Milk and fermented milk intake and Ischemic Heart Disease (IHD)

Objective

The aim of the study is to investigate associations between time updated information of milk and fermented milk consumption and risk of ischemic heart disease (IHD) as primary outcome and acute myocardial infarction (MI) as secondary outcome.

Ischemic heart disease (IHD) is a consequence of both genetic and environmental influences[1] and is the leading cause of years lost in Europe and globally.[2,3] While survival after a diagnosis of IHD has improved dramatically and is suggested to be largely explained by improved treatments,[4,5] the 40% reduction in age-standardized IHD incidence during the last two decades in Sweden, which is more accentuated in women,[6] can both be attributable to medical prevention efforts and population changes in lifestyle.[7-9] Individual lifestyle factors are by growing age becoming of stronger importance for the development of IHD[1] and one such factor is our diet. Specifically, there is uncertainty about the relevance of intakes of dairy products since previous study results have been conflicting,[10-14] without consensus in recommendations has been reached.[14,15] These discrepancies in results may be a consequence of different types of dairy products investigated, fat content and exposure width of examined intake. Both non-fermented and fermented milk are widely consumed and they may have differential effect on cardiovascular health.[10-14,16,17] Consumption patterns of different dairy products have however considerably changed during the last half century, with on average continued lowered intake of non-fermented milk in many settings[18,19] and these changes need to be accurately captured by the study design and analysis.[20]

We therefore used data from two large Swedish longitudinal cohorts consisting of women and men to assess the risk of non-fatal and fatal myocardial infarction and ischemic heart disease with a wide range of consumption patterns of non-fermented and fermented milk consumption. The aim of the study is to investigate associations between time-updated information of non-fermented and fermented milk consumption and risk of IHD and acute myocardial infarction. With use of a replication and discovery design in subcohorts of the larger cohorts, a secondary aim is to examine non-fermented and fermented milk consumption in relation to patterns of cardiovascular plasma proteomic concentrations.

Study Population

Swedish Mammography Cohort (SMC) and the Cohort of Swedish Men (COSM).
Examinations; 1987-90 (women only), 1997, 2008/2009.

Exclusions:

- Exclude due to missing baseline questionnaire. When 2009 is set to baseline, we base this on the combination of the 2008 & 2009 questionnaires, requiring participation in both.
- Exclude participants with an incorrect or a missing personal identity numbers and those with a history of cancer or death before baseline.
- Excluded those with an implausible energy intake in the baseline questionnaire (defined as 3 standard deviations from the log-transformed mean energy intake in women and men separately). This is already done for SMC 1987.
- Exclude those with an IHD diagnosis (as recorded in main and all other diagnoses) before baseline. We will exclude: ICD-8: 410-413, 426, 429; ICD-9: 410-414; ICD-10: I20-I25.

For baseline other than SMC-1987 and COSM-1997, exclusion variables based on the first three of the above criteria is called `exclude_baseline_1997` & `exclude_baseline_2009`. The last criteria is denoted by `prev_ihd1987`, `prev_ihd1997` & `prev_ihd2009`. The special treatment of this last criteria is due to a sensitivity analysis in which we ignore this exclusion criteria.

Baseline & time updates:

1987-1989: Mamdate (SMC)
1998: January 1st 1998
2008/2009: April 14th 2009

Don't update information on exposure and covariates for those with:

- Missing questionnaire
- Implausible energy intake

This applies only to the questionnaire of interest. For instance, in SMC, if 1997 questionnaire has implausible energy intake, but not 2009, then update will be done in 2009 for the analyses with baseline 1987.
Two variables are created for this purpose: `update_1997` and `update_2009` that have the value 1 if an individual's information should be updated for that questionnaire. These variables are based on the two conditions on missing questionnaire and implausible energy intake.

Main exposure

- Milk intake, including total non-fermented milk intake and milk by three fat content categories
- Fermented milk intake including yogurt and sour milk

Milk and fermented milk intake will be presented in grams.

1987:

- Milk
  - `egen milk1987 = rowtotal(gram_milk05_87 gram_milk15_87 gram_milk30_87)`
    - For one sensitivity analysis, create an indicator variable `milk_miss1987` if all three variables above are missing
  - Fat content specific
    - `gen milk_low1987 = gram_milk05_87`
    - `gen milk_medium1987 = gram_milk15_87`
    - `gen milk_high1987 = gram_milk30_87`
- Fermented milk
  - `egen fil1987 = rowtotal(gram_sourmilk05_87 gram_sourmilk30_87)`
    - For one sensitivity analysis, create an indicator variable `fil_miss1987` if both variables above are missing

1997:

- Milk
  - `egen milk1997 = rowtotal(gramf52 gramf54 gramf56)`
    - For one sensitivity analysis, create an indicator variable `milk_miss1997` if all three variables above are missing
  - Fat content specific
    - `gen milk_low1997 = gramf52`
    - `gen milk_medium1997 = gramf54`
    - `gen milk_high1997 = gramf56`
- Fermented milk
  - `egen fil1997 = rowtotal(gramf58 gramf60)`
    - For one sensitivity analysis, create an indicator variable `fil_miss1997` if both variables above are missing

2008/2009:

- Milk
  - `egen milk2009 = rowtotal(f15gr_1 f15gr_3 f15gr_5)`
    - For one sensitivity analysis, create an indicator variable `milk_miss2009` if all three variables above are missing
  - Fat content specific
    - `gen milk_low2009 = f15gr_1`
    - `gen milk_medium2009 = f15gr_3`
    - `gen milk_high2009 = f15gr_5`
- Fermented milk

- `egen fil2009 = rowtotal(f15gr_7 f15gr_9 f15gr_11 f15gr_13)`
    - For one sensitivity analysis, create an indicator variable `fil_miss2009` if all four variables above are missing
    - For some sensitivity analyses, get information without fruit yogurt:
        - `egen fil_nofruit2009 = rowtotal(f15gr_9 f15gr_11 f15gr_13)`

Outcome

Incident and previous IHD cases will be identified by linkage with the Swedish National Patient and Cause of Death Registers. To classify IHD diagnoses we will use the ICD-9 codes (International Classification of Diseases 10th Revision) 410-414 and the ICD-10 codes I20-I25. We will use first-ever primary (main diagnosis) inpatient IHD diagnosis or IHD (main) diagnosis indicated as cause of death after baseline.

Main outcome

- IHD:
    - ICD-9: 410-414
    - ICD-10: I20-I25

Subtype

- Acute myocardial infarction (MI)
    - ICD-9: 410
    - ICD-10: I21

End of follow-up; 31 December 2019

Censoring: The first of:

- Date of death
- End of follow-up
- Other subtype of IHD (only applicable in subtype-analyses)

Description of possible confounders

Sex	Stratification variable
Age	
Educational level	≤9 years, 10–12 years, >12 years, or other
Living alone	Yes/No
Leisure time exercise	<1 h/w, 1 h/w, 2-3 h/w, 4-5 h/w, >5 h/w
Walking/cycling	Never/Seldom, <20 min/d, 20-40 min/d, 40-60 min/d, 1-1.5 h/d, >1.5 h/d
Body mass index (BMI)	Weight in kg divided by height in m^2; continuous
Height	Height in cm, continuous
Total energy intake	kcal/day; continuous
Fermented milk / Non-fermented milk	Depending on the main exposure. Continuous
Cheese	Gram/day. Continuous
Alcohol consumption	Continuous
Fruits and vegetables	Servings/day of: Continuous
Red meat	Gram/day. Continuous
Soft drinks and Juice	Servings/day. Continuous
Coffee	cups/day; continuous
Total fat intake	Energy-adjusted total fat intake. Continuous
Saturated fat intake	Energy-adjusted saturated fat intake. Continuous
Smoking status	Current, former, never
Vitamin- and mineral supplements	Yes/No
Weighted Charlson Comorbidity Index.	Charlson Comorbidity Index. From baseline, no time-update. Continuous
Major cardiovascular disease other than IHD (here called "CVD")	ICD8: 412, 420-429 ICD9: 415-438, ICD10: I26-I69. Yes/No. From baseline, no time-update
Diabetes mellitus	History of diabetes (yes/no), self-reported in combination with NPR. From baseline, no time-update

Models:

Model 1: adjust for sex through stratification + age.

Model 2: Sex, Age, Educational level, Living alone, Leisure time exercise, Walking/cycling, Body mass index (BMI), Height, Total energy intake, Fermented milk / Non-fermented milk, Cheese, Alcohol consumption, Fruits and vegetables, Red meat, Soft drinks and Juice, Coffee, Total fat intake, Saturated fat intake, Smoking status, Vitamin- and mineral supplements, baseline Weighted Charlson Comorbidity Index, baseline CVD, baseline Diabetes mellitus.

Statistical analyses

Time at risk of IHD for each participant will be calculated from baseline (according to the baseline dates described above) until the date of the first IHD diagnosis, date of death, end of follow-up or other subtype of IHD (only applicable in analyses concerning subtypes of IHD), whichever came first.

In the main analyses, if possible, we will use information of exposure and covariates from the following investigations; 1987-90 (women only), 1997, and 2008/2009. In the analysis we will time-update exposure and covariates. Missing information of covariates will be imputed, unless the complete questionnaire is missing or the questionnaire has implausible energy intake, in which case no update is made.

Cox proportional hazards regression models with time since baseline as time scale will be used to estimate hazard ratios with 95% confidence intervals (CI) for categories of milk and fermented milk intake. We will investigate total IHD and MI. When investigating MI, the other subtypes of IHD will be censored.

Intake will be divided into categories <200, 200-399, 400-599, and ≥600 grams per day for milk consumption and 0, 0-199, 200-399, ≥400 grams per day for fermented milk consumption. The category with the lowest intake will be used as reference (<200 grams for milk and 0 grams for fermented milk).

We will also look at the continuous version of Milk/fermented milk, in which we estimate as a linear variable, with HR presented per 200 gram. We will also treat the continuous variables as non-linear, using restricted cubic splines using three knots (percentiles 10, 50, 90). These results will be presented as graphs and tables containing HR for 200, 400, 600, 800 gram intake, compared to 100 gram intake as reference.

Analyses

All analyses will include categorized main exposures, linear continuous main exposure and non-linear continuous main exposure (restricted cubic spline). There are two sets of adjustment variables, according to the two models above.

1) Descriptive statistics

2) Start 1987-90 (mamdate) for SMC + time updated info from 1997 and 2008/2009.
Start 1998-01-01 for Pooled, SMC, and COSM + time updated info from 2008/2009.

3) Start 1987-90 (mamdate) for SMC but no time update.
Start 1998-01-01 for Pooled, SMC, and COSM but no time update.
Start 2009-04-14 for Pooled, SMC, and COSM.

4) As in 3, but we remove those with missing information at baseline on all types of milk and all types of yogurt/sour milk respectively. The variables that defines whether or not milk is missing is called `milk_miss1987`, `milk_miss1997`, `milk_miss2009`. In the analyses where milk is the main exposure, we remove those with missing information on milk and correspondingly for the analyses where yogurt/sour milk is the main exposure.

5) As in Analysis 2 but only for fermented milk intake and in these analyses fermented milk with fruit is excluded. That is, before splitting data, `replace fil2009 = fil_nofruit2009`. Don't forget to re-categorize the categorical version of the variable.

6) As in 3 (2009) but only for fermented milk intake and in these analyses fermented milk with fruit is excluded. That is, `replace fil2009 = fil_nofruit2009`. Don't forget to re-categorize the categorical version of the variable.

7) Test the proportional hazard assumptions graphically using Schoenfeld residuals. As in 2). Use the final adjustment model 2. This is done on the first imputed data set, as if it was observed.

8) As in 3) but Complete case analyses. For SMC 1987-1990, do not adjust for the variables that are not in 1987-questionnaire, since they will be missing for everyone that died before 1997 or that didn't answer the 1997-questionnaire.

9) Substitution analyses according to Figure 2 in *doi: 10.1136/bmj.l6204*. Total dairy is defined as
milk + fermented milk + cheese.
Substitution hazard ratios to be calculated are when substituting fermented milk intake for non-fermented milk intake (meaning that we increase fermented milk and decrease non-fermented milk)
Baselines are 1987-1990, 1997 & 2009, with time updates.
The dairy product variables are all modelled linearly.

10) As in 2), but analyse each fat content of milk separately, as in analysis A in https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7284719/. Adjust for other fat contents. Spline-curves for each fat-content separately. Due to the high number of zeros, the knots for the spline-variables needs to be decided manually, deviating from Harrell's recommendation.
For each fat content, delay entry or censor those with ≥200g milk with other fat content in the sense that no one is in the risk set when drinking ≥200g milk with other fat content. In Stata-terms it might look like this:

```
replace _st = 0 if (mellanmjölk + standardmjölk) >=200
```

11) As in 2), but stratified on the variables sex, year of birth, BMI, exercise, walking, marital status, education, Charlson comorbidity index, smoking status, and energy intake, using an interaction term. Wald test concerning potential effect modification is conducted. Milk/fermented milk treated as linear. Results presented as forest plot, with cut-offs as in Figure 2, https://doi.org/10.1371/journal.pmed.1003331

12) As in 2), but exclude individuals with pre-existing cardiovascular disease (`prev_cvd`), and diabetes (`prev_diabetes`).

13) As in 2), but include individuals with pre-existing IHD.

14) Austin, RR and RD. Högsta mot lägsta kategorin. To be continued.

15) SMC, start 1987, update 1997 (not 2009). Include individuals with pre-existing IHD. Outcome is time to general death. End of followup is 31 dec 2010, 30 sep 2015, & 31 dec 2019.
Three different adjustments:
Model 2: As above
Model 3: As model 2, but exclude coffee, soft drink, alcohol, fat intake, saturated fat intake, previous diabetes, previous CVD.
Model 4: As model 3, but use linear (instead of spline) adjustment of continuous covariates. Also, exclude walking/cycling.

16) As 02, IHD, model 2, for categorical and spline exposure. Calculate p-value for interaction terms with sex.

17) Splines (three knots) and Forest plots for the dose-response association between non-fermented and fermented milk intake with plasma protein biomarkers from Olink's panels CVD II, CVD III, and Metabolism. The subcohort SMCC in Uppsala is used for discovery (FDR cutoff 0.05), and the subcohort in Västerås (SMCC + COSMC) for replication.

Variables from questionnaires

Variabelnamn	Definition 1987	Definition 1997	Definition 2008/2009	Kommentar
male				
smoke	Hämta information från 1997	tobac 1: Current 2: Former 3: Never	f27_1 1: Never 2: Current 3: Former	
bmi	weight_87 & height_87	bmi	q47/((q46/100)^2	Kräver förmodligen lite städning av data för alla tre enkäterna
height	height_87	height	q46	Enhet i cm
educat	edu_87 1-3: <=9 years 4: 10-12 years 5: >12 years 6: Missing	educat 1: <=9 years 2: 10-12 years 3: >12 years 4: Missing		Högsta värdet av 1987 och 1997 (2009 har inte den informationen)
walk	Hämta information från 1997	f37 1: Never/seldom 2: < 20 min/day 3: 20-40 min 4: 40-60 min 5: 1-1.5 hours 6: >1.5 hours/day	f1_1 1: Never/seldom 2: < 20 min/day 3: 20-40 min 4: 40-60 min 5: 1-1.5 hours 6: >1.5 hours/day	
exercise	Hämta information från 1997	f49 1: < 1 hour/week, 2: 1 hour, 3: 2-3 hours, 4: 4-5 hours, 5: >5 hours/week	f1_5 1-2: < 1 hour/week, 3: 1 hour, 4: 2-3 hours, 5: 4-5 hours, 6: >5 hours/week	
alcohol	Alco	col1	nut8	
coffee	freq_coffe	g68	rowtotal(f15fr_31 f15fr_33), miss	
alone	mar_status_87 2, 5: No 1, 3, 4: Yes	SMC: f336 1: No 2: Yes	if inlist(q221, 2, 3, 4) | (q222>1 & !missing(q222)): No	

		COSM: f1327 2: No 1, 3, 4: Yes	if ((q221==1) & (inlist(q222, 0, 1)) | (missing(q222) & !missing(q223))): Yes	
milk	rowtotal(gram_milk05_87 gram _milk15_87 gram milk30_87)	rowtotal(gramf52 gramf54 gramf56)	rowtotal(f15gr_1 f15gr_3 f15gr_5)	
fil	rowtotal(gram_sourmilk30_87 gram_sourmilk05_87)	rowtotal(gramf58 gramf60)	rowtotal(f15gr_7 f15gr_9 f15gr_11 f15gr_13)	
fil_nofruit	Only applicable to 2009	Only applicable to 2009	rowtotal(f15gr_11 f15gr_9 f15gr_13)	
cheese	gram_cheese_87	rowtotal(gramf72 gramf74)	rowtotal(f15gr_45 f15gr_47)	
vegs	rowtotal(freq_rootvegetable_87 freq_whitecabbage_87 freq_tomato_87 freq_lettuce_87 freq_spinach_87 freq_brownbeans_87), miss	rowtotal(g133 g134 g135 g136 g137 g138 g139 g140 g141 g142 g143 g144 g145 g146), miss	rowtotal(f17fr_45 f17fr_46 f17fr_47 f17fr_48 f17fr_49 f17fr_50 f17fr_51 f17fr_52 f17fr_53 f17fr_54 f17fr_55 f17fr_56 f17fr_57 f17fr_58 f17fr_59 f17fr_60 f17fr_61 f17fr_62 f17fr_63 f17fr_64), miss	
fruits	rowtotal(freq_apple_87 freq_citrusfruit_87 freq_banana_87), miss	rowtotal(g148 g150 g151 g152 g153), miss	rowtotal(f17fr_65 f17fr_67 f17fr_68 f17fr_69 f17fr_70), miss	
fv	rowtotal(fruits vegs), miss	rowtotal(fruits vegs), miss	rowtotal(fruits vegs), miss	

red_meat	rowtotal(gram_meat_87 gram_meatstew_87 gram_bacon_87 gram_mincedmeat_87 gram_sausage_87 gram_sandwichfill_87 gram_pate_87 gram_blackpudding_87 gram_liver_87), miss	rowtotal(gramf115 gramf116 gramf117 gramf118 gramf119 gramf120 gramf121 gramf122), miss	rowtotal(f15gr_51 f15gr_53 f17gr_15 f17gr_16 f17gr_17 f17gr_18 f17gr_19 f17gr_20 f17gr_21 f17gr_22 f17gr_23 f17gr_24 f17gr_25 f17gr_26), miss	
softdrink	rowtotal(freq_juice_87 freq_softdrink_87 freq_fruitdrink_87), miss	rowtotal(g62 g149), miss	rowtotal(f15fr_17 f15fr_19 f15fr_21 f15fr_23 f17fr_66), miss	
supplements	Hämta information från 1997	inlist(f194, 1, 2) if inlist(f194, 1, 2, 3)	inlist(f23_1, 2, 3) if inlist(f23_1, 1, 2, 3)	
energy_total	ener	col4	nut1	
adj_fat	adjfat87	col_eadj6	enut3	
adj_sat_fat	adjsfat87	col_eadj14	enut13	
diabetes	Finns inte information i 1987. Här får patientregistret bestämma	f260==1 | f261!=.	q41==2 | q42!=.	Denna variabel kombineras senare med prev_diabetes från patientregistret, och kommer då gå under namnet prev_diabetes

Variables from patient register. For exclusion and adjustment

	ICD-8	ICD-9	ICD-10	Kommentar
prev_diabetes	250	250	E10-E14	Både huvuddiagnos och bidiagnos. Slutenvård
cci_w (weighted)	Se Stata-kod			Både huvuddiagnos och bidiagnos. Slutenvård
cci_unw (unweighted)				
prev_cvd	412, 420-429	415-438	I26-I69	Både huvuddiagnos och bidiagnos. Slutenvård
prev_ihd	410-413, 426, 429	410-414	I20-I25	Både huvuddiagnos och bidiagnos. Slutenvård
prev_cancer	14 15 16 17 18 190 192 193 194 195 196 197 198 199 200 201 202 203 204 205	14 15 16 170 171 172 174 175 176 177 178 179 18 19 20	C0 C1 C2 C3 C40 C41 C42 C43 C45 C46 C47 C48 C49 C5 C6 C7 C8 C9	Både huvuddiagnos och bidiagnos. Slutenvård

Outcome variables from patient register. For each outcome, we create one variable per baseline, giving the first date of the outcome since baseline.

	ICD-9	ICD-10	Kommentar
ihd_outcome	410-414	I20-I25	Enbart huvuddiagnos. Slutenvård.
mi_outcome	410	I21	Enbart huvuddiagnos. Slutenvård.