Conflicting associations between dietary patterns and changes of anthropometric traits across subgroups of middle-aged women and men

Jiang, Li; Audouze, Karine; Romero Herrera, Jose Alejandro; Ångquist, Lars Henrik; Kjærulff, Sonny Kim; Izarzugaza, Jose M.G.; Tjønneland, Anne; Halkjær, Jytte; Overvad, Kim; Sørensen, Thorkild I.A.

Total number of authors: 11

Published in: Clinical Nutrition

Link to article, DOI: 10.1016/j.clnu.2019.02.003

Publication date: 2020

Document Version: Peer reviewed version


General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Original article

Conflicting associations between dietary patterns and changes of anthropometric traits across subgroups of middle-aged women and men

Li Jiang, Karine Audouze, Jose Alejandro Romero Herrera, Lars Henrik Ångquist, Sonny Kim Kjærulf, Jose M.G. Izarzugaza, Anne Tjønneland, Jytte Halkjær, Kim Overvad, Thorkild I.A. Sørensen, Søren Brunak

**Abstract**

**Background:** Individuals respond differently to dietary intake leading to different associations between diet and traits. Most studies have investigated large cohorts without subgrouping them.

**Objective:** The purpose was to identify non-uniform associations between diets and anthropometric traits that appeared to be in conflict with one another across subgroups.

**Design:** We used a cohort comprising 43,790 women and men, the Danish Diet, Cancer and Health Study, which includes a baseline examination at age 50–64 years and a follow-up about 5 years later. The baseline examination involved anthropometrics, body fat percentage, a food frequency questionnaire and information on lifestyle. From the questionnaire data we computed association rules between the intake of food groups and changes in waist circumference and body weight. Using association rule mining on subgroups and gender-specific cohorts, we identified non-uniform associations. The two gender-specific cohorts were stratified into subgroups using a non-linear, self-organizing map based method.

**Results:** We found 22 and 7 cases of conflicting rules in 8 participant subgroups for different anthropometric traits in women and men, respectively. For example, in a subgroup of women moderate waist loss was associated with a dietary pattern characterized by low intake in both cabbages and wine, in conflict with the association trends of both dietary factors in the female cohort. The finding of more conflicting rules in women suggests that inter-individual differences in response to dietary intake are stronger in women than in men.

**Conclusions:** This combined stratification and association discovery approach revealed epidemiological relationships between dietary factors and changes in anthropometric traits in subgroups that take food group interactions into account. Conflicting rules adds an additional layer of complexity that should be integrated into the study of these relationships, for example in relation to genotypes.

© 2019 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
1. Introduction

The association of dietary intake patterns and development of obesity are influenced by genetics [1–5], gene-diet and diet–diet interactions [6–9] manifesting respondent differences [10–12]. Synergistic or antagonistic effects of specific dietary factors could for example play a role that is not handled well by a conventional regression analysis [13–15].

Contributions to the complexity of diet–obesity relationships could also come from the nature of the anthropometric traits themselves that could be associated with dietary factors in different ways. Studies have for example suggested that animal protein intake is inversely related to waist circumference changes while it was positively associated with weight gain [16,17]. Also, subsequent weight changes could be dependent on the baseline metabolic state indicated by concomitant traits [18]. Adjustment for the baseline traits may lead to bias if the exposure factor is also associated with the baseline phenotype [19].

While respondent differences may be resolved by proper sub-grouping of individuals, some of the difficulties in the interpretation of dietary effects may also relate to the way dietary components are defined as food group categories. Studies have for example revealed inconsistent associations between single nutrients (carbohydrate), food substances (fibre) and food groups (whole grain cereals) with identical phenotypic outcome [17,20,21]. It has also been suggested that investigation of food patterns would benefit from methods that take combined effects of correlated food groups better into consideration [20]. In the present study, we investigate to the extent possible all combinations of food units in order to detect synergistic and antagonistic effects in the joint intake.

Earlier reviews of inconsistent associations found between BMI (Body Mass Index), obesity and dietary patterns have also linked to differences in ways of expressing the diet such as a priori designed diet indexes or groupings based on factor analyses or cluster analyses [22]. Compared to conventional food questionnaire data analysis using energy partition regression [23] that reveals linear relationships between dietary items and different traits, other data mining methods can handle large data sets in a non-linear and non-additive fashion taking exposure interactions into account [13,14].

Recently, association rule mining and clustering algorithms have been successfully used to extract useful knowledge for diabetes and high blood pressure from the National Health Examination Survey [24]. Here, we use a recently developed method, Compass [14], to capture such novel and interesting non-additive associations. This method combines Association Mining (AM) [25] and Self-Organizing Map (SOM) [26–28] based clustering. The method is able to capture heterogeneity and can stratify large cohorts into participant subgroups, and identify associations between the combination of several dietary exposures and specific anthropometric traits in each subgroup or the associations in the entire cohort.

2. Materials and methods

2.1. Population

The Danish, Diet, Cancer and Health study (DCH) is a prospective cohort study of 57,053 women and men from Denmark, enrolled in the period from 1993 to 1997 [29]. Participants were in the 50–64 year age range. Participants filled in questionnaires concerning dietary intake (FFQs) and lifestyle, and anthropometric measures were obtained. In this work, information on diet, lifestyle and anthropometric measurements, participants’ age, education and inhabitant in the cities of Aarhus or Copenhagen, Denmark, were analysed.

2.2. Dietary data

The dietary information was extracted from the FFQs. Intakes at the level of food items were directly taken from each question/answer and then multiplied with a gender-specific portion size factor [20]. The dietary variables were represented in three levels (from broad to specific): food category (FC), food group (FG) and food item (FI). The mapping between specific food items, their food groups and food categories has previously been established [20] (Supplementary Table ST1). For instance as a FC, fruits, is made up by three FGs, respectively, citrus fruits, other fruits and nuts. One specific FG, citrus fruits, comprises two FIs: orange and grape fruits. Beverages are part of the dietary variables divided into nine groups: coffee, tea, water, soft drink, fruit juice, vegetable juice, beer, wine, spirit or brandy at the food group level. The intakes at the level of FIs were derived from the FIs [20]. For the present study a total of 21 FIs [20], 47 FIs and 174 FIs, respectively, were used to examine the dietary intake profile of each individual (Supplementary Table ST1). However, not all levels were used for all analyses (see Results).

2.3. Anthropometric traits

At baseline, all individuals had weight, height, waist circumference, and hip circumference measured in light underwear and no shoes [20]. In a five-year follow-up study participants measured themselves, their weight and waist circumference according to instructions. BMI was calculated from both baseline and follow-up values.

Participants with missing values at baseline or follow-up were excluded as well as participants with extremely small (<55 cm for women or <60 cm for men) or large (>155 cm for both men and women) waist circumference. Participants belonging to either extreme (1%) of the gender-specific total energy intake were also excluded. Participants with an annual weight change larger than 5 kg, or with an annual change in waist circumference larger than 7 cm, or with a total sum of all kinds of physical activity (incl. walking) over 105 h/week (to match the total activity in some circumstances, e.g., retired/housewives), were not taken into consideration. Based on these criteria [20] the final data set contained 23,195 women and 20,595 men.

2.4. Lifestyle and other information

Lifestyle habits included frequency in alcohol intake (non-drinker, below once/month, 1–3 times/month, once/week, 2–4 times/week, 5–6 times/week, or everyday), smoking status (never, previously, or currently), daily consumption of tobacco (grams) and weekly physical activity (hours). Participants’ age, education (less than and equal to 7 years, 8–10 year, or more than 10 years), and residence area (Aarhus or Copenhagen) were also examined as potential influencing factors.

2.5. Exploration of associations between anthropometric changes and dietary factors

We studied the association between dietary factors and anthropometric changes at three levels: the broader FC level, with anthropometric changes and dietary patterns have also linked to differences in ways of expressing the diet such as a priori designed diet indexes or groupings based on factor analyses or cluster analyses [22]. Compared to conventional food questionnaire data analysis using energy partition regression [23] that reveals linear relationships between dietary items and different traits, other data mining methods can handle large data sets in a non-linear and non-additive fashion taking exposure interactions into account [13,14].

Recently, association rule mining and clustering algorithms have been successfully used to extract useful knowledge for diabetes and high blood pressure from the National Health Examination Survey [24]. Here, we use a recently developed method, Compass [14], to capture such novel and interesting non-additive associations. This method combines Association Mining (AM) [25] and Self-Organizing Map (SOM) [26–28] based clustering. The method is able to capture heterogeneity and can stratify large cohorts into participant subgroups, and identify associations between the combination of several dietary exposures and specific anthropometric traits in each subgroup or the associations in the entire cohort.

2. Materials and methods

2.1. Population

The Danish, Diet, Cancer and Health study (DCH) is a prospective cohort study of 57,053 women and men from Denmark, enrolled in the period from 1993 to 1997 [29]. Participants were in the 50–64 year age range. Participants filled in questionnaires concerning dietary intake (FFQs) and lifestyle, and anthropometric measures were obtained. In this work, information on diet, lifestyle and anthropometric measurements, participants’ age, education and inhabitant in the cities of Aarhus or Copenhagen, Denmark, were analysed.

2.2. Dietary data

The dietary information was extracted from the FFQs. Intakes at the level of food items were directly taken from each question/answer and then multiplied with a gender-specific portion size factor [20]. The dietary variables were represented in three levels (from broad to specific): food category (FC), food group (FG) and food item (FI). The mapping between specific food items, their food groups and food categories has previously been established [20] (Supplementary Table ST1). For instance as a FC, fruits, is made up by three FGs, respectively, citrus fruits, other fruits and nuts. One specific FG, citrus fruits, comprises two FIs: orange and grape fruits. Beverages are part of the dietary variables divided into nine groups: coffee, tea, water, soft drink, fruit juice, vegetable juice, beer, wine, spirit or brandy at the food group level. The intakes at the level of FIs were derived from the FIs [20]. For the present study a total of 21 FIs [20], 47 FIs and 174 FIs, respectively, were used to examine the dietary intake profile of each individual (Supplementary Table ST1). However, not all levels were used for all analyses (see Results).

2.3. Anthropometric traits

At baseline, all individuals had weight, height, waist circumference, and hip circumference measured in light underwear and no shoes [20]. In a five-year follow-up study participants measured themselves, their weight and waist circumference according to instructions. BMI was calculated from both baseline and follow-up values.

Participants with missing values at baseline or follow-up were excluded as well as participants with extremely small (<55 cm for women or <60 cm for men) or large (>155 cm for both men and women) waist circumference. Participants belonging to either extreme (1%) of the gender-specific total energy intake were also excluded. Participants with an annual weight change larger than 5 kg, or with an annual change in waist circumference larger than 7 cm, or with a total sum of all kinds of physical activity (incl. walking) over 105 h/week (to match the total activity in some circumstances, e.g., retired/housewives), were not taken into consideration. Based on these criteria [20] the final data set contained 23,195 women and 20,595 men.

2.4. Lifestyle and other information

Lifestyle habits included frequency in alcohol intake (non-drinker, below once/month, 1–3 times/month, once/week, 2–4 times/week, 5–6 times/week, or everyday), smoking status (never, previously, or currently), daily consumption of tobacco (grams) and weekly physical activity (hours). Participants’ age, education (less than and equal to 7 years, 8–10 year, or more than 10 years), and residence area (Aarhus or Copenhagen) were also examined as potential influencing factors.

2.5. Exploration of associations between anthropometric changes and dietary factors

We studied the association between dietary factors and anthropometric changes at three levels: the broader FC level, with anthropometric changes and dietary patterns have also linked to differences in ways of expressing the diet such as a priori designed diet indexes or groupings based on factor analyses or cluster analyses [22]. Compared to conventional food questionnaire data analysis using energy partition regression [23] that reveals linear relationships between dietary items and different traits, other data mining methods can handle large data sets in a non-linear and non-additive fashion taking exposure interactions into account [13,14].

Recently, association rule mining and clustering algorithms have been successfully used to extract useful knowledge for diabetes and high blood pressure from the National Health Examination Survey [24]. Here, we use a recently developed method, Compass [14], to capture such novel and interesting non-additive associations. This method combines Association Mining (AM) [25] and Self-Organizing Map (SOM) [26–28] based clustering. The method is able to capture heterogeneity and can stratify large cohorts into participant subgroups, and identify associations between the combination of several dietary exposures and specific anthropometric traits in each subgroup or the associations in the entire cohort.
variables are listed in Supplementary Table ST1. Prompted by previous studies where gender differences in body fat distribution as well as in associations between dietary factors and adiposity traits and their changes over time were observed [20,29], all analyses were gender-specific. Associations between FCs and five-year change in waist circumference were examined by association rule mining based on the gender-specific cohorts (Analysis I). Associations between dietary factors (FGs or FIs) and annual anthropometric changes were also examined by association rule mining based on gender-specific cohorts, and, in addition, based on different subgroups of participants from the gender-specific cohorts, where subgroups shared similar dietary intake according to either FGs or FIs, 7 lifestyles and 10 anthropometric traits (full profile) (Analysis II).

The five-year changes in waist circumference and weight to be applied in Analysis I were calculated as the difference between baseline and follow-up values, divided by the exact time period between baseline and follow-up to standardize for different lengths of follow-up time, and then multiplied by a five-year period. For Analysis II, average change in waist circumference and weight per year were calculated accordingly after follow-up time normalization.

In total ten traits (seven baseline and three follow-up) were in different combinations included in Analysis II: BMI (baseline), BMI (five year follow-up), body fat (kg), body fat percentage (%), height (cm), hip (baseline) (cm), waist circumference (baseline) (cm), waist circumference change (annual average) (cm), weight (baseline) (kg), and weight change (annual average) (kg).

The strategies for Analyses I and II were as follows:

- **Analysis I:** As a benchmark study, analysis of the gender-specific cohorts with the association rule mining technique was compared to the earlier applied energy partition method (EPM) [20]. The comparative correlation analysis was based on a conventional multiple linear regression model, where the associations between five-year changes in waist circumference and dietary variables were modelled in a linear and additive way using 21 FCs (including beverages) defined earlier [20]. In parallel, association mining was performed at the FG level to reveal the gender-specific cohort trends for the 47 food groups, and at the FI level to reveal the gender-specific cohort trends for the 174 food items.

- **Analysis II:** The gender-specific data sets were clustered twice into subgroups according to their dietary intake (FG or FI, separately), lifestyle factors and anthropometric traits. Association mining was performed for each subgroup to reveal associations between annual waist or weight change and FG, and associations between annual waist or weight change and FI. Gender-specific cohort association rules were compared to subgroup association rules to reveal dietary patterns in which one or two dietary factors did display a conflicting effect on the anthropometric changes, when comparing subgroups to the overall trend in each gender-specific cohort.

### 2.6. Stratification and association rule learning by Compass-2

We explored the relationships between dietary factors and anthropometric changes using the recently developed method, Compass [14]. This method hybrids two data mining approaches: Association Mining and Self Organizing Map clustering for pattern searching in large data sets, such as questionnaires, clinical or laboratory measurements, biobank or registry data. The essence of Compass is to generate variable groups, e.g. a subset of dietary factors, lifestyle factors and anthropometric traits, which discriminate a subgroup of participants from the average of a population, and then to reveal the association pattern between these variables and their statistical significance. The present study used an updated version, Compass-2, implemented in Python in order to support modularization and efficiency when mining large data sets. The software for making statistical tests was written in Python as well.

In Analysis I only the Association Mining part was used as a benchmark that was carried out for the full gender-specific cohorts (as in the previous study [20]). For Analysis II (see Fig. 1 for overview) the workflow in Compass-2 (for details, see Supplementary M&M) can be summarized in six steps:

1. Normalization of each of numeric variables in the full data set (mean = 0, standard deviation = 1).
2. Stratification of the gender-specific cohorts into groups based on SOM, which used dietary variables (47 FGs or 174 FIs, separately), 7 lifestyle variables and 10 anthropometric variables for all individuals as input to group the participants.
3. Aggregation of similar groups into large clusters (subgroups) using hierarchical clustering.
4. Within each subgroup identifying and selecting variables that a) strongly deviated from the gender-specific cohort average as characterized by the attributes H (high) or L (low), or b) were highly similar to the cohort average with strict, low variance as characterized by the M (Intermediate) attribute.
5. Converting numerical data in each subgroup into binary form: present [1] or absent (0), respectively, for each of the selected variables.
6. Learning association rules between the selected variables within each subgroup.

### 2.7. Association mining

Association rules are computed in the form:

\[ \text{Antecedent(s)}: \text{Attribute (H,M or L)} \rightarrow \text{Consequent}: \text{Attribute (H,M or L)} \]

For example:

1. **Variable A:H** $\rightarrow$ **Variable B:L**
2. **Variable A:H, Variable C:L** $\rightarrow$ **Variable B:H**

Meaning that:

- High exposure of Variable A leads to Low values of Variable B.
- High exposure of Variable A with Low exposure of Variable C leads to High values of Variable B.

Two types of associations were defined: (a) direct association between a specific diet factor and an anthropometric outcome (“1-to-1” relationship), and (b), joint effect exhibited by two diet factors and one anthropometric outcome (“2-to-1” relationship). Distinguishing these two association types allows us to identify cases of conflicting rules (explained below).

### 2.8. Strength and statistical significance of association rules

A number of parameters commonly measured in association mining were used to define each association rule [25]. These parameters were: (a) the frequency (Support) of participants in the gender-specific cohort (or in a subgroup) that fulfil the rule, and (b), the conditional probability (Confidence) of observing a rule given the Support of its respective antecedent (for details see Supplementary M&M).

Right-tailed Fisher Exact tests were used to calculate the statistical significance of each association rule. P-values were adjusted...
with Benjamini-Hochberg False Discovery Rate (FDR) estimation [31]. Rules with an adjusted p-value >0.05 were discarded.

2.9. Distance between subgroups and the gender-specific cohort averages

Distances between gender-specific cohort averages and their respective subgroups were calculated by two methods: (a) the mean of pair-wise Euclidean distances between participants in the subgroups and participants in the cohort, and (b), the Euclidean distances based on the scaled mean (z-score) of each variable in the subgroup (see Supplementary M&M). We further distinguished between two types of distances: Distance of Full Profile (DFP) that depended on all variables: diet, lifestyle and anthropometric traits, and Distance of Dietary Profile (DDP), which was only based on dietary factors.

2.10. Conflicting rules

The gender-specific cohorts were stratified into subgroups whose participants have quite similar dietary intake profiles, lifestyle and anthropometric traits. This allowed us to identify conflicting rules, defined as anthropometric change–diet associations within individual subgroups (type (b) rule) that differ from its gender-specific trend (type (a) rule) in the entire gender-specific
cohort. A general example of a pair of conflicting rules would be as
follows:

\[
\text{Diet A: H} \rightarrow \text{Waist change:L} \quad (\text{gender-specific cohort, type (a) rule})
\]

\[
\text{Diet A: H, Diet B: L} \rightarrow \text{Waist change:H} \quad (\text{gender-specific subgroup, type (b) rule})
\]

Meaning that High intake of Diet A is associated with Low Waist change as
general trend in the gender-specific cohort. However, a
gender-specific subgroup contradicts this general trend since High
intake of Diet A leads to High Waist change, when taking into account
Low exposure of Diet B.

Conflicting rules of potential high interest were detected at the
FG/FI levels following a multistep procedure (see SM&M).

2.11. Impact of non-dietary influencing factors

We considered lifestyle habits and age, education and area of
residence as potential non-dietary influencing factors. A pair of
conflicting rules was required to have an association adjusted p-
value lower (see Supplementary M&M) than the adjusted p-value
measured by replacing any of the two food groups (or two food
items) with a potential influencing factor (chosen among the
available variables).

2.12. Baseline traits in the analyses of anthropometric changes

Since there is an inherent tendency for anthropometric traits to
regress toward the mean [19], in this study, annual waist circum-
ference change and annual weight change were analysed individ-
ually in two different ways [1]: being included as the only trait in
a rule model and analysed independently, and [2] being included as
the responding trait in a rule model where the baseline level of
waist circumference, weight and BMI were considered as additional
confounding factors. The association rules that involved the follow-
up traits were further filtered using the same procedure as
described above, requiring the rule model without the baseline trait
included to have a lower adjusted p-value than when included,
eliminating influence of confounding factors.

2.13. Main driver for conflicting rules at the FI level

Each case of conflicting rules identified at the FG level was
investigated as to whether there is any dietary factor(s) at the FI
level that could contribute most to the conflicting effect of a FG.
According to the mapping between FG and FI (Supplementary
Table ST1), we transformed each pair of FG "1-to-1" association
(gender-specific cohort) and "2-to-1" association (a FG subgroup)
into FI associations, where each FG was respectively substituted by
the corresponding FI(s). The main driver(s) for a pair of FG in
conflicting rules was identified at the FI level, if 1) both the "1-to-1"
association in the gender-specific cohort and the "2-to-1" associa-
tion in a FI subgroup were established significantly at the FI level,
and 2) the conflicting trend of the dietary factor (FI) in the "1-to-1"
association was revealed when comparing its tendency to the
corresponding tendency in the "2-to-1" association.

3. Results

3.1. Analysis I: General associations in the gender-specific cohorts

The overall associations between five-year change in waist
circumference (5y-WC) and the dietary factors (21 food and
beverage FCs) were identified by association mining in the full
gender-specific cohorts using both food items and food groups as
variables. We initially performed a benchmark study comparing the
association rules identified by association mining to the significant
findings (Table 1) identified in the earlier use of the energy parti-
tion method EPM (a full list of results from the previous study can
be found in Supplementary Table ST2) [20]. Using the 21 food
categories Compass-2 retrieved the majority of the associations
found earlier including vegetables (low intake vs gain), fruits (low
intake vs gain), red meat (high intake vs moderate change), high-fat
dairy (low intake vs gain), butter (low intake vs gain) and snack
foods (low intake vs moderate change) for women; fruits (low
intake vs gain) and red meat (high intake vs moderate change) for
men (Table 1). Among these confirmed associations with 5y-WC,
high-fat dairy appeared to be less statistically significant (adjusted
p-value of 0.051). Associations involving two of the FCs were not
reproduced (processed meat for women and snack food for men).
However, these two food categories did show a similar trend where
high intake would promote 5y-WC gain as indicated by EPM.

Two associations were identified in opposite directions. In
women, low intake of potatoes increases 5y-WC, while those with
high intake of poultry display a moderate change. These results are
different compared to the earlier study, where potatoes and poultry
promoted change in WC or BMI for men [20] and women [21]. How-
ever, both dietary factors have a very weak correlation (−0.004 and
0.004, respectively) with 5y-WC, which might explain the
different results obtained by the two methods.

For comparison we went one step further and investigated the
benchmark results of the gender-specific cohorts using the 47 more
fine-grained FCs. For example, the food category, fruits, from the
EPM study was further divided into three different FGs: citrus fruits,
other fruits, and nuts. For women, the role of inhibiting waist in-
crease by the food category vegetables was confirmed in a broad
range of FGs (including fruiting vegetables, leafy vegetables, root
vegetables, cabbages, mushrooms, stalk vegetables and onions),
but not in several other vegetable FGs (including legumes and soya).

Nuts were considered part of the food category fruits that dis-
plays an overall effect of inhibiting waist gain. However, when evalu-
ating nuts as an individual food group, low intake of nuts was
associated with waist decrease. Men also displayed different trends
for individual FGs that belong to the same food category. For
example, moderate waist decrease is associated with either low
intake of chocolate and candy bars, or high intake of pork rind snacks.
The reverse association in pork rind snacks might explain why snack
(including chocolate, candy bars and pork rind snacks) as a joint food
category was not detected as a significant association with waist
change in the earlier study [20]. An overview of significant asso-
ciations between individual FGs and anthropometric changes is
shown in Supplementary Figures SF1.

3.2. Analysis II: Stratification and identification of conflicting
association rules between subgroups and the gender-specific cohort

To decipher the difference between diet--trait associations
within homogeneous subgroups and associations representing the
general pattern found in gender-specific cohort, we clustered the
gender-specific cohorts into participant subgroups and performed
association mining within each subgroup for associating dietary
factors with annual waist or weight changes.

3.2.1. Stratification

We used SOM to stratify the gender-specific cohorts based on
the similarity of individual dietary, lifestyle and anthropometric
trait profiles. The initial number of SOM units was examined,
leading to a division of the gender-specific cohorts into 784 small
subgroups (see Supplementary M&M for details). Subsequent hierar-
chical clustering (Supplementary Figure SF2, optimal clustering in
Supplementary M&M) was used to merge the 784 small groups into
78 (women) and 73 (men) larger subgroups of high internal
similarity using FGs, and into 115 (women) and 119 (men) subgroups when using FIs, respectively (see Table 2).

For further analysis, we focused on subgroups with at least 200 participants to ensure that each subgroup was sufficiently large to represent a relevant subset of the cohort. For instance, using FG dietary profiles, 51 subgroups in women had 200 or more participants. These subgroups account for 65% of all (78) subgroups, and represent 84% of all participants in the cohort of women. The proportions for the 200 participants are 58% of all subgroups and 80% of all participants for the 42 large male subgroups, respectively.

### 3.2.2. Distance network of subgroups clustered with FG dietary profiles

Figure 2 shows the DFP distances based on the mean of pairwise Euclidean distances (method a) between subgroups (circular nodes) and the cohort (hexagon node) average, and between pairs of subgroups. The variance of distances within each subgroup was not strikingly different, only approximately a two-fold difference between the minimum and maximum. Thus, the within-subgroup level of heterogeneity (variation) is similar between different subgroups, indicating a good stratification of the gender-specific cohorts when combining SOM and hierarchical clustering.

#### 3.2.2.1. Women

A number of subgroups (FG-0003, FG-00028 and FG-00031) were outliers, as they have highly dissimilar full profiles (including dietary intakes, lifestyle patterns and anthropometric traits) to the mean of cohort (Fig. 2A). Only one outlier FG-0024 has a notable, moderate waist change accompanied by high BMI in baseline. However, they are all relatively small subgroups with less than 200 participants. The largest subgroups (>500 participants) in women, including FG-0042, FG-0046, FG-0058, FG-0059 and FG-0070, displayed intermediate-to-large differences from the cohort average.

The subgroups with the most striking increase in waist circumference (subgroup border in thick pink) were found in FG-0050, FG-0040, FG-0018, FG-0006, and FG-0051. Among these subgroups, FG-0018 and FG-0051 deviated highly from the cohort average. FG-0051 was characterized by high baseline BMI (red subgroup filling). The rest of the waist-gaining subgroups had low BMI (blue subgroup filling), except FG-0050 where no distinguishable BMI feature (high, low or intermediate) was identified (subgroup filling white).

In contrast, subgroups with the most remarkable moderate change in waist circumference (subgroup border thick green) were FG-0033, FG-0010, FG-0012 having intermediate BMI (yellow subgroup filling), FG-0049 with low BMI, and FG-0074, FG-0013 and FG-0022 with high BMI, respectively. The largest subgroup FG-0058 (1927 participants) was also characterized by notable moderate change in waist circumference and high BMI in baseline, where the full profile was evidently different from the cohort mean.

The DFP distance network may be compared to the DDP based network (Supplementary Figures S3F-S4), where the large subgroups in women FG-0058 and FG-0046 displayed a dietary pattern similar to the cohort average. However, both full profiles were strongly deviating when additionally considering lifestyle factors and anthropometric traits. Compared to FG-0058, the alteration in other large subgroups (FG-0042, FG-0046, FG-0059 and FG-0070) in waist change was relatively moderate.

#### 3.2.2.2. Men

According to the DFP full profile distances (Fig. 2B), two small subgroups FG-0046 and FG-0060 were identified as outliers, having only minor waist change and no distinguishable baseline BMI feature (tagged by “H”, “M” or “L”). The largest subgroups FG-0031, FG-0018, FG-0027, FG-0011, and FG-0008 appeared to be notably different from the cohort mean. Among them, remarkable waist alteration was observed in FG-0018 (gain) and FG-0031 (moderate change), where the baseline BMI was observed, respectively, to be low and high. Compared to women where 22% of the subgroups had waist gain or moderate change, many more subgroups (38%) in men showed waist changes. The majority of the waist increase or moderate change was observed among small subgroups (<200 participants). Still, the waist gain or moderate change was a feature found in 12 larger subgroups. For example, moderate change was observed in both FG-0007 and FG-0034, with opposite levels of baseline BMI. On the other hand, FG-0022 and FG-0029 were common for the waist gain and both also had high BMI at baseline.

### 3.2.3 Identification of conflicting rules

For any association between a dietary factor and annual waist or weight change in a gender-specific cohort we searched for
subgroups where the overall trend was inconsistent when co-intake with a second dietary factor was simultaneously considered.

In the gender-specific cohorts, the analysis revealed 22 and 7 cases of conflicting rules associated with anthropometric change (waist circumference or weight) for women and men, respectively, in four subgroups of either gender (Fig. 3, and listed in Supplementary Table ST3-ST4). The number of conflicting rules found in each subgroup is shown in Supplementary Figure SF5. The majority of the conflicting rules in women have association trends differing in both dietary factors (double-dashed edge) in subgroups. Some were caused mainly by one of the dietary factors, with co-occurrence of a second dietary factor (single-dashed edge). While the conflicting rules in men were found mainly due to one of the dietary factors, the co-occurring second factor did not show a conflicting association trend when compared to its overall tendency (summarized in Table 3).

Fig. 2. The stratified cohorts for women and men, based on the full profile distances (based on dietary, lifestyle and anthropometric variables). Each of the 78 female subgroups is labelled with a number (A), similarly for the 73 subgroups for men (B). The length of the bended edge is proportional to the Euclidian distance between nodes: grey edges represent subgroup distances; yellow edges represent distances between the gender-specific cohorts and subgroups. The node border width is proportional to the average annual waist circumference change in the subgroup (moderate change: green; gain: red). The filling colour of each subgroup indicates baseline BMI features (Red: high; Blue: low; Yellow: intermediate; White: subgroup mean is not significantly different from the cohort average or there is a large variation within the subgroup). The node size is proportional to the subgroup size.
To consider the influence from the assumed diet difference, these follow-up traits related conflicting associations were further investigated for the strength of association that was additionally conditional on the reverse feature of baseline anthropometries (waist, weight and BMI). For example, examination of waist gain (moderate change) conditional on low (high) baseline BMI or low (high) baseline waist. The results showed that the strengths of these associations for anthropometric change were in general stronger when the reverse anthropometric features were present at baseline.

3.2.3.1. Women. In women, 60% (13 cases) of the conflicting rules were found in subgroup FG-0010, where five sets (double dashed edges) of dietary factors both showed conflicting trends. All of them were featured by low intake (green filling cycle) and associated with moderate change in weight (green filling rectangle), which is conflicting to their negative correlations (solid edge ending by a spot) with weight change found in the female cohort. In FG-0013, moderate change in waist was associated simultaneously with low intake in both cabbages and wine. Both dietary factors were

Fig. 3. Conflicting rules, for women (A) concerning anthropometric change (waist or weight) shown for four subgroups, 6, 10, 12 and 13. The shape of the variable-nodes indicates their type: round rectangle represents anthropometry change and circle dietary factor. The variable-node filling colour indicates the level of dietary intakes (low: blue; high: red) and the change trends (moderate change: blue; gain: red) in the subgroups. Edges represent the relationships, the solid edge link anthropometric change and a dietary factor in a subgroup that showed a conflicting association trend (indicated by node filling colours) with the co-occurrence (dashed line) of a second dietary factor compared to the overall trend (indicated by shape of line-end: arrow for positive correlation; dot for reverse correlation). (B) Conflicting rules for men relating to the male subgroups 4, 18, 24, and 34.
oppositely associated with waist change according to the female cohort. In FG-0006, weight gain (orange filling rectangle) was associated with low intake in both processed meat and soft drinks, in conflict with the trends in the female cohort, where the intake levels in both dietary factors were positively correlated (solid edge ending by an arrow) with weight change. In FG-0012, both moderate change in waist and weight was observed and associated with different sets of dietary factors.

3.2.3.2. Men. In men, all cases of conflicting rules were involved in waist change (gain or moderate change), and were all connected to a single dietary factor that differed between the subgroups. In contrast to the male cohort, where the intake of leafy vegetables was positively correlated with waist change, in FG-0024, low intake of leafy vegetables was associated with waist gain when the dietary composition was low in one of three dietary factors: processed medium-fat fish, or fatty-fish, or mayonnaises. Similarly in FG-0004, carrots (other root vegetables) displayed a conflicting tendency to its overall trend when at the same time having low intake of fresh lean fish, or onion/garlic. Together with poultry, lower intake of tea in FG-0018 was associated with waist increase, although the overall tendency was opposite. In FG-0034, the positive correlation between waist change and margarine intake (in a combination with stalk vegetables/sprouts) was against to the trend in the male cohort (reversely correlated).

3.2.3.3. Comparison of men and women. In women, conflicting rules were found in subgroups (FG-0006, FG-0010, FG-0012 and FG-0013) where based on the dietary profile they had a shorter distances ($p<0.05$, Wilcoxon rank sum test) to the cohort average. In men, the subgroups (FG-0004, FG-0018, FG-0024 and FG-0034) were not close ($p>0.3$) to the average. There were no very large subgroups (>500 participants) exhibiting conflicting trends in women. In contrast, a conflict regarding tea consumption and waist change was observed in FG-0018, a subgroup of men made up by 896 participants and characterized by waist gain and small baseline BMI. Moreover, there were more subgroups in women (FG-0050, FG-0040, FG-0018, and FG-0051) than men (FG-0031), where the most striking waist changes were observed, but no conflicting rules were identified for the anthropometric changes. The one common dietary factor in both women and men observed as a conflict between the gender-specific cohort and subgroups were leafy vegetables and root vegetables. Root vegetables were most frequently (up to five times in women) found in conflicting association rules (see counts for each dietary factor in Supplementary figure S6).

As described in Methods (the section on conflicting rules) these conflicts did not include examples where conflicts between the gender-specific cohorts and their subgroups were present regardless of the co-intake of a second dietary factor. There was however only one pair of such a conflicting rule observed in the study. It was in subgroup FG-0022 of men where high intake of lean dairy products was associated with weight gain, while the low intake was oppositely in association with weight gain in the male cohort.

3.2.3.4. Main driver for conflicting rules at the food item level. Conflicting associations identified at the FG level can be investigated in terms of conflicts at the food item level for all the items included in the group. Overlapping conflicting trends were not observed for the same anthropometric changes in gender-specific cohorts. However, conflicting dietary patterns were observed for approximately 45% (10 out of 22) of conflicting associations in women for other concomitant anthropometric traits in the data. For example, root vegetables (FG), given a co-intake pattern with non-citron fruits (FG) in women subgroup FG0010, showed a conflicting association with annual weight change compared to its overall trend. This conflicting tendency was observed in a FI dietary pattern, where carrots (FI) and peach (FI) were associated with follow-up BMI. Importantly, these overlaps are investigated using the FG-based stratification of the gender-specific cohorts. A FI-based stratification would produce somewhat different participant subgroups where the exact conflicts would differ.

4. Discussion & conclusion

In this study we explored 43,790 Danish cohort participants 50–64 years of age for conflicting associations between anthropometric changes and dietary patterns. We successfully evaluated association rule learning in a large cohort where dietary habits were assessed by food frequency questionnaires, and we were able to retrieve known association rules, such as the beneficial role of fruits and red meat in relation to changes in waist circumference in both genders [20]. Motivated by the heterogeneity that may elicit individual responses to similar dietary exposures, we stratified the gender-specific cohorts to identify associations between dietary patterns and anthropometric traits in subgroups. Focusing on the conflicting trends between subgroups and the gender-specific cohort, the study led to a number of interesting observations.

4.1. A novel combined stratification and rule finding approach to explore epidemiological relationships

Compass-2 applies the association-mining model, which can detect non-linear associations between variables, to study relationships between dietary factors and anthropometric changes. In this particular type of analysis, the quantification of the dietary factor is assigned in three levels (H, M and L), whereas it is a continuous variable assumed to have a linear association with the outcome in the two other types of earlier analyses (benchmark study) [20,30]. If the effect is non-additive and varies in different levels, for example L effects $< M effects = H effects$ (as in a dominant genetic model), then the ‘linear model’ corresponding to additive effects of L, M and H is not the best fit. On the other hand, if there are additive effects of L, M and H, then the linear model should be better, at least more sensitive. In fact, the linearity of the association between two variables can be examined by comparing the consistency of association trends at different levels. In the benchmark comparison, association mining was shown to be a sensitive approach and capable of identifying relationships confirmed by the well-established epidemiological analysis. However, since the anthropometric changes observed in response to the dietary intakes most likely is working in a non-additive way conventional linear regression methods will be unable to catch such associations [15]. Thus modelling by association mining is a good complement to the energy partition model.

4.2. Phenotyping foods at a more fine-grained level

We have examined the association of anthropometric traits with food categories (benchmark) in the gender-specific cohorts, and with FGs (in the cohorts and in the subgroups). Our data showed...
that individual FGs (e.g. legumes) could have an opposite impact compared to its broader food category FC (e.g. vegetable) for waist change. The study of the main drivers for conflicting trends at the FI level also indicates that not all individual FI belonging to a higher level FG contribute to the contradictory effect observed at the FG level. Thus, fine mapping diet–trait association patterns for more fine-grained food items (e.g. subdivide legumes into soybeans, and chickpeas etc.) could, if consistently replicated, be one of initial steps towards designing a more personalized diet, compared to dietary macronutrient studies [32,33]. The growing differentiation and precision in the diet effect on complex traits such as obesity, diabetes, cardiovascular diseases and cancer may eventually reveal phenotypic impact or disease profiling of various diets, parallel to the nutrients based food categorization [32,33].

4.3. Stratification of the gender-specific cohorts according to dietary, lifestyle and anthropometric profiles

Decomposing the gender-specific cohort revealed distinct subgroups with interesting features. Some subgroups, were for example characterized by “normal” dietary patterns similar to the cohort average (Supplementary Figure SF3), however, they led to “abnormal” anthropometric outcome, as remarkably losing waist circumference (e.g. subgroup FG-0058 in women). This observation implies that subsets of participants although having similar diets as the population average could respond differentially and develop divergent anthropometric phenotypes. There could in principle be many reasons for the heterogeneity between subgroups, embedded in the entire preceding history of the individual from before conception all the way to the baseline assessment [19]; for example this could imply different cumulative exposure to environmental factors such as physical, social, cultural and economical status, and changes of the diet by active (self) interventions [34]. Further investigation into the genetics of these subgroups and potential susceptible pathways is necessary to further understand the aetiology behind these findings [35–39].

4.4. Conflicting patterns

We identified conflicting rules deciphering heterogeneous diet–anthropometric change associations in response to specific dietary patterns. We observed a clear difference between men and women in anthropometric changes at the level of conflicts. In women, conflicting rules for weight or waist circumference changes were found in subgroups where the diet/lifestyle/trait profiles were less remarkably different from the cohort average than the subgroups in men. This implies that subgroups of women having “normal” dietary habits (dietary intake levels close to the average values of the cohort) and at the time unusual weight or waist change can be economically status, and changes of the diet by active (self) interventions [34]. Further investigation into the genetics of these subgroups and potential susceptible pathways is necessary to further understand the aetiology behind these findings [35–39].

4.5. Limitations

Despite revealing interesting (conflicting to the overall trend) associations between dietary patterns and change in weight and waist circumference in subgroups; the specific hypotheses derived from this study need to be replicated by carrying out additional observational or interventional studies. In addition, our benchmark study indicated that association mining established some dietary effects opposite to the results based on linear regression [20] when the correlation between two variables was very weak. One can improve the adjustment of the results for non-dietary influencing factors, including lifestyle and baseline measurements that we have already handled in the present study, but also for the dietary factors themselves, adjusting for the impact of other dietary factors when analysing specific one(s). As stated above, adding genetic data would make it possible to clarify to what extent germline genetics or possibly somatic mutations impact the results [35–39].

4.6. Conclusions

In conclusion, we used data mining approaches to reveal relationships between dietary factors and anthropometric changes in middle aged Danish citizens. We identified subgroups that shared similar dietary, lifestyle and anthropometric profiles. The existence of conflicting rules manifested by specific co-intake patterns in given subgroups displaying other trends than in the gender-specific cohorts, implies a complex relationship between dietary and non-dietary factors and anthropometric traits. The analysis of relationships between diet, non-dietary lifestyle, anthropometric traits, and genetics may eventually be carried out at the level of individuals, in analogy to the development within the general areas of precision medicine and pharmacogenomics [35–39]. Following this line of thought, our method is applicable as well in a clinical nutrition setting, where the focus is the nutrition in sick people, with the aim of supporting the cure or preventing the worsening of their condition, due to nutrition disorders induced by the loss of appetite that frequently accompany many different diseases.

Funding sources

This work is carried out as a part of the research program of Gene-diet interactions in obesity (GENDINOB). GENDINOB is supported by the Danish Council for Strategic Research (Grant 09–067111). The work was also supported by the Novo Nordisk Foundation (Grant agreements NNF14CC0001, NNF17OC0027594) and Innovation Fund Denmark (5184-00102B).

Statement of authorship

The authors’ responsibilities were as follows – AT, JH, KO, TIAS and SB: designed the research; LJ, KA, LHÅ SKK, JARH, JMG1 and SB: conducted the research; AT, JH, KO and TIAS: provided essential data; LJ, KA, LHA SKK, JARH, JMG1 and SB: analysed data or performed statistical analysis; LJ, TIAS and SB: wrote the paper; SB: had primary responsibility for final content; and all authors read and approved the final manuscript.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank Olga Rigina and Ulrik Plesner Jacobsen for the assistance data preprocessing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clnu.2019.02.003.

References


