Faculty of Health Sciences

HEL-8032 Register -and biobank epidemiology

Spring 2019

Candidate no: 1

Question 1.

1. *What study design is applied in Zauber's study?*
    a. RCT
    b. Cohort
    c. Case-Control
    d. Patient series

2. *Fill in the PICO-table for Zauber's study?*

P	**I**	**C**	**O**
Patients referred for initial colonoscopy (between 1980 and 1990) at NPS clinical centers who had polyps (adenomas(n=2602) and nonadenomas(n=773)).	Colonoscopic polypectomy	Patients with adenomas removed was compared with the the general population, as estimated from the Surveillance Epidemiology and End Results (SEER) Program, and with patients with nonadenomatous polyps	Among 2602 patients who had adenomas removed during participation in the study, after a median of 15.8 years, 1246 patients had died from any cause and 12 had died from colorectal cancer. Given an estimated 25.4 expected deaths from colorectal cancer in the general population, the standardized incidence-based mortality ratio was 0.47 (95% confidence interval [CI], 0.26 to 0.80) with colonoscopic polypectomy, suggesting a 53% reduction in mortality. Mortality from colorectal cancer was similar among patients with adenomas and those with nonadenomatous polyps during the first 10 years after polypectomy (relative risk, 1.2; 95% CI, 0.1 to 10.6).

3. *Are the two methods used comparable?*

    Zauber's study is an observational study, while Quintero's study is a randomized clinical trial. The observational study was not able to control for all of the confounders due to the non-randomization approach. The two studies have different study populations. For the observational study, 'a small number of trained endoscopists performed the colonoscopies

according to a study protocol that required examination to the cecum, adequate preparation, careful inspection of the colon, and removal of all identified polyps, features that are consistent with reports of high-quality performance', which is not the real-life case. The two studies have different follow-up quality, ie, the observational study had generally complete follow-up and the randomized study had low participation and poor compliance. The two studies used different reference group, ie, the observational study used general population as the reference group while the randomized study compared the colonoscopy group with the FIT group.

The two studies arrive at the similar conclusions, although they used two different methods. However, I would not say the two methods are comparable. I would say that the observational study and the randomized clinical trial can support each other and can be used synergistically to obtain more and better information about the research question.

4. *What bias are we discussing in question 3*?

a. Attrition b) detection c) withdrawal d) diagnostic

5. *If an RCT finds a difference in mortality after 10 years, what checklist question is the most critical one to trust the results?*

   a. Was group allocation done with a valid randomization procedure?
   b. Did all participants have a valid follow-up?
   c. Were the randomized group alike at study start?
   d. Were the procedures/methods for follow-up alike for both groups?

Question 2.

Analysis plan for body mass index and risk for breast cancer

1 Background

Obesity is regarded as a global epidemic, and women are more likely to be overweight and obese than their male counterparts.[1] Breast cancer is the most commonly diagnosed cancer among women.[2] Yet, to our knowledge, very little is known about the role of body mass index (BMI) in relation to the risk of breast cancer.

2 Objective

The aim of this study is to examine the association between BMI and the risk of breast cancer in Swedish women and to test if

1. There is an association between obesity and risk of breast cancer compared to normal weight
2. There is an association between overweight and risk of breast cancer compared to normal weight
3. There is an association between underweight and risk of breast cancer compared to normal weight

3 Notation

BMI: body mass index, WLH: Swedish Women's Lifestyle and Health cohort.

4 Study population

Population are from the Swedish Women's Lifestyle and Health (WLH) cohort, which includes 49,259. women in the Uppsala Health Care Region, which comprises approximately one-sixth of the Swedish population. The women were between 29 and 49 years old at the initiation of the cohort between 1991 and 1992. Details of the study design have been described elsewhere.[3]

4.1 Exclusion criteria

The following subjects will be excluded from the analysis:

1. Subject with missing data on BMI
2. Subjects with recorded breast cancer before cohort enrolment

5 Data management

5.1 Datasets and data sources

In the analyzes we will utilize the following datasets:

1. WLH cohort
2. Swedish cancer register[4]

5.2 Definition of breast cancer

Incident breast cancer will be ascertained from the Swedish cancer register from 1st September 1992 when the regional breast cancer register in Uppsala was set up, up to 12th May 2019.

5.3 Variables to analyze

Below are the variables to be considered in the analyses listed:

1. Age at enrollment (years)
2. Weight (kg) and height (m)
3. Diagnosis of breast cancer from Swedish cancer register
4. Parity (0, 1, 2, 3+)
5. Age at first birth (years)
6. Age at menarche (years)
7. Use of HRT (ever/never)
8. Menopausal status (pre-/postmenopausal)
9. History of breast cancer in mother/sisters (yes/no)
10. Duration of breastfeeding (months)
11. Oral contraceptive use(never/former/current)
12. Smoking status (never/former/current)
13. Education (0–10 years/11–13 years/>13 years)
14. Physical activity (low/moderate/high)
15. Alcohol intake (g/day)
16. Date for emigration
17. Date of death
18. Date of diagnosis of breast cancer

5.4 Derived variables

The following set of variables are calculated given the variables above:

1. BMI: weight (kg) ÷ height^2 (m^2)
2. BMI_cat: BMI will be divided into 4 categories: <18.5 kg/m^2(underweight), 18.5 to 24.9 kg/m^2 (normal weight), 25 to 29.9 kg/m^2 (overweight), and ≥30 kg/m^2 (obese).
3. Cens: 1 or 0 of the subject has been censored or not
4. Event: 1 or 0 if the subject has been diagnosed with breast cancer
5. Time: days from inclusion to either incident breast cancer or censoring due to death from other causes, emigration or end of follow-up.
6. BMI*Age: Interaction term between BMI and age at enrollment (BMI × age)
7. BMI*Phys: Interaction term between BMI and physical activity (BMI × physical activity)

6 Plans for statistical analyses

6.1 Statistical analyzes

6.1.1 Design

The study will be considered as a prospective cohort design. Participants will be followed-up from baseline (1991-1992) to 2019.

6.1.2 Statistical models

Cox proportional regression models will be used to calculate the hazard ratios of breast cancer for four categories of BMI. BMI will be used as the independent variable and breast cancer will be used as the dependent variable. Women who had BMI between 18.5 to 24.9 kg/m^2 were considered as the reference group.

Women with missing values on any of the covariates included in any of the fitted models will be excluded from all the analyses. For all models the attained age will be used as time scale.

To assess effect modification of age at enrollment and physical activity, we will conducte analyses stratified by age and physical activity, respectively.

To explore dose-dependent associations, we will also present the hazard ratios for all values of BMI using spline-based curves.[5]

The proportional hazards assumption will be tested by comparing -ln-ln survival curves and by performing tests on Schoenfeld residuals for each covariate.

Response: Breast cancer

f1.....: BMI_cat

f2.....: f1+ Smoking status + Education + Physical activity +Alcohol intake

f3.....: f2 + + Parity + Age at first birth + Age at menarche + Use of HRT + Menopausal status + History of breast cancer in first-degree relatives +Duration of breastfeeding + Oral contraceptive use

6.1.3 Table

The following tables will be produced:

Table 1. Characteristics of the study population

	Body mass index			
	<18.5 kg/m^2	**18.5-24.9 kg/m^2**	**25-29.9 kg/m^2**	**≥30 kg/m^2**
Age, years (mean±SD)				
Smoking status, n (%)				
never				
former				
current				
Education ,n (%)				
0–10 years				
11–13 years				
>13 years				
Physical activity, n (%)				
low				
moderate				
high				
Alcohol intake g/day, (mean±SD)				
University, n (%)				
Physically inactive, n (%)				
Unmarried, n (%)				
Parity				
nulliparous				
one child				
two children				
three or more children				
Age at first birth, years (mean±SD)				
Age at menarche, years (mean±SD)				
Have used HRT, n (%)				
Post-menopause, n (%)				

History of breast cancer in mother/sisters, n (%)				
Duration of breastfeeding, months (mean±SD)				
Oral contraceptive use				
never				
former				
current				

Values, mean ± standard deviation or number (percent).

Table 2. Hazard ratios (95% confidence intervals) for breast cancer during follow up according to categories of body mass index

	Events	Incidence rate[#]	Crude HR	95% CI	Adjusted HR[a]	95% CI	Adjusted HR[b]	95% CI
Body mass index (kg/m^2)								
<18.5								
18.5-24.9			1.0	Ref	1.0	Ref	1.0	Ref
25-29.9								
≥30								

[#]Incidence rate per 1000 persons-years

HR: hazard ratios, CI: confidence intervals.

HR[a]: adjusted for smoking status, education, physical activity and alcohol intake.

HR[b]: adjusted for smoking status, education, physical activity and alcohol intake, parity, age at first birth, age at menarche, use of HRT, menopausal status, history of breast cancer in mother/sisters, duration of breastfeeding and oral contraceptive use.

Table 3. HRs and 95% CIs of developing breast cancer according to BMI, stratified on age at start of follow-up

	Age 30–39 yrs	**Age 40–49 yrs**	**P value for interaction BMI ×age**
	HR(95%CI)		
BMI (kg/m^2)			
<18.5			
18.5-24.9	1.0 (Ref)	1.0 (Ref)	
25-29.9			
≥30			

Models were adjusted for smoking status, education, physical activity and alcohol intake, parity, age at first birth, age at menarche, use of HRT, menopausal status, history of breast cancer in mother/sisters, duration of breastfeeding and oral contraceptive use.

Table 4. HRs and 95% CIs of developing breast cancer according to BMI, stratified on physical activity at start of follow-up

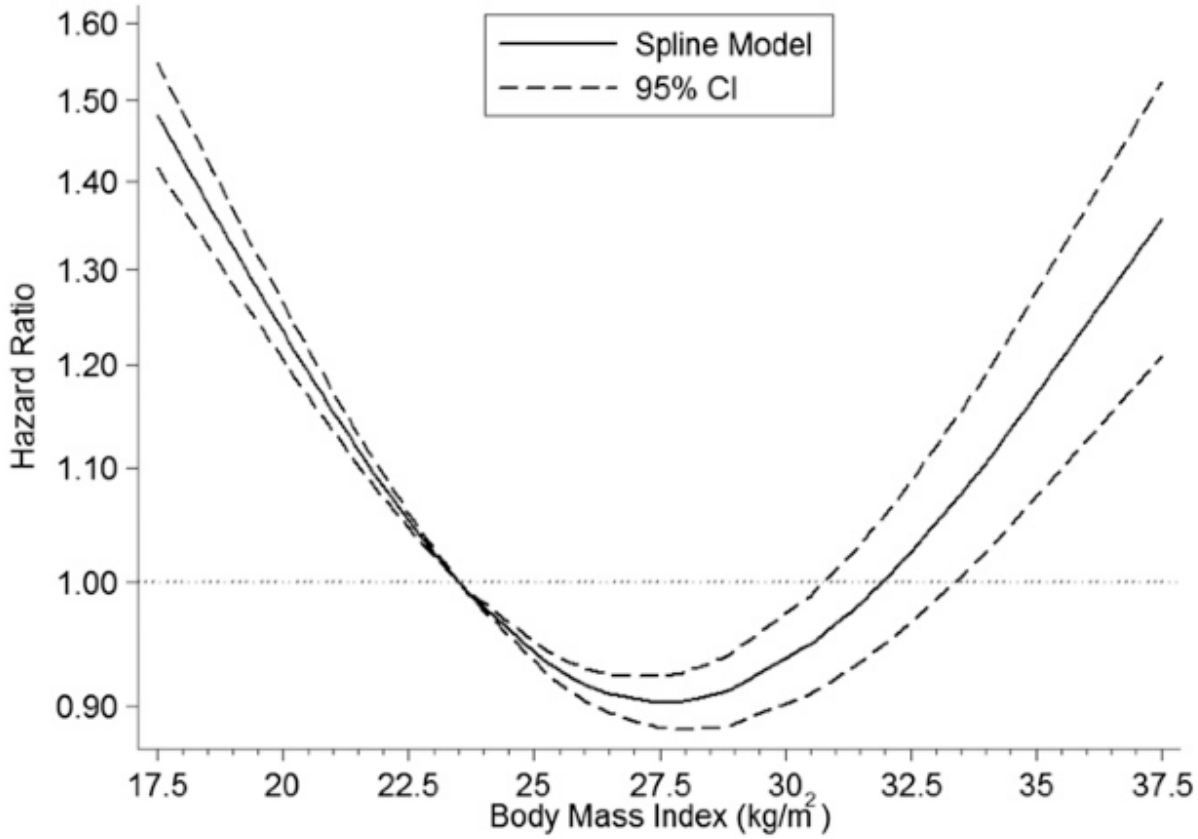
	Low physical activity	**Moderate physical activity**	**High physical activity**	**P value for interaction BMI × physical activity**
	HR (95%CI)			
BMI (kg/m^2)				
<18.5				
18.5-24.9	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)	
25-29.9				
≥30				

Models were adjusted for smoking status, education, physical activity and alcohol intake, parity, age at first birth, age at menarche, use of HRT, menopausal status, history of breast cancer in mother/sisters, duration of breastfeeding and oral contraceptive use.

6.1.4 Figure

The following figure will be produced:

Figure 1: Multivariable adjusted spline curves for dose-response relation between body mass index and risk of breast cancer.



Restricted cubic spline model adjusted for smoking status, education, physical activity and alcohol intake, parity, age at first birth, age at menarche, use of HRT, menopausal status, history of breast cancer in mother/sisters, duration of breastfeeding and oral contraceptive use.

7 References

1. Mitchell S, Shaw D. The worldwide epidemic of female obesity. *Best Practice & Research Clinical Obstetrics & Gynaecology*. 2015;29(3):289-299.
2. McPherson K, Steel C, Dixon J. ABC of breast diseases: breast cancer—epidemiology, risk factors, and genetics. *BMJ*. 2000;321(7261):624.
3. Roswall N, Sandin S, Adami H-O, Weiderpass E. Cohort profile: the Swedish Women's Lifestyle and Health cohort. *International journal of epidemiology*. 2015;46(2):e8-e8.
4. Mattsson B, Wallgren A. Completeness of the Swedish cancer register non-notified cancer cases recorded on death certificates in 1978. *Acta Radiologica: Oncology*. 1984;23(5):305-313.
5. Sleeper LA, Harrington DP. Regression splines in the Cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association*. 1990;85(412):941-949.

Question 3. *Describe three challenges and one opportunity when conducting health research that involve biomarkers.*

Challenges:

1). Ethico-legal challenges: public understanding of biobanks; rational approaches to ethical review; maintenance of privacy and security; effective consent.

2). Collaboration and standardization: more efforts are needed for standardisation and co-ordination of activities at national, regional and international levels.

3). Sustainability concerns: biobanks could be staggeringly expensive. In an environment of rapid technological change this has proven a demanding task.

Opportunity:

Biomarkers are a key resource in research focusing on the association of genetic background, lifestyle and environmental determinants with various complex diseases and health traits.

Question 4*. List and briefly describe methods used to validate registry data.*

- checking the correctness of data type
- checking the degree of coverage (ie, the proportion of the population encompassed by a registry)
- checking the degree of completeness (ie, whether information is actually collected from the persons who are included in the registry.)
- checking the use of an updated and sufficiently detailed coding system,
- checking the presence of data,
- checking logical relations in the data
- comparing the registry data with the original source, for example, patient medical journals