

Changes in intake of dairy product subgroups and risk of type 2 diabetes: modelling specified food substitutions in the Danish Diet, Cancer and Health cohort

Daniel B. Ibsen¹, Kim Overvad¹, Anne Sofie D. Laursen¹, Jytte Halkjær², Anne Tjønneland^{2,3}, Tuomas O. Kilpeläinen⁴, Erik T. Parner¹, Marianne U. Jakobsen⁵

¹Department of Public Health, Aarhus University, Aarhus, Denmark

²Danish Cancer Society Research Center, Copenhagen, Denmark

³Department of Public Health, University of Copenhagen, Copenhagen, Denmark

⁴Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

⁵National Food Institute, Division for Diet, Disease Prevention and Toxicology, Technical University of Denmark, Kgs. Lyngby, Denmark

Correspondence to:

Daniel Borch Ibsen

Department of Public Health, Aarhus University

Bartholins Allé 2, DK-8000 Aarhus C, Denmark

E-mail: dbi@ph.au.dk

ORCID: 0000-0002-7038-4770

1 Abstract

2 We investigated the association between an increased intake of one dairy product subgroup at the
3 expense of another within a 5-year period and the subsequent 10-year risk of developing type 2
4 diabetes. Effect modification by the initial level of the dairy product intake was also examined. The
5 cohort included 39,438 men and women with two measurements of diet assessed using food
6 frequency questionnaires (FFQ) administered in 1993-1997 and 1999-2003. Dairy products were
7 skimmed milk, semi-skimmed milk, whole-fat milk, buttermilk, low-fat yogurt, whole-fat yogurt,
8 cheese and butter. Type 2 diabetes cases were ascertained from the Danish National Diabetes
9 Register. The pseudo-observation method was used to calculate risk differences (RD) with 95%
10 confidence intervals (CI). Among participants aged 56-59 years at hand-in of the follow-up FFQ,
11 increased intake of whole-fat yogurt in place of skimmed, semi-skimmed or whole-fat milk was
12 associated with a reduced risk (RD% [95% CI]: -0.8% [-1.3, -0.2]; -0.6% [-1.1, -0.1]; -0.7 [-1.2, -
13 0.1]; per 50g/d, respectively). Among participants aged 60-64 and 65-72, substitution of skimmed
14 milk for semi-skimmed milk was associated with an increased risk of type 2 diabetes (0.5% [0.2,
15 0.7]; 0.4% [0.1, 0.7]; per 50g/d, respectively). No effect modification by the initial intake of milk
16 products was observed. Our results suggest that substitution of whole-fat yogurt for milk among
17 those aged 56-59 decreases the risk of type 2 diabetes and substitution of skimmed milk for semi-
18 skimmed milk was associated with an increased risk of type 2 diabetes among those aged 60-64 and
19 65-72, regardless of the initial intake level.

20 **Keywords:** diet change · substitution models · follow-up studies · dairy · diabetes

21

22 Abbreviations

23 BMI Body mass index
24 CI Confidence interval
25 FFQ Food frequency questionnaire
26 RD Risk difference

27

28 **Declarations**

29 **Funding** Daniel B. Ibsen was supported by a PhD fellowship by Aarhus University and Tuomas O. Kilpeläinen was
30 supported by the Novo Nordisk Foundation (NNF18CC0034900 and NFF17OC0026848). The Diet, Cancer and Health
31 cohort was funded by Danish Cancer Society. The funding agencies had no influence on the design, analysis or writing
32 of this paper.

33
34 **Conflict of interest** The authors declare that they have no conflict of interest.

35
36 **Ethics approval** The study was approved by Regional Ethical Committees on human studies and by the Danish Data
37 Protection Agency.

38
39 **Consent to participate** All participants gave written informed consent to participate in the study.

40
41 **Consent for publication** Not applicable.

42
43 **Availability of data and material** Data described in this article may be made available upon request pending on
44 application to and approval by the Danish Cancer Society (e-mail: dchdata@cancer.dk).

45
46 **Code availability** Code is available upon request to the corresponding author.

47
48 **Authors' contributions** The contributions were as follows—DBI, KO, ASDL, ETP and MUJ designed the study; DBI
49 analysed the data supervised by ETP; DBI, KO, ASDL, JH, AT, TOK, ETP and MUJ interpreted the data and critically
50 revised the manuscript; DBI wrote the paper; and DBI had final responsibility for the final content. All authors have
51 read and approved the final manuscript.

52

53

54 **Introduction**

55 The incidence of type 2 diabetes is increasing rapidly worldwide [1] and the prevention of type 2
56 diabetes is a top priority in public health [2]. Diet, together with other lifestyle factors, plays a key
57 role in prevention of type 2 diabetes. Indeed, risk of type 2 diabetes has been found to remain lower
58 at least 10 years after participation in lifestyle interventions that included dietary modifications
59 lasting two to six years [3].

60 Current dietary recommendations to the general population [4,5] and for prevention of type 2
61 diabetes [6,7] recommend low-fat dairy products instead of whole-fat dairy products [8]. However,
62 meta-analyses of randomized interventions have found no clear effect of increasing the total intake
63 of dairy products on glucose metabolism or insulin sensitivity and no effect of increasing the intake
64 of low-fat or high-fat dairy products on cardio-metabolic risk markers, except for a slightly
65 increased body weight [9–11]. This lack of an effect may be due to small sample sizes and short
66 durations of the included trials and to differences in types of dairy products provided and control
67 groups used, either consuming their habitual diet, a low-dairy or no-dairy diet, between the included
68 trials. Only few cohort studies have investigated the association between changes in intake of dairy
69 products on the risk of developing type 2 diabetes [12]. Of the few cohort studies that investigated
70 changes in intake of dairy products, higher consumption of yogurt, not milk products, was
71 associated with a lower risk of type 2 diabetes [13]. To be sure that assessed associations are
72 independent of changes in energy intake, it is important to adjust for energy intake in the statistical
73 model. Consequently, a food substitution is introduced where the model now reflects the influence
74 of increasing the intake of, for example, yogurt products, as in the example, at the expense of a mix
75 of other foods on the risk of type 2 diabetes; mimicking an isocaloric intervention study. As all
76 foods are either harmful, neutral or beneficial in the prevention of type 2 diabetes, the association
77 between intake of a specific dairy product subgroup and the risk of type 2 diabetes depend on the
78 substituted food. Unfortunately, this has rarely been specified, resulting in an unclear interpretation
79 of the results.

80 Another relevant question when investigating dietary changes is whether the initial intake
81 modifies the association between a dietary change and development of disease. In the context of
82 changes in dairy product subgroups, the association between substitution of, for instance, skimmed
83 milk for whole-fat milk and the risk of type 2 diabetes may be different, depending on whether the
84 initial intake of whole-fat milk was high or low.

85 Specifying food substitutions when studying associations between dietary changes and long-term
86 disease risk may add important new knowledge on the public health impact of changing diet. Thus,
87 we investigated the association between an increased intake of one dairy product subgroup at the
88 expense of another within a 5-year period, and the subsequent 10-year risk of developing type 2
89 diabetes in middle-aged men and women. In addition, we examined whether the association was
90 modified by the initial intake of these products.

91

92 **Materials and methods**

93 **Study population**

94 The Diet, Cancer and Health cohort is a Danish cohort that recruited participants from 1993-1997.
95 Middle-aged men and women from the Aarhus and Copenhagen areas were invited to participate in
96 the study. At inclusion in 1993-1997 participants answered a validated food frequency
97 questionnaire (FFQ) mailed to the participants [14,15]. Furthermore, the participants underwent a
98 physical examination and answered a lifestyle questionnaire at one of two study centres [16]. In
99 1999-2003, participants from the original cohort answered mailed follow-up questionnaires on diet
100 and lifestyle similar to the questionnaires provided in 1993-1997.

101 For the purposes of the present study, participants with a diagnosis of diabetes, myocardial
102 infarction, stroke or cancer before the time of the hand-in of questionnaires in 1999-2003 were
103 excluded. Furthermore, participants for whom we did not have information on exposures or
104 covariates were also excluded.

105

106 **Dairy product subgroups**

107 The initial intake of dairy product subgroups (skimmed milk 0.1% fat, semi-skimmed milk 1.5%,
108 whole-fat milk 3.5%, buttermilk 0.5%, low-fat yogurt <1.5%, whole-fat yogurt 3.5%, cheese and
109 butter) was assessed in 1993-1997 using a validated FFQ designed for the Diet, Cancer and Health
110 cohort [14,15]. The FFQ inquired participants about their average intake of the dairy product
111 subgroups during the previous 12 months according to 12 possible categories ranging from 'never'
112 to 'eight times or more per day'. Low-fat yogurt and whole-fat yogurt included plain yogurt types
113 and yogurt with fruit. Participants handed in the FFQ at one of the two study centres and it was
114 checked by trained personnel that immediately addressed inconsistencies of the questionnaires with
115 the participants, if necessary. In 1999-2003, a FFQ similar to the 1993-1997 FFQ was mailed to the
116 participants and the returned FFQs were checked by trained personnel. Missing values were scored

117 in terms of the type and relevance of the question (0, 0.5, 1 or 10 points). Participants with >9
118 points were contacted by phone. If participants had >55 points, the FFQ was mailed back to the
119 participants. The remaining missing values were coded as the lowest intake level “never/seldom”.
120 Harmonization of the dietary data collected in 1993-1997 and 1999-2003 is described in the
121 Supplemental text. Food and nutrient intakes in g/day were calculated using FoodCalc [17] and
122 Danish food composition tables [18] version 1996. Changes in intake of dairy product subgroups
123 were defined as the difference in reported intake in 1993-1997 (initial intake) and reported intake in
124 1999-2003, and expressed in units of one serving (50 g/day for skimmed milk, semi-skimmed milk,
125 whole-fat milk, buttermilk, low-fat and whole-fat yogurt, 20g/day for cheese and 6 g/day for
126 butter).

127

128 **Type 2 diabetes ascertainment**

129 Type 2 diabetes status of the participants was ascertained from the Danish National Diabetes
130 Register [19]. This register combines information from several Danish registers and covers the
131 entire Danish population. The date of diagnosis was the earliest date registered with (a) a diagnosis
132 of diabetes in the Danish National Patient Register; (b) chiropody as a diabetes patient in the Danish
133 National Health Service Register; (c) five blood glucose measurements in a 1-year period or two
134 blood glucose measurements per year in five consecutive years in the Danish National Health
135 Service Register; or (d) purchase of oral anti-diabetes drugs in the Danish National Prescription
136 Registry (excluding females prescribed metformin alone). The algorithm has been validated with a
137 positive predictive value of 89% as compared to medical records from the general practitioner [19].
138 The register does not distinguish between type 1 and type 2 diabetes. However, as all participants in
139 the Diet, Cancer and Health cohort were middle-aged, the vast majority of incident cases is
140 expected to be type 2 diabetic cases. The participants were followed up from the hand-in of
141 questionnaires in 1999-2003 until the date of the diagnosis of type 2 diabetes, death, emigration or
142 the date at which participants had been followed for 10 years after hand-in of questionnaires in
143 1999-2003, whichever came first. It is not known when the participants changed their dietary intake
144 during the 5 years between the two FFQs, but 10 years of subsequent follow-up was considered
145 reasonable based on evidence from clinical trials showing a lower risk of type 2 diabetes for up to
146 10 years after initiation of lifestyle interventions lasting two to six years [3].

147

148 **Assessment of covariates**

149 Information on covariates was obtained from the lifestyle questionnaires and FFQs administered in
150 1993-1997 and 1999-2003. The analyses were adjusted for known risk factors of type 2 diabetes
151 selected a priori based on a review of the literature. From the 1993-1997 questionnaires,
152 information on education, history of hypertension, history of hypercholesterolemia, total energy
153 intake and food groups associated with type 2 diabetes was obtained. From the questionnaires in
154 1999-2003, information on alcohol intake, smoking status, physical activity, body weight and waist
155 circumference and updated information on history of hypertension, history of hypercholesterolemia,
156 total energy intake and food groups was obtained. In addition, participants self-reported their family
157 history of diabetes. Height was measured at the physical examination in 1993-1997.

158

159 **Statistical analysis**

160 We calculated medians and 80% central ranges for continuous variables, and proportions for
161 categorical variables among the cohort and cases.

162 We used the pseudo-observation method [20,21] with age as the underlying time-scale to
163 calculate risk differences (RDs) and their corresponding 95% confidence intervals (CIs). A central
164 assumption of the pseudo-observation method is that there must be marginally independent entry
165 and censoring. We investigated the validity of this assumption using a previously described
166 methodology stratifying the cohort based on age and date of inclusion, which was the date at hand-
167 in of the questionnaires in 1999-2003 [22]. Because age was the underlying-time scale, a specific
168 age was set as the end of the follow-up period as required for the generation of the pseudo-values.
169 We used the median age at which participants were followed-up for 10 years after hand-in of
170 questionnaires in 1999-2003 as the end of the risk period. Risk periods were 56-67 years, 60-71
171 years and 65-76 years, in stratum 1, 2 and 3, respectively. Death during follow-up was handled as a
172 competing risk. The substitution models included all dairy product subgroups as change variables
173 and for all dairy product subgroups assessed in 1993-1997. The difference between two change-
174 variables' beta-coefficients (the beta-coefficient for the product increased minus the beta-coefficient
175 for the product decreased) and their variances and covariance were used to estimate the RDs and
176 corresponding 95% CIs for a given substitution. Hence, the association of an increased intake of a
177 specific dairy product subgroup and a concomitant decreased intake of another specified dairy
178 product subgroup (i.e. a substitution) in the 5-year period from 1993-1997 to 1999-2003, with the
179 subsequent 10-year risk of developing type 2 diabetes, was estimated. In multivariable analyses we
180 included total energy intake (1993-1997 and 1999-2003, kJ/day; continuous) and sex (man, woman)

181 in Model 1a. Total energy intake at both time-points was included to assess the association between
182 substitutions independent of total energy intake at both time-points and to reduce measurement
183 error of reported food intake [23]. Consequently, an unspecified residual substitution was
184 introduced, which is equal to the difference in energy provided by the substituted products. Model
185 1b was further adjusted for the duration of education (≤ 7 , 8-10, > 10 years), assessed in 1993-1997,
186 and alcohol intake (g/day; continuous, as a restricted cubic spline with three knots), smoking status
187 (never, former, current < 15 g tobacco/day, current 15-25 g tobacco/day, current > 25 g tobacco/day),
188 physical activity (participating in leisure-time sport activities or not), and family history of diabetes
189 (yes, no, do not know) assessed in 1999-2003. Model 2 was further adjusted for other foods
190 associated with type 2 diabetes: fruit, vegetables, red and processed meat, sugar-sweetened
191 beverages, whole grains, potato chips and coffee (all in g/day; continuous) assessed in 1999-2003.
192 Lastly, body mass index (BMI) (kg/m^2 ; continuous), waist circumference (cm; continuous), history
193 of hypertension (yes, no, do not know) and history of hypercholesterolemia (yes, no, do not know),
194 assessed in 1999-2003, were added in a separate model (Model 3), as these covariates may be
195 considered as mediators rather than confounders. Linearity for the dairy product variables was
196 investigated using restricted cubic spline of each dairy product subgroup. We constructed a
197 correlation matrix of dairy product intake, assessed in 1993-1997, and changes in dairy product
198 intake in order to assess potential multicollinearity (i.e. one variable almost completely explains
199 another variable) as changes in dairy product variables may be highly correlated with the initial
200 intake of dairy product subgroups.

201 We investigated potential effect modification by the initial intake level of the replaced dairy
202 product group assessed in 1993-1997 in subgroup analyses. For instance, the association between
203 substitution of skimmed milk for whole-fat milk and risk of type 2 diabetes in subgroup analyses by
204 initial low or high whole-fat milk intake assessed in 1993-1997 (cut at the 50th percentile). We also
205 conducted sex-stratified analyses.

206 As a sensitivity analysis, we excluded participants who developed hypertension and/or
207 hypercholesterolemia between 1993-1997 and 1999-2003. This was done because these participants
208 are likely to have changed their dietary intake as a result of diagnosis and treatment. Furthermore,
209 non-participation at follow-up may result in selection bias. Thus, in another sensitivity analysis, we
210 adjusted for potential selection bias by using inverse probability weighting [24]. The probability
211 weights were based on sex, age, education, alcohol, smoking, physical activity, BMI, waist
212 circumference, history of hypertension, history of hypercholesterolemia, history of diabetes and

213 alternate healthy eating index-2010 score [25] assessed in 1993-1997. Also, we compared
214 characteristics assessed with the 1993-1997 questionnaires between those who participated at both
215 time-points and those who only participated in 1993-1997. Changing two dietary components may
216 be followed by other dietary changes as well. We constructed spider plots to visually assess
217 differences in changes in intake of different food groups between participants that substituted one
218 for another dairy product subgroup, compared with participants who maintained a stable intake of
219 both dairy product subgroups. Substitutions were defined as increased intake of one dairy product
220 subgroup by ≥ 1 serving/day and decreased intake of another dairy product subgroup by ≥ 1
221 serving/day. To investigate the time-scale sensitivity of our results, we compared risk ratios from
222 the analyses with age as the underlying time-scale with risk ratios from analyses using time in the
223 study as the underlying time-scale. We compared risk ratios and not RDs as risk ratios are generally
224 less sensitive across studies than risk differences. Hence, differences between time-scales using risk
225 ratios would be of a greater concern than dissimilarities using risk differences. Lastly, we excluded
226 7179 participants with >9 missing values in their 1999-2003 FFQ coded as “no intake or less than
227 once per month” in the main analysis.

228 All analyses were performed in Stata version 15.1 (StataCorp, College Station, Texas, USA).
229 Forest plots were, however, coded in R version 3.5.0, R Core Team, Vienna, Austria, using the
230 ggepi package [26].

231

232 **Results**

233 In total, two measures of dietary intake were available for 45,247 participants (79% of the initial
234 cohort) (Figure 1). We excluded individuals who had a diagnosis of diabetes ($n=1971$), myocardial
235 infarction or stroke ($n=1518$), or cancer ($n=1621$) before hand-in of the questionnaires in 1999-
236 2003, and participants for whom there were missing values in exposures or covariates ($n=699$),
237 leaving 39,438 eligible participants.

238 The assumption of independent entry within the entire cohort did not hold (hazard ratio [95% CI]
239 for type 2 diabetes comparing first and third tertile of age at inclusion: 1.26 [1.11, 1.42]). Based on
240 this result and a previous study [27], we analysed substitutions within three strata based on age at
241 hand-in of the questionnaires in 1999-2003 and three strata based on calendar date at hand-in of the
242 questionnaires in 1999-2003 (in total 9 strata). Within each stratum, no violation of the assumptions
243 was observed. Thus, the pseudo-values were generated in each of these strata. Because the pseudo-
244 values can only be combined if they span over the same time-period within the used time-scale, the

245 data were analysed in three separate age-strata. Within each age-stratum, calendar date-strata were
246 combined after generation of pseudo-values with similar start and end ages for cumulative risks
247 within each age-stratum. However, as a consequence, we excluded 45 individuals with risk-time
248 outside these specified intervals. Stratum 1, aged 56-59 years, included 13,125 participants of which
249 1043 developed diabetes, stratum 2, 60-64 years, included 13,140 participants and 1188 cases and
250 stratum 3, 65-72 years, included 13,128 participants and 1281 cases. The end of each follow-up
251 period was set at the ages 67, 71 and 76 years with a median follow-up time of 9.5, 9.1 and 8.3
252 years for strata 1, 2 and 3, respectively.

253 Those who developed type 2 diabetes were more likely to be less educated, be smokers, not
254 participate in leisure-time sports activities, have a higher BMI and waist circumference and have a
255 history of hypertension or hypercholesterolemia and a family history of diabetes at hand-in of the
256 follow-up questionnaires (Table 1). There were no substantial differences in the median changes in
257 intake of dairy product subgroups between the cohort and those who developed type 2 diabetes.
258 However, the ranges of changes in intake (p10-p90) were different between the two (Table 1).

259 For the main substitution analyses we found different patterns across the age-strata (Figure 2 and
260 Supplemental Figure 1). After multivariable adjustment (Figure 2), an increased risk of type 2
261 diabetes was observed when skimmed milk replaced semi-skimmed milk among those aged 60-64
262 (RD [95% CI]: 0.5% [0.2, 0.7] per 50 g/d) and 65-72 (RD [95% CI]: 0.4% [0.1, 0.7] per 50 g/d) and
263 when skimmed milk replaced whole-fat milk (RD [95% CI]: 0.4% [0.0, 0.8] in those aged 60-64
264 and 0.4 [0.0, 0.7] in those aged 65-72). When whole-fat yogurt replaced milk products, a reduced
265 risk of type 2 diabetes was observed among those aged 56-59 (RD [95% CI]: whole-fat yogurt for
266 skimmed milk: -0.8% [-1.3, -0.2]; semi-skimmed milk: -0.6% [-1.1, -0.1]; whole-fat milk: -0.7% [-
267 1.2, -0.1]; per 50 g/d). Replacing skimmed milk, semi-skimmed milk or whole-fat milk with low-fat
268 yogurt was associated with a reduced risk of type 2 diabetes (RD [95% CI]: -0.4% [-0.9, 0.0]; -0.3%
269 [-0.7, 0.2]; -0.3% [-0.8, 0.1]; per 50 g/day, respectively), although not statistically significantly so.
270 In contrast, replacing low-fat yogurt with semi-skimmed or whole-fat milk was associated with an
271 increased risk in those aged 65-72 years (RD [95% CI]: 0.6% [0.1, 1.1]; 0.6% [0.0, 1.1]; per 50
272 g/day, respectively). The same pattern of associations was found after adjustment for BMI, waist
273 circumference, history of hypertension and hypercholesterolemia (Supplemental Figure 1, Model
274 3). No deviation from linearity of the dairy product variables and no multicollinearity was detected
275 during the model validation (data not shown).

276 Figure 3 shows the associations for substitutions between dairy product subgroups and risk of
277 type 2 diabetes stratified by the initial intake level of the replaced dairy product group. The patterns
278 of associations were largely similar between those with an initial low and high intake of the
279 substituted dairy product subgroup. When the analyses were stratified by sex, no associations
280 between substitutions of whole-fat yogurt for milk and risk of type 2 diabetes were observed among
281 women aged 56-59 whereas substitution of whole-fat yogurt for milk was associated with a reduced
282 risk in men aged 56-59. Substitution of skimmed milk for semi-skimmed milk was associated with
283 an increased risk among women aged 60-64 and 65-72 and only among men aged 60-64
284 (Supplemental Figure 2).

285 After excluding participants who developed hypertension or hypercholesterolemia between the
286 two dietary assessments, replacement of milk products with yogurt became more strongly
287 associated with a reduced risk of type 2 diabetes (Supplemental Figure 3). Otherwise, similar results
288 as in the main analyses were observed. Supplemental Figure 4 shows the flow of participants that
289 participated both in 1993-1997 and in 1999-2003 and those who only participated in 1993-1997.
290 Non-participants (i.e. those who only participated in 1993-1997) were more likely to be younger,
291 have a lower education, consume less alcohol, be smokers, be less physically active, have a higher
292 BMI and waist circumference and have a lower healthy eating index score in 1993-1997 compared
293 to those who answered both 1993-1997 and 1999-2003 questionnaires (Supplemental Table 1).
294 Adjusting for non-participation using inverse probability weighting provided similar results as the
295 main analyses (Supplemental Figure 5). Supplemental Figure 6 illustrates differences in intake of
296 other foods related to type 2 diabetes comparing participants that have increased their intake of ≥ 1
297 serving/day of, for example skimmed milk, and decreased their intake of ≥ 1 serving/day of, for
298 example semi-skimmed milk (i.e. substituted skimmed milk for semi-skimmed milk), with
299 individuals that did not change their intake of the dairy product subgroups. Subjects that substituted
300 low-fat dairy products, particularly skimmed milk or low-fat yogurt for other dairy products,
301 increased their intake of fruit and vegetables to a greater extent than those with a stable intake.
302 Comparison of different underlying time-scales and exclusion of participants with >9 missing
303 values in the 1999-2003 FFQ revealed similar patterns of associations compared with the main
304 analyses (Supplemental Figure 7 and 8, respectively).

305

306 **Discussion**

307 In this large cohort of middle-aged Danish men and women with repeated measures of diet, we
308 found different patterns of associations across age-strata for specified substitutions of dairy product
309 subgroups. We found that replacing milk products with whole-fat yogurt reduced the risk of type 2
310 diabetes among those aged 56-59 years. Replacing semi-skimmed milk with skimmed milk was
311 associated with an increased risk of type 2 diabetes among those aged 60-64 and 65-72 years. No
312 clear associations were observed for replacements with buttermilk, cheese or butter.

313 The primary exposures of this study were changes in intake of dairy product subgroups. To
314 investigate changes we used two measures of diet. The dietary intakes were self-reported and thus
315 measurement error is present. However, the FFQ was validated for nutrient intake and similar
316 FFQs, validated on a food level, tend to show a high correlation with food records for dairy
317 products ($r = 0.53-0.90$) [28–30]. We thoroughly adjusted for established risk factors of type 2
318 diabetes, but residual confounding cannot be excluded. Furthermore, after excluding those who
319 developed hypertension or hypercholesterolemia between 1993-1997 and 1999-2003, replacement
320 of milk with yogurt became more strongly associated with a reduced risk of type 2 diabetes. This
321 raise concerns about potential confounding by indication. However, the point in time when the
322 change in diet and development of hypertension or hypercholesterolemia was unknown. To our
323 surprise, we observed an increased risk of type 2 diabetes when replacing semi-skimmed milk with
324 skimmed milk among those aged 60-64 and 65-72 years. This is not consistent with the current
325 evidence, which mostly suggests no association between intake of milk, regardless of fat content,
326 and type 2 diabetes risk [12,31]. Other indicators of deteriorating health status (such as physical
327 disability or chronic inflammation) were not taken into account and thus, we cannot exclude the
328 possibility of confounding by indication here as well. All Danish men and women between ages 50-
329 64 years, without any registered cancer diagnosis, living in the two largest cities in Denmark were
330 invited to participate in the Diet, Cancer and Health cohort. It has previously been shown that those
331 who chose to participate generally had a higher socioeconomic position [16], which limits the
332 generalizability of our results. We observed similar tendencies among participants eligible for this
333 study. Participants had longer duration of education than non-participants. Hence, non-participation
334 at follow-up of about 17% may have introduced selection bias. However, after using inverse
335 probability weights of participation in our analyses, we found similar results compared to the main
336 analyses, suggesting that selection bias is not a major concern.

337 Only few previous studies have investigated changes in intake of dairy product subgroups on risk
338 of developing type 2 diabetes. Interestingly, we found that replacing milk with yogurt products was

339 associated with a reduced risk of type 2 diabetes in those aged 56-59, but not those aged 60-72
340 years. As the effect of dietary intake on future risk of developing type 2 diabetes is likely
341 cumulative, changes among older populations may not have the same impact as dietary changes
342 among younger populations. For instance, although intake of fruit and vegetables have been
343 consistently associated with a lower risk of type 2 diabetes [32], these associations were not found
344 in a meta-analysis of cohorts of elderly individuals (>60 years) [33]. Alternatively, higher baseline
345 hazards among the elderly may also explain the findings from this meta-analysis and none of the
346 included studies investigated changes in intake of fruit and vegetables. Current dietary
347 recommendations emphasize substitution of high-fat dairy product with low-fat dairy products,
348 although few cohort studies have investigated the health impact of changes in intake of dairy
349 products. In a cohort of Spanish men and women aged 55-80 years, Diaz-Lopez *et al.* [13], reported
350 that an increased intake of low-fat and whole-fat yogurt from below the median intake to more than
351 or equal to the median intake was associated with a lower risk of type 2 diabetes compared with a
352 stable intake below the median, whereas no similar association was observed for low-fat or whole-
353 fat milk. These findings support our findings among those aged 56-59. However, the authors did not
354 stratify according to different age groups and they did not specify relevant substitutions for the
355 changes in dairy intake. In a 9-week randomized intervention comparing commercially available
356 low-fat yogurt with a soy pudding, reduced levels of some, not all, measured biomarkers of chronic
357 inflammation and endotoxin exposure were observed in the low-fat yogurt group [34]. Overall,
358 these findings support a benefit of yogurt products on risk of development of type 2 diabetes
359 through mechanisms that may be mediated by changes in the gut microbiota, as suggested by others
360 [35].

361 An advantage of specifying the food substitution is that it provides a clear interpretation of the
362 results. We specified the association for increasing the intake of one dairy product subgroup at the
363 expense of another. Our main model was Model 2, which also adjusted for other foods. However, it
364 could be argued, for public health recommendations, that Model 1b, not adjusted for other foods,
365 may be more relevant as this model did not constrain the underlying dietary pattern. For example,
366 substitution of low-fat yogurt for skimmed milk may be associated with changes in the intake of
367 other foods. We illustrated this using spider plots and generally found that those who increased their
368 intake of low-fat yogurt also increased their intake of fruit, vegetables and coffee compared to those
369 who did not change their intake of these dairy products. Indeed, after further adjustment for other
370 foods (Model 2) some of the observed associations (Model 1b) disappeared. Although the

371 magnitude of the RDs observed was rather small, around -0.9% to -0.5% for a substitution of 50 g/d
372 for whole-fat yogurt with 50 g/d for milk products, this is not surprising, as we only modelled a
373 single dietary change consisting of two different products.

374 We also investigated the modifying effect of the initial intake of the dairy product subgroup
375 being replaced. Overall, the pattern of associations was similar among those with an initial high or
376 low intake of the substituted dairy product. The low average initial intake of whole-fat dairy
377 products in this cohort may partly explain the lack of a modifying effect. In our sex-stratified
378 analyses we observed different patterns of associations between men and women. However, as there
379 is little evidence of biological reasons for such differences, we interpret this with caution and
380 encourage replication of this finding in other cohorts.

381 Future research should focus on randomized interventions and cohort studies with repeated
382 measures of diet to evaluate these suggested differences between yogurt and milk products on risk
383 of developing type 2 diabetes or risk markers hereof.

384 In conclusion, this study suggests a reduced risk of type 2 diabetes when replacing milk with
385 whole-fat yogurt among men and women aged 56-59 years and an increased risk of type 2 diabetes
386 when replacing skimmed milk for semi-skimmed milk among those aged 60-64 and 65-72,
387 regardless of initial intake level.

Acknowledgements We thank the Danish Cancer Society, the staff at the Danish Diet, Cancer and Health study for collection of the data and the study participants for their contribution to the study. We thank data manager Nick Martinussen from the Danish Cancer Society for his large contribution with the harmonisation of the food frequency questionnaire data. We also thank data manager Katja Boll from the Danish Cancer Society for her work with the database and data administration.

References

1. Danaei G, Finucane MM, Lu Y, Singh GM, Cowan MJ, Paciorek CJ, et al. National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants. *Lancet*. 2011;378:31–40.
2. Zimmet PZ, Magliano DJ, Herman WH, Shaw JE. Diabetes: A 21st century challenge. *Lancet Diabetes Endocrinol*. 2014;2:56–64.
3. Hemmingsen B, Gimenez-Perez G, Mauricio D, Figuls M, Metzendorf M-I, Richter B. Diet, physical activity or both for prevention or delay of type 2 diabetes mellitus and its associated complications in people at increased risk of developing type 2 diabetes mellitus. *Cochrane Database Syst Rev*. 2017;12:CD003054.
4. U.S. Department of Health and Human Services and U.S. Department of Agriculture. Dietary Guidelines for Americans 2015-2020 [Internet]. [cited 2018 Apr 18]. Available from: <https://health.gov/dietaryguidelines/2015/guidelines/chapter-2/a-closer-look-at-current-intakes-and-recommended-shifts/#food-groups>
5. Nordic Council of Ministers. Nordic Nutrition Recommendations 2012. Integrating nutrition and physical activity. Copenhagen: Nord 2014; 2014.
6. Dyson PA, Twenefour D, Breen C, Duncan A, Elvin E, Goff L, et al. Diabetes UK evidence-based nutrition guidelines for the prevention and management of diabetes. *Diabet Med*. 2018;35:541–7.
7. Salas-Salvado J, Martinez-Gonzalez M, Bullo M, Ros E, Salas-Salvado J, Martinez-González MÁ, et al. The role of diet in the prevention of type 2 diabetes. *Nutr Metab Cardiovasc Dis*. 2011;21 Suppl 2:B32-48.
8. Weaver CM. How sound is the science behind the dietary recommendations for dairy? *Am J Clin Nutr*. 2014;99 (5 Supp):1217S-22S.
9. Connor SO, Turcotte A, Gagnon C, Rudkowska I. Glucose Concentrations and Glycated Hemoglobin: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Adv Nutr*. 2019;10:262–79.
10. Turner KM, Keogh JB, Clifton PM. Dairy consumption and insulin sensitivity: A systematic review of short- and long-term intervention studies. *Nutr Metab Cardiovasc Dis*. Elsevier B.V; 2015;25:3–8.
11. Benatar JR, Sidhu K, Stewart RAH. Effects of High and Low Fat Dairy Food on Cardio-Metabolic Risk Factors: A Meta-Analysis of Randomized Studies. *PLoS One*. 2013;8:e76480.
12. Soedamah-Muthu SS, de Goede J. Dairy Consumption and Cardiometabolic Diseases: Systematic Review and Updated Meta-Analyses of Prospective Cohort Studies. *Curr Nutr Rep*. Current Nutrition Reports; 2018;7:171–82.
13. Díaz-López A, Bulló M, Martínez-González MA, Corella D, Estruch R, Fitó M, et al. Dairy product consumption and risk of type 2 diabetes in an elderly Spanish Mediterranean population at high cardiovascular risk. *Eur J Nutr*. 2016;55:349–60.
14. Overvad K, Tjønneland A, Haraldsdóttir J, Ewertz M, Jensen O. Development of a semiquantitative food frequency questionnaire to assess food, energy and nutrient intake in Denmark. *Int J Epidemiol*. 1991;20:900–5.
15. Tjønneland A, Overvad K, Haraldsdóttir J, Bang S, Ewertz M, Jensen O. Validations of a semiquantitative food frequency questionnaire developed in Denmark. *Int J Epidemiol*. 1991;20:906–12.
16. Tjønneland A, Olsen A, Boll K, Stripp C, Christensen J, Engholm G, et al. Study design, exposure variables, and socioeconomic determinants of participation in Diet, Cancer and Health: a population-based prospective cohort study of

- 57,053 men and women in Denmark. *Scand J Public Health*. 2007;35:432–41.
17. Lauritsen A. FoodCalc. <https://www.cancer.dk/dchdata/access-to-data-and-biobank/foodcalc/> (2019). Accessed 18 May 2020.
18. DTU Food. Danish National Food Tables. <https://frida.fooddata.dk/?lang=en> (2015). Accessed 29 May 2020.
19. Carstensen B, Kristensen JK, Marcussen MM, Borch-Johnsen K. The National Diabetes Register. *Scand J Public Health*. 2011;39:58–61.
20. Parner ET, Andersen PK. Regression analysis of censored data using pseudo-observations. *Stata J*. 2010;10:408–22.
21. Andersen PK, Pohar Perme M. Pseudo-observations in survival analysis. *Stat Methods Med Res*. 2010;19:71–99.
22. Mortensen L, Hansen C, Overvad K, Lundbye-Christensen S, Parner E. The pseudo-observation analysis of time-to-event data. Example from the Danish Diet, Cancer and Health Cohort illustrating assumptions, model validation and interpretation of results. *Epidemiol Method*. 2018; <https://doi.org/10.1515/em-2017-0015>.
23. Subar AF, Freedman LS, Tooze JA, Kirkpatrick SI, Boushey C, Neuhauser ML, et al. Addressing Current Criticism Regarding the Value of Self-Report Dietary Data. *J Nutr*. 2015;145:2639–45.
24. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615–25.
25. Guenther PM, Casavale KO, Kirkpatrick SL. Update of Healthy Eating Index: HEI-2010. *J Acad Nur Diet*. 2014;113:1–20.
26. Johnston L. ggepi: ggplot2 Extensions for Visualizing Epidemiological Analyses (Chronic Diseases). R package version 0.01.9000; 2019. <https://github.com/lwjohnst86/ggepi>. Accessed 29 May 2020.
27. Mortensen LM, Lundbye-Christensen S, Schmidt EB, Calder PC, Schierup MH, Tjønneland A, et al. Long-chain n-3 and n-6 polyunsaturated fatty acids and risk of atrial fibrillation: Results from a Danish cohort study. *PLoS One*. 2017;12:e0190262.
28. Byers T, Marshall J, Anthony E, Fiedler R, Zielezny M. The reliability of dietary history from the distant past. *Am J Epidemiol*. 1987;125:999–1011.
29. van Liere M, Lucas F, Clavel F, Slimani N, Villemainot S. Relative validity and reproducibility of a French dietary history questionnaire. *Int J Epidemiol*. 1997;26:S128-36.
30. EPIC group of Spain. Relative validity and reproducibility of a diet history questionnaire in Spain. *Int J Epidemiol*. 1997;26:S91-9.
31. Vissers LE, Sluijs I, Schouw YT van der, Forouhi NG, Imamura F, Burgess S, et al. Dairy product intake and risk of type 2 diabetes in EPIC-InterAct: A Mendelian randomization study. *Diabetes Care*. 2019;42:568–75.
32. Schwingshackl L, Hoffmann G, Lampousi A-M, Knuppel S, Iqbal K, Schwedhelm C, et al. Food groups and risk of type 2 diabetes mellitus: a systematic review and meta-analysis of prospective studies. *Eur J Epidemiol*. 2017;32:363–75.
33. Mamluk L, O’Doherty M, Orfanos P, Saitakis G, Woodside J, Liao L, et al. Fruit and vegetable intake and risk of incident of type 2 diabetes: results from the consortium on health and ageing network of cohorts in Europe and the United States (CHANCES). *Eur J Clin Nutr*. 2017;71:83–91.
34. Pei R, DiMarco DM, Putt KK, Martin DA, Gu Q, Chitchumroonchokchai C, et al. Low-fat yogurt consumption reduces biomarkers of chronic inflammation and inhibits markers of endotoxin exposure in healthy premenopausal women: a randomized controlled trial. *Br J Nutr*. 2017;118:1043–51.

35. Fernandez MA, Marette A. Novel perspectives on fermented milks and cardiometabolic health with a focus on type 2 diabetes. *Nutr Rev.* 2018;76:16–28.

Tables

Table 1 Characteristics at hand-in of follow-up questionnaires (1999-2003) and changes in intake of dairy product subgroups of participants in the Diet, Cancer and Health cohort in three strata divided into tertiles of age¹.

Characteristics	Stratum 1 (56-59 years)				Stratum 2 (60-64 years)				Stratum 3 (65-72 years)			
	Cohort (n=13,125)		Cases (n=1043)		Cohort (n=13,140)		Cases (n=1188)		Cohort (n=13,128)		Cases (n=1281)	
	Median	p10-p90	Median	p10-p90	Median	p10-p90	Median	p10-p90	Median	p10-p90	Median	p10-p90
Women (%)	53		45		54		46		56		48	
Age (years)	57	56-59	57	56-59	61	59-63	61	59-63	67	64-70	67	64-70
Low education level (%)	22		30		32		38		37		44	
Alcohol intake	14	2-50	13	2-59	14	2-50	14	2-54	14	2-46	12	2-48
Current smoker (%)	29		34		28		33		26		27	
No leisure-time sports activity (%)	55		65		57		65		56		60	
Body mass index (kg/m ²)	25	21-31	28	23-36	25	21-31	28	23-34	25	21-31	27	23-34
Waist circumference (cm)	92	77-108	100	85-119	92	78-108	100	85-117	94	78-108	100	84-116
History of hypertension (%)	20		37		23		37		27		40	
History of hypercholesterolemia (%)	14		22		18		24		21		27	
Family history of diabetes (%)	14		24		13		20		13		21	
<i>Change in intake</i>												
Skimmed milk (g/d)	0	-5-198	0	-12-198	0	0-198	0	0-300	0	0-198	0	0-198
Semi-skimmed milk (g/d)	2	-192-116	2	-199-122	3	-186-123	3	-194-117	3	-184-158	3	-200-173
Whole-fat milk (g/d)	1	-24-42	1	-30-46	1	-34-47	1	-58-45	2	-44-58	2	-29-63
Buttermilk (g/d)	0	-15-19	0	-19-19	0	-22-23	0	-23-31	0	-30-33	0	-32-39
Low-fat yogurt (g/d)	-3	-45-113	-3	-71-97	-3	-71-139	-3	-71-109	-3	-74-155	-3	-71-181
Whole-fat yogurt (g/d)	-4	-106-14	-3	-88-14	-4	-108-15	-3	-88-18	-3	-129-28	-3	-156-25
Cheese (g/d)	-7	-41-17	-6	-43-19	-6	-40-17	-7	-39-18	-6	-38-17	-4	-37-20
Butter (g/d)	-1	-19-10	-1	-20-10	-1	-18-10	-1	-19-10	-1	-18-10	-1	-18-12

¹All medians and percentiles are reported as pseudo-medians and pseudo-percentiles (i.e. an average of the five surrounding values of each percentiles) in order to comply with European data protection regulations.

Figures

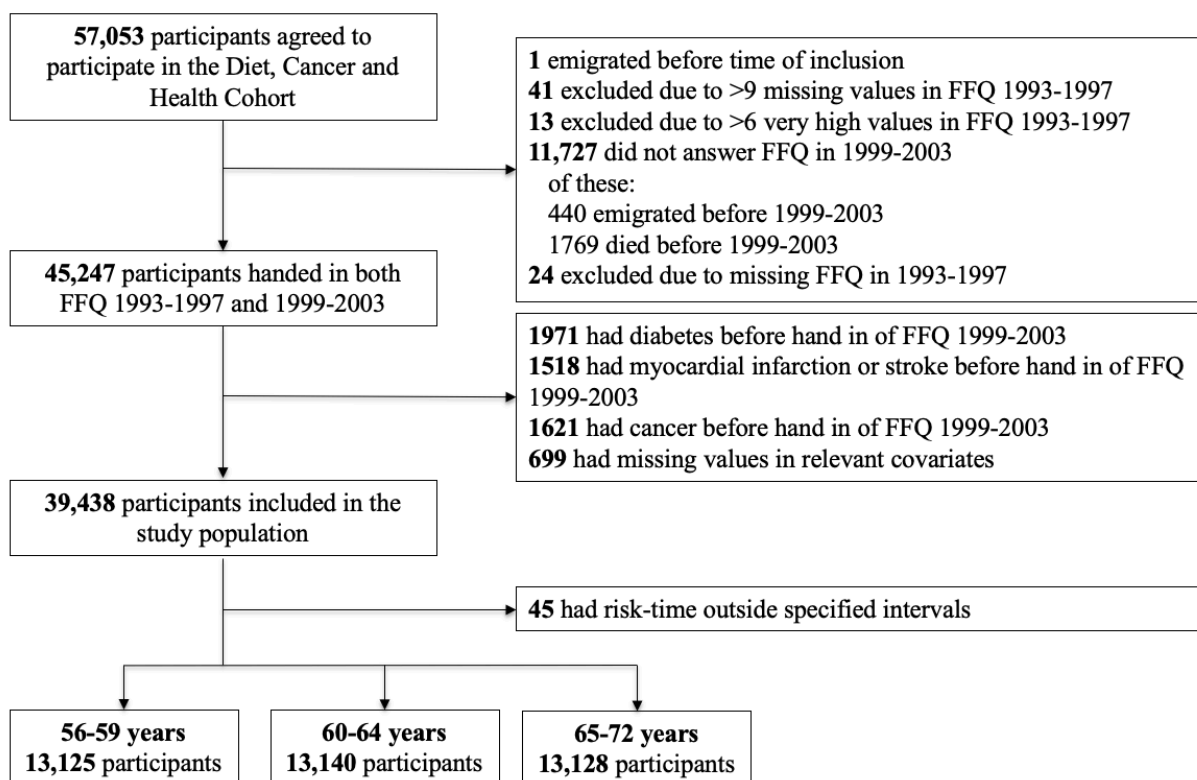


Figure 1 Flowchart of the study sample in the Diet, Cancer and Health cohort. FFQ: food frequency questionnaire.

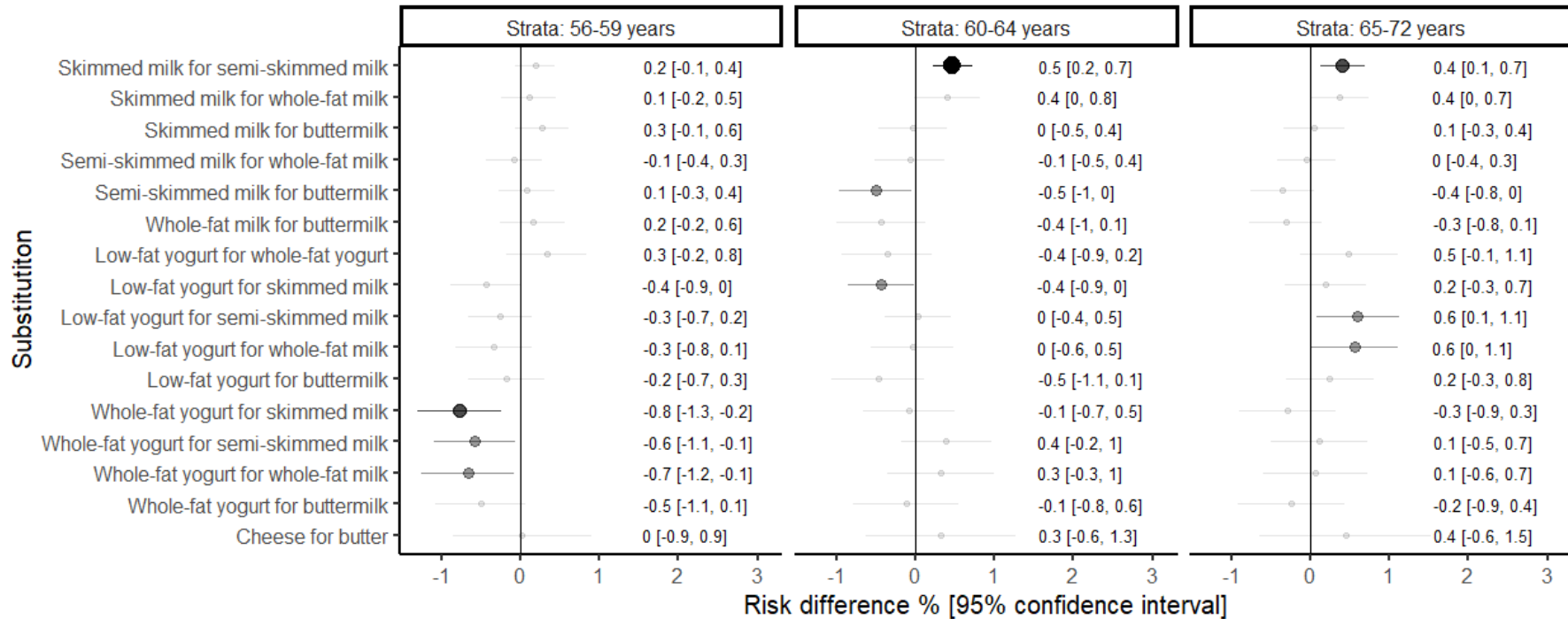


Figure 2 Forest plot of substitutions between dairy product subgroups on 10 years risk of type 2 diabetes in the Diet, Cancer and Health cohort across three strata divided into tertiles of age at hand-in of questionnaires in 1999-2003. 10-year risk periods were 56-67 years, 60-71 years and 65-76 years, in strata 56-59, 60-64 and 65-72 years, respectively. Intakes were modelled per 50 g/day for all dairy products except for cheese (20 g/day) or butter (6 g/day). Adjusted for Model 2 covariates: total energy intake (assessed in 1993-1997 and 1999-2003), sex, duration of education (assessed in 1993-1997) and alcohol intake, smoking status, leisure-time sports activity and family history of diabetes, fruit, vegetables, red and processed meat, sugar-sweetened beverages, whole grains, potato chips and coffee (assessed in 1999-2003). The larger and darker the dots, the lower the p-value for a given association.

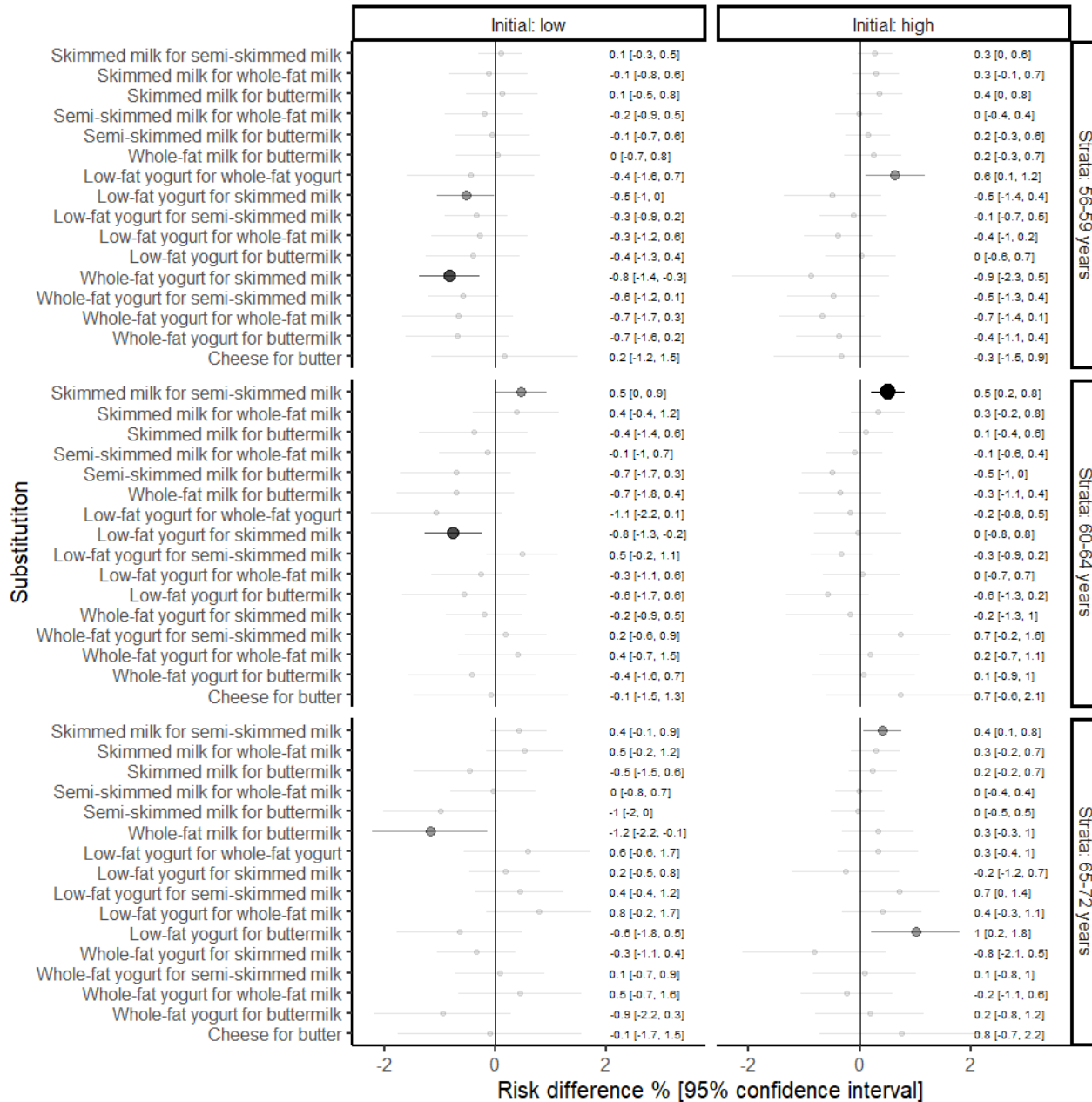


Figure 3 Forest plot of substitutions between dairy product subgroups on risk of type 2 diabetes, followed-up for subsequent 10 years, in the Diet, Cancer and Health cohort across three strata divided into tertiles of age at hand-in of questionnaires in 1999-2003 and low or high initial intake of the substituted dairy product, cut at the 50th percentile. 10-year risk periods were 56-67 years, 60-71 years and 65-76 years, in strata 56-59, 60-64 and 65-72 years, respectively. Intakes were modelled per 50 g/day for all dairy products except for cheese (20 g/day) or butter (6 g/day). Adjusted for Model 2 covariates: total energy intake (assessed in 1993-1997 and 1999-2003), sex, duration of education (assessed in 1993-1997) and alcohol intake, smoking status, leisure-time sports activity and family history of diabetes, fruit, vegetables, red and processed meat, sugar-sweetened beverages, whole grains, potato chips and coffee (assessed in 1999-2003). The larger and darker the dots, the lower the p-value for a given association.