between income and SRH whereas, smoking, underweight, and obese mediated the association between education and SRH on Korean middle-aged and older adults (44). This Korean study supports the individual impact of lifestyle factors on mediating the association between SES and SRH but fails to provide the overall impact of these lifestyle factors taken together. Similar results were observed in the Chinese General Social Survey; lifestyle factors significantly mediated International Socio-Economic Index (ISEX) and health where lifestyle factors were measured with a 6 item Likert scale rather than differentiating individual lifestyle parameters (45). Both studies have a different country and population setting than the Norwegian population. While assessing the comparability within the population, macro factors such as governance, health insurance, and social capital play a role in differentiating SES's impact on health. Even though the issue of direct comparability of the relationship between SES and SRH across the population persists, it supports our results in the direction and impact of lifestyle factors on the association between SES and SRH. However, the prospective cohort design of this study have a clear advantage over the cross sectional design used in both of the studies above. A Finnish study on the influence of behavioral factors on the occupational class difference in health showed that behavioral factors depending on the material factor(income) were about half of their independent effects, which supports the study of this results on lifestyle factors going through income (67). In a British study, the education variable was attenuated by 27% with the addition of lifestyle factors while assessing the association between SES and SRH (46). In a study to assess socioeconomic inequalities in health through lifestyle factors, path analysis showed that higher SES was directly associated with better SRH through healthier lifestyle habits in Spain. In addition, SES can affect incentives or motivation for healthy behaviors. As paying for tobacco cessation aids, joining fitness clubs, and buying more expensive fruits, vegetables, and lean meats comes with certain costs and time, trade-offs in low SES groups can be quite challenging. Inspection of social determinants of inequalities in self-reported health in Europe from European social survey showed in Nordic countries the combination of occupational and behavioral factors explained more social inequalities in health instead of living conditions factors (material and psychosocial factors) (68). In context to Norway, behavioral factors were more prominent factors explaining the social inequalities in general health (68).

Smoking is one of the powerful predictors of health. Smoking had a social difference (69, 70). A study shows maintaining good health practices such as high PA and no smoking are associated with positive health (71). A Swedish study exploring health behaviors and SRH found that smoking, obesity, and underweight were significantly associated with poor SRH(35). Physically active people have lower Coronary Heart Disease (CHD) and CVD rates, which leads to better health (37). Apart from lifestyle factors, the impact of SES on health may operate through the physical environment in which one lives and works and associated exposure to pathogens, carcinogens, and other environmental hazards; the degree of access to social resources and supports; socialization and experiences that influence psychological development (23). These factors and lifestyle factors (health behaviors) may have an individual effect on developing health. Adjustment of these factors would help obtain the exact extent of mediation as some pathways may be mutually related.

The dynamics of the change in the lifestyle factors are quite potent in Norway. Lifetime smoking habits of Norwegian men and women born between 1890 and 1994 showed men had long and intense exposure to cigarettes. After the 1970s, smoking declined in all cohorts of men and women, where women smoking peaked around 20 years later than men in around 1940-1949(72). Now daily smoking prevalence in the general population is around 9% as per statistics Norway with October 2020 estimate (73). Despite the decrease in the proportion of smoking across the population, educational inequality was observed in smoking as the proportion of smokers in the group who only completed lower secondary education is 24 %, but just 5 % in the group who completed tertiary education(12). The proportion of adults with overweight/obesity increased significantly from the mid-1960s among men and mid-1985 among women (12). As BMI change over time in general with age, in NOWAC, women gained weight on an average of 0.5kg/year during a 6year follow-up study, where smoking cessation and decreased PA were the significant factors driving the weight gain (74). The relationship between smoking cessation and PA with weight change highlights the interrelation between different lifestyle factors. Looking at the BMI trend across education groups over the last two decades, there is an increase in BMI in all compulsory education, upper secondary education, and tertiary education group (75). In Norway, the proportion of overweight and obese individuals is also highest among adults in groups with the lowest educational attainment (12). As the present population of Norway has different dynamics in the education, income, and lifestyle factors information collected in the study, the results from this study would be more generalized to older cohorts as used in the study. Despite the change in SES and lifestyle factors, this provides insights on the relationship between SES and SRH and lifestyle factors mediating the relationship between SES and SRH.

4.2 Bias

Inspection of variables used in the study, their self-reported nature brings the question of bias and validity of the data collected. When self-reported information is collected with proper tools, it can provide sufficient and reliable data to study results. In this study, self-reported lifestyle factors, socioeconomic measures, and health are of interest.

4.2.1 Selection bias

Selection bias occurs when the subjects are entered into a study. Observational, as well as cohort studies, are prone to selection bias. Response rate often plays a crucial role in claiming the role of selection bias by raising a question is their difference with information from the non-responders (76). In NOWAC response rate was around 60% and comparable to other population-based studies (54). Response rates may be shaped due to the healthy volunteer effect where people with higher education, good health, resulting in higher SES and SRH. The cohort profile of NOWAC showed higher educated women participated in NOWAC; a higher response rate was observed in northern Norway (54). But it does not necessarily lead to selection bias in the exposure-outcome study. Non-response always creates a black box of information yet to be explored. When accounting for a multivariable study, the missing information on several variables can lead to bias if individuals do not respond to each question. In this study, complete case analysis was performed, which then lead to the loss of some individuals who did not have complete information for the multivariable study. On inspection of missing data due to the need of complete case analysis, similar prevalence of good and poor SRH was observed (not reported), which does not imply the loss of individuals was acceptable, but in general, they were comparable.

4.2.2 Information bias

Information bias can occur if any information used in the study is measured or recorded inaccurately (76). In a cohort, study exposure is collected prospective to the outcome, as the exposure information may have information bias but not the outcome; the degree of misclassification is likely to be the same resulting in non-differential misclassification (76).

A study conducted to assess the validity of self-reported BMI among NOWAC shows self-reported weight and height provide valid rankings for BMI as the cumulative distribution curves for measured and self-reported values closely followed each other (77). But under-reporting of weight among the overweight and obese group was observed (77). In general, BMI is debated as it is not a true measure as it is derived from measures on height and weight. In contrast, other measures, such as body fat or fat distribution, can have a significant advantage over general crude BMI (78). Using BMI does not let us capture the role of fat distribution in the prediction of medically significant mortality and morbidities (78). We do not have any guidelines for the use of specific measures to date, which brings us to a relatively more comfortable measure to get information on a large population scale, BMI.

The nature of question on smoking habits differentiating never, current and former may not capture the dose relationship; information on duration (years of smoking) and intensity (packs smoked per day) of smoking, i.e., packyears, would be more reliable and convenient in establishing the dosage relationship of smoking on health.

As NOWAC assessed PA as a ten category scale from very low to very high with self-report of PA, respondents may have a different perception of the scale. Still, a study conducted on assessing criterion validity has demonstrated the NOWAC scale appears to be valid in ranking PA in the NOWAC population (79). Information on smoking is not validated in NOWAC but a general study in the Norwegian population showed self-reported smoking habits could be trusted (80). In addition, the study's lifestyle factors do not capture the entire spectrum of lifestyle factors that are significantly associated with health.

4.2.3 Confounding

Confounding is the distortion of the association between exposure and outcome by the third variable called confounder; that has to be associated with both exposure and outcome while not being in the causal pathway between said exposure and outcome (81). Residual confounding occurs when the adjustment for confounding factors is performed inadequately. Any unidentified confounder and the known confounder with bias and measurement error leave the space for residual confounding (82). Confounding is also a form of bias. Proper identification and adjustment of confounders are necessary to have a reliable estimate of the results. In general, age, gender, and race/ethnicity are potential confounders. As the study population comprises only females, there cannot be confounding by gender, whereas age is adjusted for. As no information on race/ethnicity were available, that could be adjusted for apart from that immigrant population may behave in a different manner in perceiving their health. As the study could not incorporate the wider dimension of SES by only having education and income separately, other SES measures such as occupation may confound the relationship. Alcohol use, diet, stress, and sleep, other proven lifestyle factors not being accounted for in this study leaves the space for information bias and confounds the association between SES and SRH. Drift hypothesis association reflects the influence of illness on SES rather than SES on illness, as the baseline population is collected at a point of life, functional limitations and chronic conditions are proven to be predictors of SRH, which were not accounted for in this study.

4.3 Methodological considerations

A systematic review of RCTs using mediation analysis showed relatively low use of the counterfactual framework in mediation analysis studies (83). Most of the existing literature on mediation analysis uses traditional approaches such as Baron and Kenny's mediation analysis approach, bound by statistical power and only continuous mediator and outcome limitations (83). Despite different extensions and processes such as Hayes, we cannot use any categorical variable as mediator, outcome, and exposure, which brings to counterfactual's opportunities (53). Incorporating all covariates and adjustment variables in the model will provide sufficient condition for the validation of mediation analysis assumptions(53). However, the lack of sensitivity analysis for RCT limits the assessment of assumptions. Incorporating an inbuilt sequential approach for mediators remains a major limitation on the package where residual plots could be generated. But general goodness of fit for the model with the observational study also seems to be needed as the present study could not simply assess the model fit estimate just by residual plots and ways of validating the assumptions for mediation analysis, which seems to raise a question on the results of the study. Russo et al. discussed inferring causality through counterfactuals in observational studies in the epistemological context highlighted issues with the counterfactual framework (52). Issues such as the counterfactual model measures effects of cause whereas other models concerned instead with causes of effects, undermine a sound empirical basis as the counterpart of the cause are not observed, complex mechanism and critical assessment of parallelism are discussed as a major issue with counterfactuals(52). Despite there are several techniques on mediation analysis, the need for consensus-based planning for the conduct and reporting of mediation analysis was highlighted due to heterogeneous reporting of mediation analysis (83). Nevertheless, the incorporation of counterfactuals has enabled to decompose the direct and indirect effect in situations where traditional approaches would fail to do so.

4.4 Strength and weakness

This study's major strength lies within a nationally representative cohort, which provides a higher degree of statistical power. As most studies are cross-sectional, they are often hindered by temporality and reverse causation. A prospective cohort with only good SRH at baseline resolves the issue of temporality and prevents reverse causation. Also, validated measures of physical activity and BMI adds up to the strength of this study. The use of the counterfactual framework provides an advantage over traditional methods existing in the literature on calculating the proportion of mediation jointly and the impact of the individual mediators.

Despite several strengths, this study also possesses some limitations. One of the limitations of this study arises from its population as the cohort has only a female population. It is not generalizable to the general population and cannot say anything about the male population. Also, with the spectrum of lifestyle factors, all potential lifestyle factors that affect health are not included. Despite having information on diet and alcohol consumption in the cohort due to difficulty in modeling, they were omitted, whereas information on sleep and stress is not collected in the cohort. This study can only point out effect estimates of education and household income separately as proxies of SES, but not as an overall SES measure. Composite measures and life course SES would enable to conclude SES in general. Limitations on the methodological aspect, as the package Medflex used in the study, helps to generate plots to assess model adequacy, but some tests specific to observational study to assess model adequacy and fulfillment of assumptions are missing. Also, selection bias, information bias, and residual confounding are also inbuilt weaknesses to observational or cohort studies of any kind. As income and education would not capture the entire spectrum of SES, interpreting the result from education and income as SES, in general, should be avoided rather just state as proxies of SES. One must consider how potential important unmeasured SES factors may affect conclusions(84).

4.5 Conclusion

Data from NOWAC showed that socioeconomic variables (Education and Household income level) had a statistically significant association with SRH. Better education and higher income level were associated with good SRH. Mediation analysis showed each lifestyle factor (smoking, BMI, and PA) individually and jointly mediated the association between SES and SRH in a statistically significant way. Looking at the proportion of mediation, smoking and BMI greatly contributed to the model with education, where PA contributed relatively less. Whereas in the model with income, PA had the greatest contribution, followed by BMI and smoking. The joint model with all lifestyle factors (smoking, BMI, and PA) showed an increasing gradient of the proportion of mediation with education as a proxy of SES with range (20%-30%) whereas with household income level as a proxy of SES similar proportion of mediation around 20% was observed. Any interventions focusing on improving the healthy lifestyle of the population should comprehend the impact of SES.

4.6 Future recommendation and implications of the study

Future studies incorporating a composite measure of SES or life course model would help to generalize as a whole picture of SES rather than individual proxies of SES, which is more prevalent in recent literature. But more issues may arise if the composite measures information is collected improperly; confounding plays a bigger role in such cases. A cautious approach with these composite measures could add substantial value and missing pictures due to individual SES proxy. Furthermore, with a broader range of lifestyle factors, information can enable us to have a more reliable estimate. With a better measure of SES and a wider array of lifestyle factors, the mediation effect estimate could be achieved close to a true estimate within the population. Also, mediation of the association between SES and SRH may be operating through other channels as physical environment, social environment, social support, and psychological (cognitive, emotional, and social capabilities) development. The inclusion of these factors will adjust the extent of mediation if some effects are translated with common pathways.

Despite the value addition that future studies can add to the results, we can have some practical implications from this study’s results. This study confirms as supports the existing literature on healthy behaviors tends to accumulate with higher SES. A substantial amount of the effect of SES on SRH was mediated via lifestyle factors.

References

1. Idler EL, Benyamini Y. Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies. Journal of Health and Social Behavior. 1997;38(1):21-37.
2. Bonsaksen T, Ekeberg Ø, Skogstad L, Heir T, Grimholt TK, Lerdal A, et al. Self-rated global health in the Norwegian general population. Health and Quality of Life Outcomes. 2019;17(1):188.
3. Idler EL, Hudson SV, Leventhal H. The Meanings of Self-Ratings of Health: A Qualitative and Quantitative Approach. Research on Aging. 1999;21(3):458-76.
4. Idler EL, Angel RJ. Self-rated health and mortality in the NHANES-I Epidemiologic Follow-up Study. Am J Public Health. 1990;80(4):446-52.
5. Sundquist J, Johansson SE. Self reported poor health and low educational level predictors for mortality: a population based follow up study of 39,156 people in Sweden. Journal of Epidemiology and Community Health. 1997;51(1):35.
6. Fayers PM, Sprangers MAG. Understanding self-rated health. The Lancet. 2002;359(9302):187-8.
7. Shields M, Shooshtari S. Determinants of self-perceived health. Health reports. 2001;13(1):35-52.
8. Jylhä M, Guralnik JM, Ferrucci L, Jokela J, Heikkinen E. Is self-rated health comparable across cultures and genders? The Journals of Gerontology Series B: Psychological Sciences and Social Sciences. 1998;53(3):S144-S52.
9. Cullati S, Mukhopadhyay S, Sieber S, Chakraborty A, Burton-Jeangros C. Is the single self-rated health item reliable in India? A construct validity study. BMJ Global Health. 2018;3(6):e000856.
10. Gallagher JE, Wilkie AA, Cordner A, Hudgens EE, Ghio AJ, Birch RJ, et al. Factors associated with self-reported health: implications for screening level community-based health and environmental studies. BMC Public Health. 2016;16(1):640.
11. Fylkesnes K, Førde OH. The Tromsø study: Predictors of self-evaluated health—Has society adopted the expanded health concept? Social Science & Medicine. 1991;32(2):141-6.
12. Norwegian Institute of Public Health. Public Health Report: Health Status in Norway 2018. Oslo: Norwegian Institute of Public Health; 2018.
13. Lorem GF, Schirmer H, Wang CEA, Emaus N. Ageing and mental health: changes in self-reported health due to physical illness and mental health status with consecutive cross-sectional analyses. BMJ Open. 2017;7(1):e013629-e.
14. Wu S, Wang R, Zhao Y, Ma X, Wu M, Yan X, et al. The relationship between self-rated health and objective health status: a population-based study. BMC public health. 2013;13:320-.
15. Idler E, Cartwright K. What Do We Rate When We Rate Our Health? Decomposing Age-related Contributions to Self-rated Health. Journal of Health and Social Behavior. 2018;59(1):74-93.
16. Schnittker J. When Mental Health Becomes Health: Age and the Shifting Meaning of Self-Evaluations of General Health. The Milbank Quarterly. 2005;83(3):397-423.
17. Thomas ML, Kaufmann CN, Palmer BW, Depp CA, Martin AS, Glorioso DK, et al. Paradoxical Trend for Improvement in Mental Health With Aging: A Community-Based Study of 1,546 Adults Aged 21-100 Years. J Clin Psychiatry. 2016;77(8):e1019-e25.
18. Chandola T, Ferrie J, Sacker A, Marmot M. Social inequalities in self reported health in early old age: follow-up of prospective cohort study. BMJ. 2007;334(7601):990.

19. Gomez-Baya D, Salinas-Perez JA, Rodero-Cosano ML, Alvarez-Galvez J. Socioeconomic Inequalities in Health Through Lifestyles: Analysing Gender and Age Differences in Andalusia, Spain. Journal of Community Health. 2020;45(4):836-45.
20. Lim W-Y, Ma S, Heng D, Bhalla V, Chew SK. Gender, ethnicity, health behaviour & self-rated health in Singapore. BMC public health. 2007;7:184-.
21. Wilson K, Elliott SJ, Eyles JD, Keller-Olaman SJ. Factors Affecting Change Over Time in Self-reported Health. Canadian Journal of Public Health / Revue Canadienne de Sante'e Publique. 2007;98(2):154-8.
22. Oakes M. Measuring Socioeconomic Status. e-Source. Behavioral & Social Sciences Research. 2011.
23. Adler NE, Boyce T, Chesney MA, Cohen S, Folkman S, Kahn RL, et al. Socioeconomic status and health: The challenge of the gradient. American Psychologist. 1994;49(1):15-24.
24. Adler NE, Ostrove JM. Socioeconomic Status and Health: What We Know and What We Don't. Annals of the New York Academy of Sciences. 1999;896(1):3-15.
25. Prus SG. Comparing social determinants of self-rated health across the United States and Canada. Social Science & Medicine. 2011;73(1):50-9.
26. Szaflarski M, Cubbins LA. Self-Reported Health in Poland and the United States: A Comparative Analysis of Demographic, Family and Socioeconomic Influences. Health. 2004;8(1):5-31.
27. Kawada T, Wakayama Y, Katsumata M, Inagaki H, Otsuka T, Hirata Y, et al. Patterns in self-rated health according to age and sex in a Japanese national survey, 1989–2004. Gender Medicine. 2009;6(1):329-34.
28. Cislaghi B, Cislaghi C. Self-rated health as a valid indicator for health-equity analyses: evidence from the Italian health interview survey. BMC Public Health. 2019;19(1):533.
29. Ross CE, Wu C-L. Education, Age, and the Cumulative Advantage in Health. Journal of Health and Social Behavior. 1996;37(1):104-20.
30. Subramanian SV, Huijts T, Avendano M. Self-reported health assessments in the 2002 World Health Survey: how do they correlate with education? Bulletin of the World Health Organization. 2010;88:131-8.
31. Kurtze N, Eikemo TA, Kamphuis CBM. Educational inequalities in general and mental health: differential contribution of physical activity, smoking, alcohol consumption and diet. European Journal of Public Health. 2012;23(2):223-9.
32. Schoenborn CA. Health habits of U.S. adults, 1985: the "Alameda 7" revisited. Public Health Rep. 1986;101(6):571-80.
33. Warren GW, Alberg AJ, Kraft AS, Cummings KM. The 2014 Surgeon General's report: "The Health Consequences of Smoking–50 Years of Progress": A paradigm shift in cancer care. Cancer. 2014;120(13):1914-6.
34. Banks E, Joshy G, Korda RJ, Stavreski B, Soga K, Egger S, et al. Tobacco smoking and risk of 36 cardiovascular disease subtypes: fatal and non-fatal outcomes in a large prospective Australian study. BMC Medicine. 2019;17(1):128.
35. Manderbacka K, Lundberg O, Martikainen P. Do risk factors and health behaviours contribute to self-ratings of health? Soc Sci Med. 1999;48(12):1713-20.
36. Moore SC, Lee IM, Weiderpass E, Campbell PT, Sampson JN, Kitahara CM, et al. Association of Leisure-Time Physical Activity With Risk of 26 Types of Cancer in 1.44 Million Adults. JAMA Intern Med. 2016;176(6):816-25.
37. Shiroma EJ, Lee I-M. Physical activity and cardiovascular health: lessons learned from epidemiological studies across age, gender, and race/ethnicity. Circulation. 2010;122(7):743-52.

38. Wardle J, Steptoe A. Socioeconomic differences in attitudes and beliefs about healthy lifestyles. Journal of Epidemiology & Community Health. 2003;57(6):440-3.
39. Hanson MD, Chen E. Socioeconomic status and health behaviors in adolescence: a review of the literature. Journal of behavioral medicine. 2007;30(3):263.
40. Shankar A, McMunn A, Steptoe A. Health-related behaviors in older adults relationships with socioeconomic status. Am J Prev Med. 2010;38(1):39-46.
41. Lynch JW, Kaplan GA, Salonen JT. Why do poor people behave poorly? Variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic lifecourse. Social science & medicine. 1997;44(6):809-19.
42. Xue X, Cheng M. Social capital and health in China: exploring the mediating role of lifestyle. BMC Public Health. 2017;17(1):863.
43. Cockerham WC. Health lifestyles: Bringing structure back. The new Blackwell companion to medical sociology2009.
44. Kim J. The Mediating Effects of Lifestyle Factors on the Relationship between Socioeconomic Status and Self-Rated Health among Middle-Aged and Older Adults in Korea. International journal of aging & human development. 2011;73:153-73.
45. Wang J, Geng L. Effects of Socioeconomic status on physical and psychological health: lifestyle as a mediator. International journal of environmental research and public health. 2019;16(2):281.
46. Contoyannis P, Jones AM. Socio-economic status, health and lifestyle. Journal of Health Economics. 2004;23(5):965-95.
47. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology. 1986;51(6):1173-82.
48. Jacob R. Neiheisel. The SAGE Encyclopedia of Communication Research Methods. Thousand Oaks, California: SAGE Publications, Inc; 2017 2020/11/25.
49. Hayes AF. PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling. University of Kansas, KS; 2012.
50. VanderWeele TJ. Mediation Analysis: A Practitioner's Guide. Annu Rev Public Health. 2016;37:17-32.
51. Höfler M. Causal inference based on counterfactuals. BMC Medical Research Methodology. 2005;5(1):28.
52. Russo F, Wunsch G, Mouchart M. Inferring Causality through Counterfactuals in Observational Studies - Some Epistemological Issues. Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique. 2011;111(1):43-64.
53. Steen J, Loeys T, Moerkerke B, Vansteelandt S. Medflex: an R package for flexible mediation analysis using natural effect models. Journal of Statistical Software. 2017;76(11).
54. Lund E, Dumeaux V, Braaten T, Hjartåker A, Engeset D, Skeie G, et al. Cohort Profile: The Norwegian Women and Cancer Study—NOWAC—Kvinner og kreft. International Journal of Epidemiology. 2007;37(1):36-41.
55. Benzeval M, Judge K, Shouls S. Understanding the Relationship between Income and Health: How Much Can be Gleaned from Cross-sectional Data? Social Policy & Administration. 2001;35(4):376-96.
56. Fritzell J, Nermo M, Lundberg O. The impact of income: assessing the relationship between income and health in Sweden. Scand J Public Health. 2004;32(1):6-16.
57. Mackenbach JP, Martikainen P, Looman CW, Dalstra JA, Kunst AE, Lahelma E, et al. The shape of the relationship between income and self-assessed health: an international study. International Journal of Epidemiology. 2004;34(2):286-93.

58. Dahl E. Social inequalities in ill-health: the significance of occupational status, education and income-results from a Norwegian survey. Sociology of Health & Illness. 1994;16(5):644-67.
59. Layes A, Asada Y, Kephart G. Whiners and deniers – What does self-rated health measure? Social Science & Medicine. 2012;75(1):1-9.
60. Subramanian S, Huijts T, Avendano M. Self-reported health assessments in the 2002 World Health Survey: how do they correlate with education? Bulletin of the World Health Organization. 2010;88:131-8.
61. Shavers VL. Measurement of socioeconomic status in health disparities research. J Natl Med Assoc. 2007;99(9):1013-23.
62. Winkleby MA, Jatulis DE, Frank E, Fortmann SP. Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. Am J Public Health. 1992;82(6):816-20.
63. Krokstad S. Socioeconomic inequalities in health and disability.: Social epidemiology in the Nord-Trøndelag health study (HUNT), Norway. 2004.
64. Arntzen A, Bøe T, Dahl E, Drange N, Eikemo TA, Elstad JI, et al. 29 recommendations to combat social inequalities in health. The Norwegian Council on Social Inequalities in Health. Scandinavian journal of public health. 2019;47(6):598-605.
65. Nygård G. Facts about education in Norway 2020. Key figures 2018: Statistics Norway; 2020 07/01/2020. 36 p.
66. Cho HE, Wang L, Chen J-S, Liu M, Kuo C-F, Chung KC. Investigating the causal effect of socioeconomic status on quality of care under a universal health insurance system - a marginal structural model approach. BMC Health Services Research. 2019;19(1):987.
67. Laaksonen M, Roos E, Rahkonen O, Martikainen P, Lahelma E. Influence of material and behavioural factors on occupational class differences in health. Journal of Epidemiology and Community Health. 2005;59(2):163.
68. Balaj M, McNamara CL, Eikemo TA, Bambra C. The social determinants of inequalities in self-reported health in Europe: findings from the European social survey (2014) special module on the social determinants of health. European Journal of Public Health. 2017;27(suppl_1):107-14.
69. Power C, Matthews S, Manor O. Inequalities in self-rated health: explanations from different stages of life. The Lancet. 1998;351(9108):1009-14.
70. Gregoraci G, van Lenthe FJ, Artnik B, Bopp M, Deboosere P, Kovács K, et al. Contribution of smoking to socioeconomic inequalities in mortality: a study of 14 European countries, 1990–2004. Tobacco Control. 2017;26(3):260.
71. Belloc NB, Breslow L. Relationship of physical health status and health practices. Preventive Medicine. 1972;1(3):409-21.
72. Lund I, Lund KE. Lifetime smoking habits among Norwegian men and women born between 1890 and 1994: a cohort analysis using cross-sectional data. BMJ Open. 2014;4(10):e005539.
73. Statistics Norway. Tobacco, alcohol and other drugs 2020 [updated 01/10/2020. Available from: https://www.ssb.no/en/helse/statistikker/royk.
74. da Silva M, Weiderpass E, Licaj I, Rylander C. Factors Associated with High Weight Gain and Obesity Duration: The Norwegian Women and Cancer (NOWAC) Study. Obesity Facts. 2018;11:381-92.
75. Norwegian Institute of Public Health. Overweight and obesity (25-79 years), self-reported, by educational attainment (NH) 2020 [Available from: http://norgeshelsa.no/norgeshelsa/index.jsp?headers=AAR&VEKTslice=3_4_OvervektOgFedme&stubs=GEO&stubs=Utdann&measure=common&virtualslice=Adjusted_value&layers=VEKT&layers=virtual&GEOsubset=0&VEKTsubset=3_4_OvervektOgFedme&study=http%3

A%2F%2F10.1.5.31%3A80%2Fobj%2FfStudy%2FKMI-LKU-L-UTD&mode=cube&virtualsubset=Adjusted_value&v=2&AARsubset=1998_1998+-+2015_2015&measuretype=4&cube=http%3A%2F%2F10.1.5.31%3A80%2Fobj%2FfCube%2FKMI-LKU-L-UTD_C1&Utdannsubset=1_Grunns+-+3_Univer&top=yes.

76. Bhopal RS. Concepts of epidemiology : integrating the ideas, theories, principles, and methods of epidemiology. Oxford, England: Oxford University Press; 2016.

77. Skeie G, Mode N, Henningsen M, Borch KB. Validity of self-reported body mass index among middle-aged participants in the Norwegian Women and Cancer study. Clin Epidemiol. 2015;7:313-23.

78. Nuttall FQ. Body Mass Index: Obesity, BMI, and Health: A Critical Review. Nutrition Today. 2015;50(3):117-28.

79. Borch KB, Ekelund U, Brage S, Lund E. Criterion validity of a 10-category scale for ranking physical activity in Norwegian women. International Journal of Behavioral Nutrition and Physical Activity. 2012;9(1):2.

80. Foss O, Haug K, Hesla P, Lund-Larsen P, Vasli L. Can we rely on self-reported smoking habits? Tidsskrift for den Norske Laegeforening: Tidsskrift for Praktisk Medicin, ny Raekke. 1998;118(14):2165-8.

81. Gordis L. Epidemiology. 5th ed. Philadelphia: Elsevier Saunders; 2014.

82. Fewell Z, Davey Smith G, Sterne JAC. The Impact of Residual and Unmeasured Confounding in Epidemiologic Studies: A Simulation Study. American Journal of Epidemiology. 2007;166(6):646-55.

83. Vo T-T, Superchi C, Boutron I, Vansteelandt S. The conduct and reporting of mediation analysis in recently published randomized controlled trials: results from a methodological systematic review. Journal of Clinical Epidemiology. 2020;117:78-88.

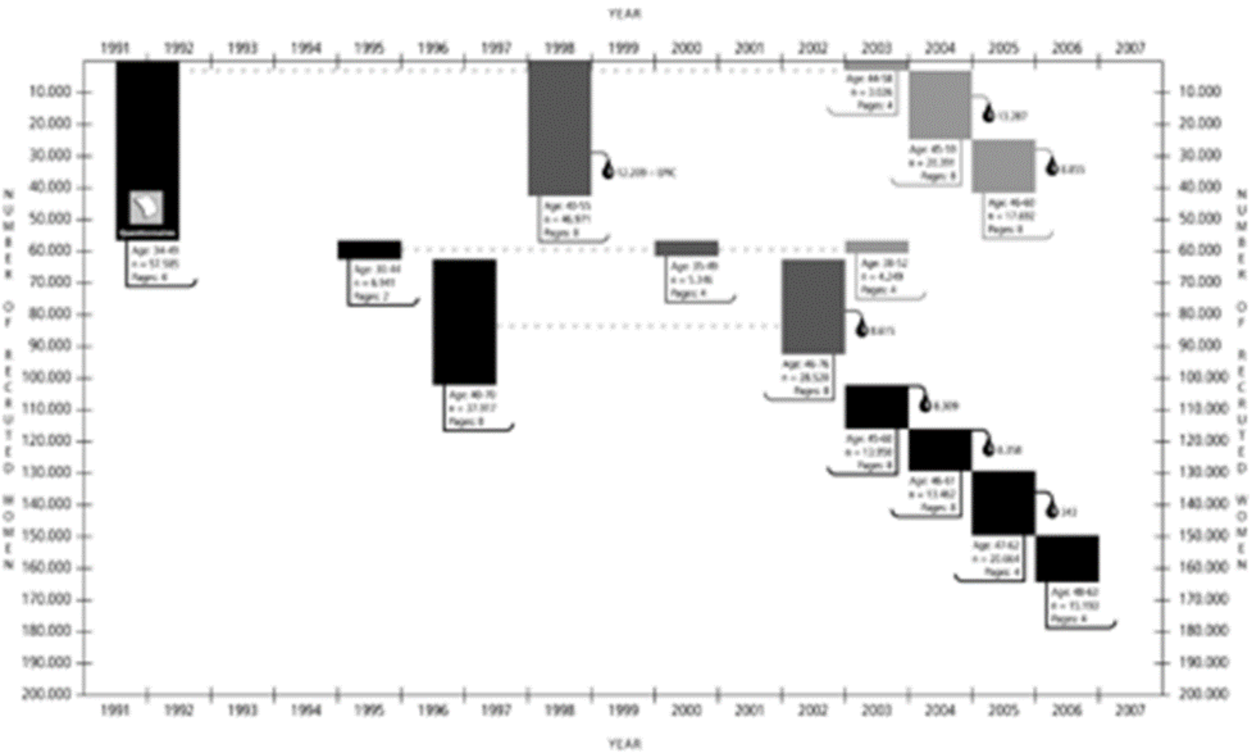
84. Braveman PA, Cubbin C, Egerter S, Chideya S, Marchi KS, Metzler M, et al. Socioeconomic Status in Health ResearchOne Size Does Not Fit All. JAMA. 2005;294(22):2879-88.

Appendix

Appendix 1

Inclusion flow in the Norwegian Women and Cancer study

Number of women recruited (black boxes), timing of second (grey boxes) and third questionnaire mailings (light boxes) and collection of blood samples within the Norwegian Women and Cancer (NOWAC) study according to year of enrolment, age and length of questionnaires with number of blood samples in the EPIC and post-genome cohort biobanks.



Reference:

Lund E, Dumeaux V, Braaten T, Hjartåker A, Engeset D, Skeie G, et al. Cohort Profile: The Norwegian Women and Cancer Study—NOWAC—Kvinner og kreft. International Journal of Epidemiology. 2007;37(1):36-41.

Appendix 2

The counterfactual framework

This is a description of the counterfactual framework that provides information on the estimation of direct and indirect effects. As the counterfactual framework is relatively used less in literature; this information is targeted to understand how counterfactuals provide the estimates on statistical grounds.

Let Yi(x) denote the potential outcome for subject i that had been observed if, possibly contrary to the fact, i had been assigned to treatment (or exposure level) x.

For a binary exposure (with X = 1 for the exposed and X = 0 for the unexposed), the individual-level causal effect can then be expressed by comparing Yi(1) to Yi(0), whereas the population average total causal effect can be expressed as $E\{Y(1) - Y(0)\}$. Similarly, direct and indirect effects have been defined in terms of counterfactual outcomes. For instance, the definition of the so-called controlled direct effect reflects the traditional notion of measuring the effect of exposure while fixing the mediator M at the same value m for all subjects.

Using counterfactual notation, this effect can be expressed as $CDE(m) = E\{Y(1, m) - Y(0, m)\}$, where Y (x, m) denotes the potential outcome that would have been observed under exposure level x and mediator value m. That invoked so-called composite or nested counterfactuals, $Y(x, M(x^*))$.

For instance, the (pure) natural direct effect $NDE(0) = E\{Y(1, M(0)) - Y(0, M(0))\}$ expresses the expected exposure-induced change in outcome when keeping the mediator fixed at the value that had naturally been observed if unexposed.

By considering potential intermediate outcomes $M(x^*)$ rather than a fixed mediator value m, these authors offered a 4 Medflex: flexible mediation analysis in R definition of direct effect that both allows for natural variation in the mediator and provides a complementary operational definition for the indirect effect (which the definition of the controlled direct effect does not). That is, under the composition assumption, which states that $Y(x, M(x)) = Y(x)$, the difference between the average total effect $E\{Y(1) - Y(0)\}$ and the (pure) natural direct effect yields an expression for the (total) natural indirect effect $NIE(1) = E\{Y(1, M(1)) - Y(1, M(0))\}$. This reflects the expected difference in outcome if all subjects were exposed but their mediator value had changed to the value it would take if unexposed.

Total effect of treatment:

= E [Y (1, (M (1))] – E [Y (0, M (0))]

= E [Y (1, (M (1))] – E [Y (1, M (0))] + (E [Y (1, M (0))] – E [Y(0, M(0))])

= Natural indirect effect + natural direct effect

Reference:

Steen J, Loeys T, Moerkerke B, Vansteelandt S. Medflex: an R package for flexible mediation analysis using natural effect models. Journal of Statistical Software. 2017;76(11).